The shape of discourse:
How gesture structures conversation

by

Schuyler Laparle

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Linguistics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Eve Sweetser, Chair
Professor Line Mikkelsen, Co-chair
Assistant Professor Elise Stickles
Professor Dor Abrahamson

Fall 2022

The shape of discourse:
How gesture structures conversation

Abstract

The shape of discourse:
How gesture structures conversation

by

Schuyler Laparle

Doctor of Philosophy in Linguistics

University of California, Berkeley

Professor Eve Sweetser, Chair

Professor Line Mikkelsen, Co-chair

This dissertation examines the capacity of interactive gesture to contribute to discourse structure independent of accompanying speech. The close relationship between gesture and speech in face-to-face interaction is, at this point, well-established and accepted, especially within gesture studies and certain linguistic frameworks (e.g. Embodied Construction Grammar and Embodied Conversation Analysis). However, the integration of gesture into formal linguistic theory more generally is still in its early stages of development. This is especially true for interactive gesture and formal theories of discourse structure. To the author's knowledge, this dissertation serves as the first in-depth exploration of the ability to formalize a theory of interactive meaning in gesture using a predictive model.

The proposed model is built upon the use and management of an 'Interaction Space' – the physical space in which co-present interlocutors interact and the metaphoric space in which they co-construct a goal-based, hierarchical discourse structure together. I propose an inventory of management actions performed upon this space, each of which is enacted by a particular kind of interactive gesture. PRESENT, REFER and REMOVE actions constitute the three primary actions performed by interactive hand gestures and serve to introduce, organize, and remove discourse topics as metaphoric objects from the Interaction Space. ENGAGE and DISENGAGE actions are the two primary actions performed by non-manual gestures and serve to manage interlocutors' roles as speaker or addressee throughout a discourse. Three additional management actions, SEPARATE, COMBINE and REQUEST are composite actions used to achieve more complex forms of discourse management, such as expressing a particular discourse relation or the desire for an interlocutor to perform particular actions in response.

After an initial explanation of the model and associated formalism, I apply the model in

three case studies. Each case study looks at the use of gesture to express a particular discourse move – *topic-shifting, digression,* and *specification*, as co-expressed by the lexical discourse markers, *anyway*, *by the way*, and *here's the thing.* All data comes from interviews and monologues on the American talk show *The Late Show with Stephen Colbert* and were collected using UCLA'S Communication Studies Archive in collaboration with the Red Hen Lab. Through quantitative and close qualitative analyses, I demonstrate the capacity of interactive gesture to perform discourse management, and the ability of the Interaction Space to parsimoniously model how this is done.

Because of the interdisciplinary nature of this dissertation, I draw from a range of theories in my analyses and discussions. For describing the form and function of gesture, I rely on action schematic approaches to gesture meaning (Cienki 2013; Mittelberg 2018; Müller 2017). For interpreting gestures in context, I use Conceptual Metaphor Theory (Lakoff & Johnson 1980) and Blending (Fauconnier & Turner 1998). For analyzing discourse structure, I borrow from question-based approaches (van Kuppevelt 1995; Roberts 1996, 2012). It is uncommon to find such disparate frameworks together in a single work. In bringing these approaches together, I hope to demonstrate the value of inter-theory and inter-field dialogues to the development of a comprehensive model of multimodal language use.

In loving memory of Coni Henson

who always knew that I could do it,
even when I was sure that I couldn't

Contents

Contents **ii**

List of Figures **iv**

List of Tables **vii**

1 In pursuit of a multimodal linguistics **1**
1.1 Introduction 1
1.2 Contributions of this work 3
1.3 What is gesture and what does it do? 6
1.4 Is gesture a part of language? 12
1.5 Multimodal linguistics as a stronger linguistics 26

2 Describing gesture and multimodal data **28**
2.1 Introduction 28
2.2 Data collection and presentation 28
2.3 Strengths and limitations of a one-show data set 34
2.4 An integrated methodology for gesture description 38
2.5 Interpretation in context 56
2.6 Applying the methodologies 59

3 Constructing a discourse **60**
3.1 Introduction 60
3.2 Communicative pressures in face-to-face conversation 60
3.3 Modeling discourse goals and contributions 65
3.4 Metaphor in discourse theory 71
3.5 An embodied discourse structure 77

4 The Interaction Space **78**
4.1 Introduction 78
4.2 Traditional approaches to gesture position 80
4.3 Defining the Interaction Space 82
4.4 Managing engagement 90

4.5 Managing contents 111
4.6 Discussion 122

5 What is multimodal topic-shifting *anyway*? 125
5.1 Introduction 125
5.2 Data & Methodology 129
5.3 Variation in the verbal mode 130
5.4 Variation in the gestural mode 155
5.5 Topic Dismissal 159
5.6 Topic Introduction 180
5.7 Complex sequences 193
5.8 Summary 201

6 *By the way,* digression is multimodal 202
6.1 Introduction 202
6.2 Data & Methodology 207
6.3 Variation in the verbal mode 208
6.4 Variation in the gestural mode 220
6.5 Discussion 253

7 *Here's the thing* about multimodal specification 255
7.1 Introduction 255
7.2 Data & Methodology 257
7.3 Variation in the verbal mode 258
7.4 Variation in the gestural mode 277
7.5 Conclusion 314

8 Concluding remarks 315
8.1 Summary of findings 316
8.2 Theoretical implications 318
8.3 Future directions 319

Bibliography 320

List of Figures

1.1 Movement trajectories compatible with the lexical affiliate "up" 8
1.2 Three gestures with "go up" 15

2.1 Search page in the UCLA News Archive 29
2.2 Result page in the UCLA News Archive 30
2.3 Export file contents for corpus search results 30
2.4 Excel annotation sheet for *here's the thing* data set 32
2.5 Data presentation for verbal context only (example taken from Ch.7) 33
2.6 Data presentation for manual and non-manual gesture in verbal context (example taken from Ch.6) 33
2.7 The Late Show with Stephen Colbert in the Ed Sullivan Theater 34
2.8 Differences in spatial organization across interaction types 35
2.9 Unique spatial affordances across interaction types 36
2.10 An integrated methodology for gesture description, inspired by proposed procedures in Bavelas (1994) 39
2.11 Palm orientations as performed by actor Jennifer Lawrence 42
2.12 Annotated hand shapes with fixed forms 43
2.13 Annotated hand shapes without fixed forms 44
2.14 Annotated hand movements with fixed forms 45
2.15 Examples of hand movements annotated as "other" 45
2.16 Forms of presentational gestures 48
2.17 Forms of removal gestures 50
2.18 Forms of stopping gestures 52
2.19 Forms of referring gestures 53
2.20 Forms of reorientation 55

3.1 Levels of analysis 72

4.1 The delineation of the Interaction Space 83
4.2 The Interaction Space as a blended concept 86
4.3 Disengagement with the Interaction Space and breaking of the speaker-hearer line. 87
4.4 Locating topics within the Interaction Space 88
4.5 Gesture sequence corresponding to Transcript 1: Elizabeth Warren, ex. 35. . . . 94

4.6 Gesture sequence corresponding to Transcript 2: Eddie Redmayne, ex. 38. . . . 98
4.7 Gesture sequence corresponding to TRANSCRIPT 3: JENNIFER LAWRENCE, ex. 40. 101
4.8 Blocking gestures by Jennifer Lawrence (left; ex. 40) and Stephen Colbert (right; ex. 38) 103
4.9 Gesture sequence corresponding to TRANSCRIPT 4: KATHY GRIFFIN, ex. 42. . 106
4.10 Gesture sequence corresponding to TRANSCRIPT 5: LIN-MANUEL MIRANDA, ex. 44 109
4.11 Gesture sequence corresponding to TRANSCRIPT 6: NATALIE PORTMAN, ex. 46 114
4.12 Gesture sequence corresponding to TRANSCRIPT 7: ANDREW SCOTT, ex. 48 . 117
4.13 Gesture sequence corresponding to Transcript 8: John Dickerson, ex. 50 120

5.1 Three classes of dismissal gesture 159
5.2 Removal gestures corresponding to TRANSCRIPTS 17-20, ex. 79-86 161
5.3 Gesture sequence corresponding to TRANSCRIPT 23: STEPHEN COLBERT, ex. 87.168
5.4 Gesture sequence corresponding to TRANSCRIPT 24: STEPHEN COLBERT, ex. 88 173
5.5 Gesture sequence corresponding to TRANSCRIPT 25: DREW BARRYMORE, ex. 89.175
5.6 Gesture sequence corresponding to TRANSCRIPT 26: KATHY GRIFFIN, ex. 90. 177
5.7 Classes of topic-introduction gestures aligned with "anyway" 180
5.8 Sub-types of presentation gestures 183
5.9 Gesture sequence corresponding to TRANSCRIPT 30: EMMA THOMPSON, ex. 94. 188
5.10 Gesture sequence corresponding to TRANSCRIPT 31: STEPHEN COLBERT, ex. 97.191
5.11 Gesture sequence corresponding to TRANSCRIPT 32: CONAN O'BRIEN, ex. 98. 192
5.12 Gesture sequence corresponding to TRANSCRIPT 33: RENEE ZELLWEGER, ex. 99 194
5.13 Gesture sequence corresponding to TRANSCRIPT 34: NICK KROLL, ex. 100 . . 197
5.14 Gesture sequence corresponding to TRANSCRIPT 35: SAOIRSE RONAN, ex. 101 200

6.1 Gesture sequence corresponding to TRANSCRIPT 14: JOHN DICKERSON, ex. 120. 223
6.2 Gesture sequence corresponding to TRANSCRIPT 15: ELIZABETH BANKS, ex. 121.224
6.3 Gesture sequence corresponding to TRANSCRIPT 16: EDDIE REDMAYNE, ex. 122.225
6.4 Four classes of hand gestures associated with BTW 227
6.5 Gesture sequence corresponding to TRANSCRIPT 17: NANCY PELOSI, ex. 123. 232
6.6 Gesture sequence corresponding to TRANSCRIPTS 18-20, ex. 124 234
6.7 Gesture sequence corresponding to TRANSCRIPT 21: STEPHEN COLBERT, ex. 125. 235
6.8 Gesture sequence corresponding to TRANSCRIPT 22: SUSAN RICE, ex. 126. . . 237
6.9 Gesture sequence corresponding to TRANSCRIPT 23: ELIZABETH WARREN, ex. 127. 240
6.10 Trajectories of locating gestures with BTW 241
6.11 Addressing gestures with BTW 242
6.12 Gestures corresponding to 246
6.13 Addresse-directed presentation with BTW 247

6.14 Gesture sequence corresponding to Transcript 30: Rachel Maddow, ex. 135. 250
6.15 Gesture sequence corresponding to Transcript 31: John Lithgow, ex. 136. 251
6.16 Gesture sequence corresponding to Transcript 32: Andrea Savage, ex. 137. 253

7.1 Types of presentation gestures as observed with HTT 283
7.2 Gesture sequence corresponding to Transcript 15: Colbert, ex. 153 286
7.3 Gesture sequence corresponding to Transcript 16: Elizabeth Warren & Colbert, ex. 157 290
7.4 Gesture sequence corresponding to Transcript 17: Kamala Harris, ex. 161 295
7.5 Gesture sequence corresponding to Transcript 18: Stephen Colbert, ex. 163 297
7.6 Gesture sequence corresponding to Transcript 19: Stephen Colbert, ex. 165 300
7.7 Gesture sequence corresponding to Transcript 20: Stephen Colbert & Mindy Kahling, ex. 168 302
7.8 Types of stopping gestures as observed with HTT 305
7.9 Gesture sequence corresponding to Transcript 23: Jon Lovett, ex. 172 . . 309
7.10 Gesture sequence corresponding to Transcript 24: Nancy Pelosi, ex. 175 . 312

List of Tables

2.1 Primary action schemas 47
2.2 Frame structure of presentational gestures 49
2.3 Frame structure of removal gestures 51
2.4 Frame structure of stopping gestures 52
2.5 Frame structure of stopping gestures 54
2.6 Frame structure of engagement actions 56
2.7 Basic metaphoric mapping of physical interaction to discourse 57
2.8 Metaphoric mapping of presentational gesture to discourse 57
2.9 Metaphoric mapping of referring gesture to discourse 58
2.10 Metaphoric mapping of removal gesture to discourse 58

3.1 Communication as Object manipulation metaphoric mappings 74
3.2 Discourse as a Journey metaphoric mappings 75
3.3 Discourse as a Physical space metaphoric mappings 76

5.1 Distribution of *anyway* senses 132
5.2 Distribution of shift types in data set 138
5.3 Shift triggers by shift type 143
5.4 Turn structure by shift type 148
5.5 Turn structure by shift trigger 149
5.6 Lexical discourse markers adjacent to "anyway" 152
5.7 Summary of predictions for gestural expression of topic-shifting 155
5.8 Gestural expression of topic dismissal and introduction with *anyway* 156
5.9 Temporal alignment of dismissal and introduction gestures with *anyway* 157
5.10 Types of topic-shifting gestures accompanying anyway 158
5.11 Gestural expression of topic dismissal *anyway* 160
5.12 Distribution of removal gesture by shift type 169
5.13 Distribution of removal gestures by shift trigger 170
5.14 Distribution of removal gestures by turn position 171
5.15 Distribution of non-manual disengagement by shift type 178
5.16 Distribution of non-manual disengagement by shift trigger 179
5.17 Distribution of non-manual disengagement by turn position 179
5.18 Gestural expression of topic introduction with *anyway* 181

5.19 Distribution presentation gestures by shift type 186
5.20 Rate of locating gestures by shift type 189

6.1 Use-contexts of BTW 209
6.2 Distribution of use-contexts with BTW 210
6.3 Clause position of BTW 216
6.4 Clause position of BTW 217
6.5 Lexical discourse markers accompanying BTW 219
6.6 Recurrent gestures aligned with BTW 220
6.7 Temporal alignment of gestural discourse markers with lexically-signaled digressions 221
6.8 Temporal alignment of interactive gesture with signaled digression 226
6.9 Gestures accompanying BTW by interaction type 228
6.10 Gestures accompanying BTW by discourse type 229
6.11 Forms and functions of gesture classes associated with BTW 230

7.1 Discourse relations underlying HTT 259
7.2 Discourse relations underlying HTT by discourse type 260
7.3 Co-occurring lexical discourse markers by relation 273
7.4 Reccurrent gestures accompanying HTT 278
7.5 Gestures accompanying HTT by discourse type 279
7.6 Gestures accompanying HTT by functional context 282
7.7 Gestures accompanying HTT by turn position 283
7.8 presentation gestures accompanying HTT and the proceeding specification . . . 285
7.9 Stopping gestures accompanying HTT by functional context 307

8.1 Gestural expression across discourse moves 317

Acknowledgments

As with any work of significant time and effort, this dissertation would not have been possible without the support of an incredible cast of kind, thoughtful, and generous people. I am thankful to more people than I can name.

An obvious and huge thank you goes to my committee, Dor Abrahamson, Line Mikkelsen, Elise Stickles, and Eve Sweetser, for their patience and support over the last two years. It has been a long and occasionally halting journey, but they have stuck with me, offering their wisdom, advice, and time throughout.

I owe so much to my co-advisors, Eve Sweetser and Line Mikkelsen. Without them, I am fairly confident that I wouldn't have made it to candidacy, never mind to the end of a dissertation. Eve, I aspire to your wisdom and generosity. You made me feel at home from the moment I arrived in Berkeley, and have guided me in my development as a teacher and cognitive linguist in every step since that arrival. I hope that I am as surefooted of a guide one day to a future advisee, and that I have even a small fraction of your wisdom to offer them. Line, I aspire to your open-mindedness and thoughtfulness. You made me feel valued and capable when I doubted myself most, and talked me through all kinds of intellectual, professional, and emotional quandaries. I hope that I am as articulate and responsive of a conversationalist one day, and that I can offer the people in my life a fraction of your warmth and conviction.

I am also very grateful for the continued support of Rob Truswell, my undergraduate advisor at the University of Edinburgh, who somehow finds the time for ongoing collaboration and encouragement.

I owe many thanks to Belén Flores and Paula Floro for always solving every problem, very quickly and with incredibly good humor. You are heroes of the Berkeley Linguistics Department, and I am so grateful to have overlapped with you. Thank you also to Johnny Morales Arellano and Susan Luong for holding the department together after they left, and helping me finish without any unexpected disasters. A special thank you to Martine Alexander for always having my back, especially when things got strange.

I would also like to thank all of the (former) undergraduate students who I have had the honor of teaching and working with during my years at Berkeley. Your creativity, curiosity, and insight never ceases to impress me. A very special thank you to all of my incredible undergraduate research assistants: Kahini Achrekar, Kiana Aryan, Miranda Cheung, Char Juin Chin, Annabel Davis, Kat Huynh, Pranav Jayachand, Toby Kafin, Anjali Kantharuban, Yi Xuan Kueh, Karsen Paul, Sarah Roberts, Sophia Stremel, and Irene Yi.

I am thankful to the Berkeley graduate student linguistic community for providing an open-minded, caring, and constructive space through my six years in the program. I am especially grateful for time spent and conversations had with Madeline Bossi, Emily Clem, Meg Cychosz, Emily Drummond, Edwin Ko, Tyler Lemon, Gabriella Licata, Ben Papadopoulos, Dakota Robinson, Tessa Scott, and Amalia Skilton. I owe further thanks to Virginia Dawson, for dive bar visits and philosophical musings; to Myriam Lapierre, for never being anything other than Myriam Lapierre; to Erik Hans Maier, for an abundance of warmth, absurdity,

and aimless walking; and to Martha Schwarz, for kindness and a constant willingness to daydream when we should have been working.

Finally there is, of course, my family, given and chosen, without whom this work would not exist. A special thank you to Victoria Dearborn, whose silliness and wisdom is unparalleled, and Oona Dooley, whose magic is felt across oceans. Also T Haddow, for doing their best to get me out of the house once in a while, and Kelly Jones, for always finding the metaphor. A further thank you to Ivy Armstrong, Leah Busby, Kai Yui Samuel Chan, Rachel Chen, Stephen Dockery, Rachel Forgash, Hamish Finlay, Inés Lozano Palacio, Sadie Meade, Tania Osorio, Natalya Tuabin, Ivaylo Youmerski and all of the others who I've embarrassingly forgotten to name. You have shown me so much kindness and brought me so much joy.

The debt I owe to my parents, Christine Aubrey and John Laparle, is unpayable. Despite not always knowing what I was doing, or why I would want to do it, their support was unwavering. A special thank you as well to Connie Capacchione, who is always there when you need her, and to Maragaret, Geoff, Pat and Bill.

Finally, I cannot express the love and gratitude I feel toward my husband, Jason Borga. He has seen me through the worst moments and has been with me for the best. He has, without complaint, kept me caffeinated and fed, more often than not with unnecessarily complex and delicious meals. He has driven me hours into the mountains and forests to help me find moments of peace. He has played dozens of games of Terraforming Mars, watched countless Star Trek episodes, and sat through innumerable hours of gesture-related rambling. All because he loves me, which appears to be his only flaw. Jason, you are my rock, my anchor, my sun, and all of the other cheesy metaphors for emotional support.

A very special thank you also to Shantih, our perfect cat, who always knows when to come and curl up on my lap.

Chapter 1

In pursuit of a multimodal linguistics

1.1 Introduction

Every day, we depend on language to interact with one another. We have conversations, write emails, solve problems, and complete tasks, often in a relatively reasoned and orderly way. That we do so with apparent ease is nothing short of incredible. Despite the distractions, despite the possibility for misunderstanding, despite the impossibility of translating the richness and multidimensionality of our subjective embodied experiences into a finite linear expression, we communicate. More often than not, we communicate quite well. A fundamental goal of linguistics as a discipline is to understand what makes this possible. The recognition of language as inherently multimodal is changing how we pursue that goal.

In the last two decades, this recognition of the multimodal nature of language has gained traction. Indeed, within the past ten years or so, we have seen a dramatic increase in the number and range of 'traditional' linguists interested in exploring the potential of a truly multimodal grammar, especially within cognitive (e.g. Cienki 2017; Kok & Cienki 2016; Schoonjans 2017; Stickles 2016; Zima & Bergs 2017; Zima 2017) and formal semantic (e.g. Alahverdzhieva et al. 2017; Lascarides & Stone 2009b; Schlenker 2020) approaches. Though the study of gesture and its relationship to verbal language is not new, its incorporation into mainstream linguistics is. So why now? Why is gesture finally making its nascent appearances in semantic conferences and syntactic journals? There are, no doubt, some simple reasons, such as better technology for multimodal data collection and better access to large multimodal data sets (such as UCLA's Communication Studies Archive which has made this dissertation possible). More important than these practical shifts, I believe, are three theoretical shifts: (i) the 'embodied turn' in the cognitive sciences, (ii) the acceptance of iconicity as a legitimate aspect of linguistic communication, as necessitated by sign language linguistics, and (iii) the development of descriptive and analytic frameworks for describing gesture on its own terms, rather than in its relationship to accompanying speech.

First is the so-called 'embodied turn' in the cognitive sciences (Mondada 2016; Nevile 2015), of which linguistics is a core member. This turn shifted cognitive science and its

constituent fields from considering cognition as an abstract, intangible system, to something that is directly observable and grounded in our experiences as physical agents in a physical world. The subfield of Cognitive Linguistics, which is dedicated to the study of language in its relation to cognition and interaction, has contributed directly to the embodied turn, perhaps most notably through work on metaphor and conceptual integration (e.g. Lakoff & Johnson 1980; Fauconnier & Turner 2008). Though politically and socially driven tensions between Cognitive and Generative linguistic frameworks have seemed insurmountably high for half a century (Harris 1993), this larger embodied turn is providing a new opportunity for conversation and co-development of new linguistic theories.

Second is the long overdue incorporation of sign languages into general linguistic sub-disciplines (i.e. phonology, morphology, semantics, and syntax), rather than maintaining it as a marginalized sub-discipline unto itself. This has been a long-fought and often fraught process (see Hockett 1978; Johnston & Schembri 2007; Reagan 2011; Stokoe 1960; Stokoe et al. 1976; Wilcox & Occhino 2017 for potent discussions). It appears that sign language linguistics has finally won the political battle – sign languages are now uncontroversially *viewed* as full languages and within the domain of linguistics proper. Though there are arbitrary form-meaning pairs in all sign languages, sign languages also take advantage of iconicity to an extent that is simply not possible for spoken languages (Hockett 1978). The resulting acceptance of iconicity as a legitimate source of linguistic meaning opens the door for reconsidering the legitimacy of gesture, which relies largely on iconicity and embodied metaphor, as contributing linguistic meaning in spoken languages.

Third, recent methodological and theoretical developments within gesture studies have made the incorporation of gesture studies into linguistics proper appear more relevant to the formal linguist. Early influential work on gesture focused disproportionately on the *differences* between speech and gesture (McNeill 1992). These differences are, without a doubt, both numerous and worth attention. However, recent work demonstrating the systematic contribution of gesture to face-to-face communication suggests that the focus on *differences* may be counter-productive (e.g. Harrison 2018; Kendon 2014; Perniss 2018; Streeck 2009 for discussion).[1] Equipped with the increasing knowledge that gesture and speech consistently work in concert during communication, we are now tasked with exploring how gesture contributes to communication *despite* its differences with speech, and how those differences are exploited to make our communicative system more efficient.[2]

In the remainder of this introductory chapter, I discuss how these three developments have contributed to the current work, and how I intend this work to further inform these

[1] Bavelas (1994) and Kendon (2000) were early advocates of this view. It is not until much more recently that it has been taken seriously by a more general audience within either gesture studies or linguistics.

[2] There is, at this point, substantial experimental work demonstrating the increased efficiency of multimodal communication. For example, gesture has been shown to increase spoken fluency when providing spatial information (e.g. Rauscher et al. 1996), increase learning outcomes in child-directed and second-language education (e.g. Church et al. 2004; Goldin-Meadow 2017; Singer & Goldin-Meadow 2005), decrease processing time and surprisal (e.g. Drijvers & Holler 2022; Holler et al. 2018; Holler & Levinson 2019), and aid in completing novel tasks (e.g. Holler et al. 2011; Macuch Silva et al. 2020).

three developments. I begin by outlining my intended contributions and the strategies I use (Section 1.2). I then turn to situating this dissertation in relation to three "big questions":

1. What is gesture and what is it for? (Section 1.3)
2. Is gesture a part of language and how can we tell? (Section 1.4)
3. How might the consideration of gesture as a part of language impact our existing linguistic theories? (Section 1.5)

These discussions are necessarily abbreviated, and I do not intend to offer full answers. However, I do think that taking the time to relate *any* substantial work, such as a dissertation, to these larger issues helps to clarify the work's worth and purpose.

1.2 Contributions of this work

In this dissertation I hope to contribute to our understanding of gesture meaning while also making efforts to bridge the gap between our understanding of gesture and our existing linguistic theories. My focus is on *discourse structure* and how the class of *interactive gestures* contributes to that structure. I have chosen this focus not just because it happens to be what interests me most, but also because these two areas remain *messy*. In the study of discourse structure, for example, the definitions of basic concepts such as *topic* and *comment* are still subject to debate (e.g. Krifka 2008). As is how to describe relationships between utterances (e.g. Hovy 1990; Rohde et al. 2018) and how those relationships may or may not be overtly expressed (e.g. Taboada 2006, 2009). To be clear, I consider this messiness, and the willingness to acknowledge it, a virtue. In tidiness, there seems to me a certain reticence toward novelty and major change. There is too much at risk to question fundamental premises. In *messiness* there is an open-mindedness, even an enthusiasm, to embrace new theories and approaches, to question premises. The incorporation of gesture studies into linguistics proper depends on such open-mindedness and a willingness to change. In the pursuit of a truly multimodal linguistics, the interface between gesture and discourse structure thus seems a particularly productive place to start.

In order to understand the interface between gesture and discourse structure, we have to have a working theory of how both work independently. *How* gestures convey meaning is still a subject of significant debate, one that I'll briefly overview in Section 1.3. In Chapter 2, I describe and commit to the 'action schema' approach to gesture meaning which holds that gestures convey meaning metonymically and metaphorically through the physical actions they enact or depict (Müller 2017). The nature of discourse structure, how it is organized and why it is organized as it is, is also still debated. Chapter 3 is dedicated to outlining the hierarchical, question-based, and goal-oriented approach that I use in this work (Grosz & Sidner 1986; Roberts 1996, 2012; van Kuppevelt 1995). In Chapter 4, I apply this understanding of gesture meaning and discourse structure to develop a formal model of interactive

meaning in gesture. I introduce the *Interaction Space* (Laparle 2022) to do so – the physical space in which a discourse is maintained and constructed by co-present interlocutors. I show how a relatively small inventory of 'management actions' can represent the ways in which gesture acts upon discourse structure, systematically and predictably. This step of formalization is necessary if we are to assess the extent to which our theories of gesture meaning and discourse structure are compatible. With theories of discourse structure and interactive meaning in gesture established, Chapters 5-7 test their compatibility, focusing on the ways in which interactive gesture can improve our understanding and modelling of discourse structure.

The proposal underlying each part of this work is that discourse management, like language more generally, is fundamentally multimodal. By this I mean that language users can employ all strategies at their disposal, in spoken, signed, and gestural modes, to ensure that conversation progresses as smoothly as possible. I propose the following four *Principles of Multimodal Discourse Management* to help guide our discussions of how gesture and speech work in concert to convey discourse-structural meaning.

(1) **Principles of Multimodal Discourse Management**

a. Multiple strategies: There are multiple strategies, within both the verbal and gestural modes, for expressing discourse management.

b. Optionality of expression: The use of each expressive strategy is optional and subject to contextual variation.

c. Independence of contribution: Strategies may be employed independently, and simultaneously employed strategies may profile different aspects of discourse management.

d. Compositional management: The strategies employed in both modes are integrated systematically and predictably into a single coherent multimodal message.

The principle of *multiple strategies* states that there are potentially many ways to express a single type of discourse management. For example, consider how you might present information about a sequence of events. All of the verbal variants in (2) seem to be acceptable. In each case, we are variably relying on order of presentation and lexical markers (*before, after, then,* etc.) to convey that these events happened in a particular order.

(2) a. I made coffee, ate breakfast, and went to the store.

b. I made coffee and ate breakfast before I went to the store.

c. I went to the store after I made coffee and ate breakfast.

d. Before I went to the store, I made coffee and ate breakfast.

In addition to these verbal strategies, gestural strategies may be employed, such as counting the events on your fingers, or tapping from left to right on the table as if indicating each event on a timeline in front of you.

The principle of *optionality of expression* states that each strategy in each mode is optional. This does not rule out the possibility of 'optimal' sets of strategies. Instead, this

principle suggests that there is no particular set of strategies that is *obligatory*. This means that we expect strategies to vary both systematically and idiosyncratically. *Who* is doing *what* and *when* directly impacts which strategies are employed, and which are not.

The principle of *independence of contribution* maintains that the meaning of concurrent strategies are non-redundant, and as such, each strategy profiles a different aspect of the given communicative context. Consider (2d) above. In this example, the lexical items *before* and, potentially, *and*, are providing information about the sequence of events. The preposing of the chronologically last event to the front of the utterance conveys something about the topic of conversation – something like *what did you do this morning before going out?*. Similarly, a concurrent gesture may convey information about the chronology of events, the topic of conversation, or some other contextually salient information. For example, perhaps I'm telling you about my day, and I forgot to include *making coffee* and *eating breakfast*. In this context, I may hold up a finger, as if to say "wait a moment" while I correct my list of events by saying the sentence in (2d).

Finally, the principle of *compositional management* simply states that each expressive strategy is interpreted as a part of a single coherent message, as in Enfield's (2009) conception of the 'composite utterance'.[3] This principle is particularly important when considering the apparent polysemy of particular gestures. Just as compositionality in the verbal mode gives rise to apparent polysemy and multifuntionality of its constituents, so too does multimodal compositionality give rise to apparent polysemy and vagueness in the gestural mode. This will be discussed in detail in Section 1.4, and will remain a primary focus throughout this work.

Based on these four proposed principles of multimodal discourse management, I ask two empirical questions: (i) are there expressive strategies that tend to group together, and (ii) what contextual factors contribute to a preference for one expressive strategy over another? To address these two questions, I focus on three lexical discourse markers, *anyway, by the way*, and *here's the thing*, and the interactive gestures that accompany them. I employ both quantitative and qualitative approaches. First, I identify general patterns in gesture-lexical discourse marker alignments in 100-150 examples of each discourse marker, as they occur on the television talk show *The Late Show with Stephen Colbert*. I then conduct close qualitative analyses on particular examples, showing how other discourse structural factors can account for variation in discourse management strategies.

By demonstrating the systematicity with which gesture provides discourse structural information, I aim to contribute to the transition from a speech-centric to a truly multimodal linguistics. In particular, I intend this work to demonstrate the value of gesture studies to the formal analysis of discourse structure.

[3]This principle also helps to distinguish discourse management gestures from so-called 'body language', movements which are claimed to subconsciously express aspects of the speaker's psychological state rather than the intended communicative message (e.g. Beattie 2004).

1.3 What is gesture and what does it do?

Before we can fully explore the multimodal expression of discourse management, we have to establish a working understanding of what gesture is and how it relates to verbal communication. This section provides a brief overview of the role of "gesture" in communication as it is understood in established literature.

I follow Kendon's broad definition of gesture as any "visible action when it is used as an utterance or as part of an utterance" (2004:7). This means that any meaningful movement of the hands, head, body, or face can be considered gesture. This definition, of course, includes the behaviors that likely come to mind when one hears the word "gesture" – index finger points, thumbs up, and the elaborate hand movements that so often co-occur in animated face-to-face interactions. Also included in this definition are facial expressions, such as smiles and eyebrow raises, as well as relatively subtle body movements such as when an interlocutor leans toward their addressee as they recount the most dramatic part of a story, or leans back as they finish their turn as speaker. This definition also takes an important step toward integrating gesture into linguistic study on its own terms, rather than in its relation to speech. In taking this step, I reject the term 'co-speech gesture', despite its predominance in current gesture literature, in favor of simply 'gesture'. Non-verbal communicative actions are considered gesture, independent of their alignment (or not) with speech. A 'thumbs up' to signal agreement is a gesture, whether it is used with an accompanying verbal affirmation (e.g. "good" or "yup") or in silence. Similarly, leaning toward one's interlocutor signals engagement and interest in the current discourse, regardless of whether it is performed by the speaker or listener. No distinction is made between gestures occurring with and without speech. Apparent differences in functionality of gestures with and without speech arise from multimodal compositionality, rather than differences in the gestures themselves.

It is worth noting that Kendon's definition does not clearly differentiate between communicative *effect* and communicative *intent*.[4] In order to assess whether a visible action is "used as an utterance or part of an utterance", we have to have a clear notion as to what constitutes an 'utterance'. Kendon (2004:7), following Goffman (1963), goes on to define an utterance as "any unit of activity that is treated by those co-present as a communicative 'move', 'turn' or contribution". Rather than serving as a clarification, this offered definition opens more questions including (i) what should count as a "'move', 'turn' or contribution", and (ii) how to determine whether something is "treated" as a contribution by those co-present. In an attempt to resolve this, I offer the following condition: if a visible movement can be interpreted as contributing to a single, coherent, communicative message, then it can be considered gesture. This successfully excludes involuntary movements (e.g. sneezing, itching) and "body language" from the definition of gesture. Though these movements may

[4]The question of whether or not the use of gesture is "intentional" has resulted in an extended debate as to whether gesture is "for the hearer" (intentional, communicative) or "for the speaker" (unintentional, cognitive). At this point, it seems safe to say that the answer is *both*, and that we do not have sufficient methodologies for measuring speaker "intent". See Church et al. (2017) for an array of approaches considering this question. Also see discussion in Cooperrider (2017) for a reconciliation of the two views.

convey information (e.g. that someone is sick or anxious), they cannot be integrated into a coherent communicative message within an ongoing interaction.

Despite this relatively broad definition of gesture, the core arguments for the inclusion of gesture in linguistic research is its consistent alignment with speech at all levels of linguistic representation, from physical articulation to meaning (e.g. Abner et al. 2015; McNeill 1992, 2005; Kendon 2004; Wagner et al. 2014). I will outline some of the most relevant ways in which the alignment between speech and gesture has been demonstrated. Because the purpose of the present work is to argue for the inclusion of gesture in linguistics research on its own terms, I will focus on ways in which gesture-speech alignments can reflect the capacity of gesture to contribute independent meaning to a single coherent multimodal message.

1.3.1 Gesture in verbal communication

If one performs a hand gesture while saying "she ran up the hill", it will likely involve movement in an upward trajectory.[5] Part of the upward movement will occur as the word "up" is uttered. The word "up" can be considered the gesture's *lexical affiliate* (McNeill 2005; Schegloff 1984), the phrase most closely related in meaning.[6] Moreover, the gesture will certainly *not* involve movement in a downward trajectory. This consistent alignment in meaning and timing has been used to argue that "gesture and speech share a computational stage" (McNeill 1985:353), that is that they are both part of the cognitive system that gives rise to communication. More recently, gesture researchers have been interested in looking beyond simple alignment to understand the ways in which gestures *add* information to an utterance. This is an important step in establishing gesture as a part of communication proper, rather than only as a reflection of underlying cognitive processes. In service to this goal, I will briefly describe three ways in which gestures communicate information independent from, but in concert with, accompanying speech by (i) providing non-redundant semantic information, (ii) performing discourse management, and (iii) signalling listener feedback.

1.3.1.1 Non-redundant semantic information

First, consider again the upward gesture you imagined aligning with "she ran up the hill". It was likely a fairly simple gesture, following a more or less straight trajectory, such as that in (A) or (B) in Figure 1.1 below. The movement represented by (A) and (B) can be considered "fully redundant" with the accompanying speech; no new information would be provided by such gestures. The movement represented by (A) would be fully redundant with "up", and (B) would be redundant with basic features of the full proposition, namely that

[5]Note that this would not necessarily hold for native speakers of verb-framed languages (e.g. Turkish) in which the trajectory of an event is preferentially coded in the verb. See Kita & Özyürek (2003) and Özyürek & Kita (1999) for discussion.

[6]Gestures frequently begin prior to their lexical affiliate (e.g. Schegloff 1984:275). Previous work has shown that when a gesture occurs prior to its lexical affiliate, it is often 'held' until its affiliate is completed (e.g. Harrison 2010, 2018; McNeill 2005).

there was an event involving both vertical and lateral movement. However, the trajectory of this imagined gesture can be modified such that the speed, angle, and curvature of movement provides additional non-redundant information. For example trajectory (C) below may convey information about the steepness of the hill, and (D) may convey information about the path the runner took (zig-zag rather than straight).

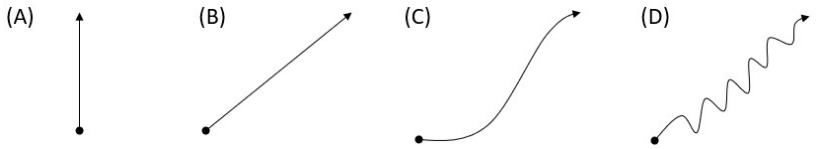


Figure 1.1: Movement trajectories compatible with the lexical affiliate "up"

This type of variation in gesture performance is considered partially redundant with the concurrent speech. The meaning of the gesture is compatible with the meaning of the spoken utterance. Indeed, the gesture often cannot be fully interpreted without considering the meaning conveyed in accompanying speech. However, the gesture also provides additional truth-conditional information that is otherwise not overtly expressed. Kendon (2014:9) provides a particularly good example of this in natural conversation with gestures accompanying the verb "throw". Consider the two throwing events described in (3). In each case, a gesture depicting "throwing" is performed, aligned with the underlined segments of the utterance.

(3) a. He used to go down there and throw ground rice over it
b. an' throw oranges and chewing gum all off the lorries to us kids in the streets
(Kendon 2014:9, Figures 6 & 7; formatting simplified)

In (3a), the throwing gesture is performed at waist level and involves repeated small lateral movements. The gesturer's palm is facing up, and his hand is cupped, as if holding grains of rice. In (3b), the throwing gesture involves a large repeated movement from the gesturer's waist and up to head level, as if to grab small objects from his lap and lob them backward over his shoulder. Here we can see complementary information about the event conveyed across the two modes. What is being thrown is expressed in the verbal mode by the phrases "ground rice" and "oranges and chewing gum". Where it is being thrown is expressed in the verbal mode by the phrases "go down there", "over it", "off the lorries" and "to us kids in the streets". The manner of movement is expressed in the gesture through position, trajectory, and repetition.

In cases such as these, we can infer based on real-world experience that different movements are involved, even when expressed by the same lexical item. The point is that in many cases this information is implicit if only the speech is considered. This implicit information can then be overtly expressed through manual gestures such as those described above.[7]

[7]See also Beattie & Shovelton (1999), Holler & Beattie (2003), Hostetter & Alibali (2011), Melinger & Levelt (2004), and Rowbotham et al. (2012) for discussion of independent semantic meaning gained from gesture.

1.3.1.2 Discourse management

Gestures are also used to convey information about discourse structure and interactional context, helping to "maintain the conversation as a social system" (Bavelas et al. 1992: 469). These *interactive gestures*[8] often convey information that is unexpressed or underspecified in the verbal mode, including intention to begin or end speaking (e.g. Brône et al. 2017; Jokinen et al. 2013; Streeck & Hartge 1992), information status (e.g. Enfield et al. 2007; Holler 2009; Holler et al. 2011; Holler & Bavelas 2017; McNeill et al. 1993), the relationship between adjacent utterances (e.g. Hinnell 2019; Jannedy & Mendoza-Denton 2005; Laparle 2021; Müller 2004), viewpoint (e.g. Parrill 2009, 2010, 2012), and speaker attitude (e.g. Bressem & Müller 2014, 2017; Calbris 2008; Teßendorf 2014; Wehling 2017). All of these functions can be expressed in the verbal mode through particular lexical markers, syntactic structures, and prosodic cues. However, manual gestures are an efficient way to convey this information without necessarily taking the time in speech to do so.

Consider the role of gesture in managing turn-structure. To signal a desire to begin speaking, a discourse participant may lean in and raise their hand toward their interlocutor (Bohle 2014; Streeck & Hartge 1992). Head nods have been observed to perform a similar function (Dittmann & Llewellyn 1968). This allows the participant to signal their intention without interrupting the current speaker or having to ask overtly for a turn to speak. Moments of 'mutual gaze', when the eyes of interlocutor's meet, have also been shown to align with turn-transitions such that the current speaker will look at their addressee as they are readying to end their turn (Kendon 1967; Jokinen et al. 2009, 2013). All of these gestural strategies seem to facilitate smooth transitions between speakers without one having to always overtly say "what do you think" or "what do you have to say".

1.3.1.3 Feedback

The efficiency of using gesture to convey independent information without disrupting the verbal mode is particularly apparent in backchannels, signals used to 'check-in' on how the discourse is going (e.g. Duncan 1974; Goodwin 1981; Heylen et al. 2011). Frequent gestural backchannels include nodding (McClave 2000; Stivers 2008) and smiling (Bavelas & Chovil 2018; Brunner 1979), which can be used to signal agreement, understanding, and continued attention. These may or may not be accompanied in the verbal mode by small vocalizations like *mhm* and confirmations such as *yeah* or *right*. Requests for confirmation of understanding frequently include nodding (McClave 2000), presentational or 'delivery' hand gestures (Bavelas et al. 1992) and establishing momentary mutual gaze (Bavelas et al. 2002). Similarly, these gestural check-ins may or may not be accompanied in the verbal mode by

[8]There is significant variation in functional class terminology in the gesture literature. For example, 'interactive gestures' are also referred to as 'pragmatic gestures' (e.g. Kendon 1995, 2017) and 'discourse management gestures' (e.g. Wehling 2017). I favor the term 'interactive gesture' in this work for its transparent connection to 'the Interaction Space' model proposed in Chapter 4.

phrases like *y'know*. Finally, Holler & Wilkin (2011) also find evidence that gesture mimicry play an active role in confirming common ground between experiment participants.

1.3.2 How gestures convey meaning

There is an abundance of work demonstrating the ways in which gesture aligns with speech and seems to contribute directly to communication. However, that these alignments exist does not tell us anything about the nature of gesture as a semiotic system. Since a goal of this work is to justify the inclusion of gesture into general linguistic study, it is not enough to know *that* gestures convey meaning. We also have to know *how* gestures convey meaning. Moreover, we have to clarify what exactly we "mean by meaning" (Parrill & Sweetser 2004). To do this, it is necessary to distinguish models of *gestural meaning* from models of *multimodal meaning integration*. The first type of model asks simply "what does this gesture mean?". The second asks, in the words of Bavelas (1994), "what does this gesture do and how?". In attempts to formalize gesture's contribution in verbal communication, the first question, "what does this gesture mean?", often remains unarticulated (e.g. Alahverdzhieva et al. 2017; Lascarides & Stone 2009a; Schlenker 2020). This results in an implicit understanding of gesture meaning *as* its integrated meaning with accompanying speech. If gesture is to be integrated into linguistic theories as a full semiotic system, it is necessary to foreground a model of gesture meaning before considering a model of integration.

By and large, gestures convey complex meaning through *iconicity*[9] – some aspect of their form resembles some aspect of their meaning.[10] Consider once more the upward gesture you imagined aligning with "she ran up the hill". We can say that this gesture means "upward movement" by resembling the upward movement of the event being described. This is an *iconic* representation rather than a purely literal representation because we do not interpret the gesture as meaning "upward movement of the gesturer's hand". Instead we are mapping the upward trajectory of the gesture to a very different kind of upward movement (running up a hill). We could also imagine a situation in which the same gesture is used while saying "and I move my hand upward like this". In this case the 'same' gesture may be considered a full enactment of the meaning, rather than an iconic representation.

This type of ambiguity is what leads to the conflation between gesture meaning and multimodal meaning. We just can't seem to fully interpret the meaning of the gesture without interpreting the meaning of the accompanying speech. Indeed, this apparent inability for a gesture to *mean* without accompanying speech has been upheld as evidence that gesture is not a part of our language system proper. I will return to this criticism in the following

[9]In my approach, the meaning of the gesture is always *iconic*. When a gesture is used to convey interactive meaning, its iconicity is understood as metaphoric. The process of deriving contextualized metaphoric meaning from a gesture's iconic representation will be discussed in Chapter 2.

[10]There are, of course, exceptions. 'Emblems' for example, are a class of gestures that convey meaning through arbitrary convention, rather than iconicity (Ekman & Friesen 1969; McNeill 1992). These include gestures like 'thumbs up' to mean "okay" or 'the middle finger' to mean an obscenity.

section. For now, let's consider how we might move forward in developing an understanding of independent gestural meaning.

I argue that the overwhelming focus on semantically-oriented gestures, such as the upward gesture we've been considering, has negatively impacted efforts to develop an independent model of gesture meaning. When analyzing semantic gestures, researchers are working, often implicitly, with the hypothesis that the meaning conveyed by the gesture is derived from the meaning conveyed in the accompanying speech. The researcher takes the meaning conveyed in speech to be equivalent to the meaning conveyed by the message as a whole, and works backward to interpret the meaning of gesture. This creates a bias – the gesture will be analyzed as conveying some portion of the meaning conveyed in speech and nothing else. This, in turn, results in a view of gesture as semiotically impoverished, unable to convey meaning independently and on its own terms. In order to prevent this bias in analysis, we need to develop ways of interpreting gesture meaning without assuming that it is the same meaning conveyed in speech. One way to do that is to focus on gestures that lack clear lexical affiliates, gestures that cannot be interpreted via the semantic information conveyed in speech.

Conveniently, the class of interactive gestures (Bavelas et al. 1992), gestures that convey primarily pragmatic meaning rather than semantic meaning, are often non-redundant with accompanying speech (Bavelas 1994:216-218). Interactive gestures more often express meaning that must be *inferred* from the discourse structural and social context. When interactive gestures *are* partially redundant with accompanying speech, it is most often with lexical discourse markers, verbal phrases whose meanings are notoriously difficult to pin down and agree upon (e.g. Fraser 1999; Jucker & Ziv 1998; Schourup 1999). This has resulted in the perception of interactive gestures as a particularly 'unruly bunch' (Streeck 2009:181) – if you are beginning with the hypothesis that you can deduce the meaning of the gesture from the meaning of the speech, as so many have, then interactive gestures are going to seem intractably random and non-systematic. It is exactly this seeming intractability that makes interactive gesture a particularly good place to start in pursuing the first question we have to ask – *what does this gesture mean?*.

In the present work, I use image schemas and action schemas as the basis for gestural meaning (e.g. Cienki 2005, 2013; Mittelberg 2018, 2019; Müller 2017). This approach holds that gestures mean by depicting some image or enacting some action. For example, we can say that the upward gesture we have been imagining for the sentence "she ran up the hill" depicts upward movement by tracing the path of something actually moving upward. Most of the gestures discussed in this work mean through action schemas – people are acting upon the discourse's structure by enacting the presentation, removal, and organization of metaphoric objects. Importantly, we can identify a gesture's action schema by looking at properties of its physical form (e.g. Calbris 2013). This methodology will be exemplified and described in detail in Chapter 2.

1.4 Is gesture a part of language?

This brings us to our second "big question" – *is gesture a part of language and how can we tell?* Even if we can show that gestures *mean* independently from accompanying speech, how can we tell that it is the type of meaning that we associate with *language*? Underlying this question is an even larger one, a question that underlies all of linguistic study – *what is language?* The answer to this question determines what belongs in the discipline of linguistics, and what doesn't. It also has implications for what methodologies we use, what generalizations we make (intentionally or otherwise), and how much attention we as researchers pay to different aspects of communication. If 'language' is defined such that gesture is a part of it, than our theories and models of 'language' must accommodate multimodality. If language is defined such that gesture is *not* a part of it, than the study of gesture can be more or less ignored in our linguistic pursuits – our theories do not need to account for it.

In much of this work, I will *assume* that gesture is a part of language, and treat it as such. In order to justify this, I will use the remainder of this chapter to argue that we do not have sufficient methodologies for distinguishing, in a principled and unbiased way, 'language' from communicative systems more generally. I will do so by presenting 'language' as a *contested concept*, and demonstrating the ways in which feature-based approaches to defining language are insufficient given this contestation.

1.4.1 Language as a contested concept

That we can ask the the question "what is language?" at all suggests that language is a *contested concept* (Gallie 1955). By this I mean that there is no objective, in-the-world way to answer the question. The very concept of "language" is a construct of thought and reflection, subject to change in the same way our concepts of *art* (Gallie 1956), *democracy* (Collier & Levitsky 1997), and even *science* (Jasanoff 1987) are subject to change across social contexts and time.[11] If we accept *language* as an essentially contested concept, then we are charged with explicating its use in our work. We cannot take the definition of 'language' for granted, just as we don't take for granted theory-specific terms such as 'topic' and 'comment' (Krifka 2008).

Kendon (2014) argues that there is no meaningful study of 'language' without acknowledging its function in 'languaging', i.e. the use of language to communicate.[12] In gesture studies this claim was a radical position until quite recently.[13] In linguistics, it remains a radical position. Accepting this position in linguistics would require us to question fundamental principles, such as the *competence-performance* distinction (Chomsky 1965) which maintains language as an abstract system distinct from its embodied use. I argue that

[11] See Harris (1980, 1981, 2002) for extensive discussion and critique of 'language' as an academic construct.

[12] This perspective is also convincingly argued for in, for example, Bavelas (1994), Cienki (2022), Ferrara & Hodge (2018), Kendon (2000), and Perniss (2018).

[13] See, for example, Parrill (2008:195-201) for discussion of most gesture being uncontroversially 'non-linguistic', compared to the more recent critiques of this position in Ferrara & Hodge (2018).

achieving a multimodal linguistics will require an even more radical position, in line with that advocated for in Ginzburg & Poesio (2016); 'grammar', the concept that has for so long maintained the understanding of language as an abstract system, is indistinguishable from the rules that govern language use in interaction.[14] This position does not necessitate the abandonment of models of language that represent linguistic phenomena through abstraction. In fact, recognizing our models as abstractions rather than direct representations of reality is both liberatory and productive. In an approach to language as a fully embodied and interactional system, abstract models are used to *direct* us toward answers in the real world, rather than constitute the answers themselves. Imperfections are embraced for their usefulness in identifying the unexpected, rather than rejected as 'breaking' the overall model.

In this section, I explore and critique the 'features of language' that have been used explicitly to argue *against* the inclusion of gesture in language.[15] I discuss the ways in which a feature-based approach to defining language is (i) illegitimate given its contested status, and (ii) detrimental to the understanding of both verbal and gestural meaning-making processes. First I focus on four diagnostics for 'language-like' properties, as described in McNeill (2005:48-49) and summarized in (4).[16]

(4) **'Language-like' properties** (as described in McNeill 2005:48-49)
   a. *Obligatory form*: features of a linguistic unit cannot be altered and still convey the same meaning (Section 1.4.2)
   b. *Prespecified meaning*: a linguistic unit has a single meaning conventionalized and entrenched in the grammar of the language (Section 1.4.3)
   c. *Arbitrary form*: the form-meaning mapping of a linguistic unit is one of learned convention rather than iconicity (Section 1.4.4)
   d. *Cultural specificity*: the form-meaning mapping of a linguistic unit is specific to a community, and is not universally interpretable (Section 1.4.5)

I will discuss the ways in which gestures *succeed* in meeting each of these requirements, despite common perception. I also demonstrate the ways in which speech very often *fails* to meet each of these requirements. In addition to these four diagnostics I also discuss notions of *discrete meaning* and *compositionality*, as they have been (mis)understood in traditional gesture literature (1.4.6). Through this discussion I hope to show that using criteria to determine a semiotic system's status as 'language' is counterproductive to understanding the nature of language more generally.

[14]This position is widely accepted in approaches to the study of language that are generally considered outside of the domain of 'mainstream' linguistics. Most notably, this includes embodied conversation analysis (e.g. Deppermann 2013; Goodwin 2000; Mondada 2016; Streeck et al. 2011) and embodied multimodal construction grammars (e.g. Bergen & Chang 2013; Oakley 2017; Schoonjans 2017; Steen & Turner 2013).

[15]In this framing of gestures as 'non-linguistic', emblematic gestures (e.g. the 'thumbs up' gesture) are generally treated as exceptional in their language-like properties (Ekman & Friesen 1969; Kendon 2004; McNeill 2000, 2005; Parrill 2008).

[16]These diagnostics are likely derived from Hockett & Hockett's (1960) proposed design features of communication.

1.4.2 Obligatory form

The first diagnostic requires that linguistic units have a set form, and that any change in form results either in ungrammaticality or the realization of a different linguistic unit. For example, the concept MAP (a visual representation of a spatial region) is realized in English as "map". The linguistic unit "map" cannot be altered without changing its meaning. If we change the place of articulation of the first sound from /m/ to /n/, we get a distinct linguistic unit, "nap", which maps to a distinct concept NAP (a short period of sleep, usually outside of standard sleeping hours). If we change the first sound /m/ to a /b/ instead, we get ungrammaticality; "bap" is not a conventionalized linguistic unit in English and has no form-meaning mapping.

After providing an example of the supposed failure of gesture to meet this diagnostic, I will argue that this diagnostic is incoherent on four grounds: (i) the claim that gestures do not have obligatory forms presupposes that we know what counts as a single gesture, which is unlikely given our lack of a comprehensive model of gesture meaning; (ii) it is difficult to separate the concept of obligatory form from a requirement for a given meaning to always map to one form, which is undermined by phenomena like synonymy and syntactic alternations in speech; (iii) given that variation is a natural part of all language production[17], it is unclear how to determine what type of variation is permissible while still maintaining a claim to 'obligatory form', and (iv) if we permit schematic unpronounceable units, such as syntactic constructions, to count as linguistic units,[18] then a concept of 'obligatory form' becomes practically undefineable.

To demonstrate the extent of formal variation in gesture, let us consider again the 'upward movement' gesture that we have imagined elsewhere in this chapter. A gesture conveying the notion of 'upward movement' can vary radically in regards to the position of the gesture relative to the speaker's body, the size of the gesture, the speed of movement, and the hand shape used, all while maintaining the core meaning. Under such an analysis, tracing an upward line with one's index finger while keeping the hand at rest on a table, and moving an open hand across one's entire body in an upward trajectory could be considered, in some way, the 'same gesture'. To illustrate, consider the three gestures in Figure 1.2, all of which occur with the lexical affiliate "go up". Despite occurring with the same phrase, they are articulated in very different ways.

[17]The prevalence of formal variation in language production is reflected in the existence of variationist approaches to linguistics. See, for example, reviews in Chambers & Schilling (2018) and Tagliamonte (2011).

[18]This position is advocated for most clearly in Construction Grammar approaches. See, for example, Fried & Östman (2004), Goldberg (2006), and Hoffmann & Trousdale (2013) for review.

Figure 1.2: Three gestures with "go up"

Though the narrow lexical affiliates for these three gestures are the same, the broader verbal contexts are different, as shown in (5). First, the contexts differ in concreteness. Gestures 1 & 2 correspond to metaphoric upward movement rather than literal upward movement. Gesture 3, on the other hand, depicts a very particular type of literal upward movement, namely a plane taking off. Gesture 2, unlike Gesture 3, is in the immediate presence of a contrasting movement ("go down"), which is represented in the immediately following 'downward movement' gesture (not depicted).

(5) Verbal utterances aligned with Gestures 1-3 in Figure 1.2
G1: "but will their taxes go up?"[19]
G2: "you know, ratings go up and ratings go down"[20]
G3: "did you go up and fly with the jets or anything like that"?[21]

Given these contextual differences, it may be argued that the formal differences in the three 'upward movement' gestures are functional, rather than idiosyncratic. As a result, analyzing these three gestures as the 'same gesture' would not make much sense. The three gestures certainly share a formal feature, *upward movement*, but this does not mean they are underlyingly the same gesture, just as a shared morpheme does not mean two words are underlyingly the same word. Indeed, a claim that "unconditional", "untie", and "unsubscribe" are the same words because they all contain the morpheme "un-" would be poorly received.

One may refute the legitimacy of this example by conceding that these are indeed different gestures, while maintaining that there *are* 'single' gestures that demonstrate radical idiosyncratic formal variation across articulations. In response, I would challenge one to demonstrate that the formal variation observed is, in fact, not functional. When gesture analysts look closely at variation in a 'single' gesture, they recurrently find alignments between form and function. For example, connections between formal variants and functional variants have been demonstrated for several cross-linguistically common gesture routines, including 'cyclic' gestures (Ladewig 2011, 2014a), finger points (Enfield et al. 2007), and shrugs (Debras 2017). The position and size of gesture variants in each of these cases has been shown to map to particular functional variants, suggesting that there is more than one underlying gesture routine.

[19]UCLA Communication Studies Archive UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335,1954
[20]UCLA Communication Studies Archive UID:a46f90b2-d40e-11e7-9367-089e01ba0335,2141
[21]UCLA Communication Studies Archive UID:a3441a4a-db2b-11ec-84a9-089e01ba0770,1892

It seems to me that it is this failure to differentiate *functional* and *idiosyncratic* variation that has led to the perception of gestures as not having 'obligatory forms'. This failure is related to the insufficiency of our models of gesture meaning and is reinforced by falsely equating gesture *families* (e.g. Fricke et al. 2014) with single underlying gestures. The study of gesture families, i.e. gestural routines that share some kinesic and functional core, is an important step in developing a model of how gestures mean. As the name suggests, these studies do not intend to argue that the gesture variants that constitute a proposed gesture family are underlyingly the same gesture. For example, Bressem & Müller's work (2014; 2017) on the 'away' gesture family demonstrates the ways in which a single shared feature (movement away from the speaker's body) combines with other formal features to create functionally and formally distinct gestures, such as 'holding away', 'throwing away', and 'sweeping away' gestures.

Problems arise when a gesture family is identified and named before all possible *members* of the family have been identified and named. This can result in a very general gestural behavior being mistaken for a single underlying gesture, when in reality the behavior constitutes an entire formally and functionally diverse gesture family. Perhaps the paradigm example of this is the palm-up open-hand (PUOH) gesture family which has been noted for its pervasiveness within and across languages, as well as its apparent functional diversity (e.g. Cooperrider et al. 2018; Müller 2004). Because PUOH gestures are so common and so functionally diverse, analysts have yet to fully understand how the titular palm orientation and hand shape combine with other formal features to constitute different gestures.[22] Without understanding this, it is impossible to judge whether a particular PUOH gesture exhibits idiosyncratic variation or functional variation. This in turn means that there is no way to assess whether or not a particular PUOH gesture adheres to or violates some notion of 'obligatory form'.

I've so far argued that we cannot claim that gestures do not have obligatory forms because we do not yet have a sufficient model of what should constitute a single gesture – until we fully understand how gestures mean, we cannot know whether or not variations in formal features are idiosyncratic and spontaneous or functional. I'd now like to turn briefly to the absence of a coherent notion of 'obligatory form' in speech. Variation in form occurs at all levels of linguistic representation, and yet we do not take this variation to be a threat to the status of speech forms as linguistic units.

The "same" sound (i.e. phoneme) is pronounced differently at different times and by different people. Indeed, entire theoretic models have been constructed explicitly to contend with the mapping of rampant articulatory variation to a single abstract linguistic unit. For example, usage-based approaches to phonology hold that individual experiences of the same phoneme can vary quite dramatically in their phonetic realizations – different lexical contexts, different physiological features of the speaker, and different social and environmental factors can all result in measurably different articulations (e.g. Bybee 2000; Hayes 1999; Pierrehumbert 2001). Still, theorists accept this variation as permissible deviations from a

[22]Though see Ladewig (2011; 2014a) for discussion of PUOH combining with cyclic movement.

some abstracted linguistic form. Phonetic variation is a particularly apt comparison because of its continuous and gradient nature. The possible positions of the tongue or the exact degree of rounding of the lips are continuous rather than discreet, just as the position and shape of the hands are. We accept such variation in phonology because we have full working models of phonological representations. We do not yet have full working models of gestural representations.

Grammaticality in speech has also been shown to be gradient, subject to inter-speaker variation, and context-dependent (e.g. Sorace & Keller 2005). For example, in previous experimental work, I demonstrated that the acceptance of the non-canonical syntactic construction known as *Locative Inversion* (e.g. "Into the room ran a bear" rather than "A bear ran into the room") is variably accepted by speakers based on other contextual factors such as verb type and grammatical aspect (Laparle 2020). This suggests that formal variation in a 'single' syntactic construction does not directly determine the unit's status as grammatical or not. As with phonological representations, we have a range of working theories we can use in the analysis syntactic constructions.[23] Through these, we have the means to account for formal variation as functional rather than idiosyncratic, alleviating the risk of deeming syntactic constructions as somehow non-linguistic. Once again, we cannot yet say the same for gesture, despite some recent attempts (e.g. Schlenker 2020).

Somewhat more dramatically, the 'same' meaning can be produced using different words and different syntactic constructions. Take the phrase accompanying Gesture 1 in Figure 1.2. The linguistic unit "but will their taxes go up" could change quite radically while still communicating the same proposition. I could use the "going to" future construction rather than "will". I could replace "go up" with "rise" or "increase". I could specify the referent of the pronoun "their" by saying "people in the middle class" instead. I could even rearrange the word order. Suddenly I have the sentence "for people in the middle class, are taxes going to increase?" instead of the original "will their taxes go up?" – the same meaning conveyed by radically different forms. Similarly to the other cases discussed, we have a range of working theories regarding the interfaces between syntax, semantics, and pragmatics (e.g. Asher & Lascarides 2003; Erteschik-Shir 2007; Kamp & Reyle 1993; Lambrecht 1996), as well as the effects sociolinguistic factors on lexical and syntactic variation (e.g. Tagliamonte 2011; Trudgill 2001). These theories help us to account for variation without using variation as an excuse to dismiss propositional content as non-linguistic.

What all of this suggests is that the diagnostic of obligatory form for determining linguistic status is illegitimate. Given that variation is an inevitable part of all language production, we would have to determine what counts as permissible variation while still achieving the 'obligatory form'. Because the factors determining "permissible" variation are mode and module specific – phonetic variation and lexical variation are different – one cannot simply import restrictions on variation understood for the verbal mode into the gestural mode.

[23]For example, compare the robust traditions of Minimalism (e.g. Chomsky 1965; Hornstein et al. 2005), Construction Grammar (e.g. Hoffmann & Trousdale 2013), and Lexical Functional Grammar (e.g. Dalrymple 2001).

Worse yet, until we can reliably distinguish idiosyncratic variation from functional variation in gesture forms, we cannot claim to know the inventory of underlying gestural variation nor the formal variation they are subject to in production.

1.4.3 Pre-specified meaning

The second diagnostic is more or less the inverse of the first, necessitating that a given linguistic unit must always evoke the same meaning. Using the same example from before, the linguistic unit "map" in English must always evoke the concept MAP. Given the similarity of this diagnostic to the previous, my rejections largely remain the same: (i) this diagnostic presupposes that we have a sufficient model of gesture meaning, which is not yet true; (ii) variation, not only in form, but also in meaning of spoken linguistic units, is a natural part of language, making distinguishing permissible and impermissible variation difficult; and (iii) if schematic linguistic units, such as syntactic constructions, are permitted, then 'pre-specified meaning' becomes difficult to maintain as a coherent notion.

Let's consider the 'upward movement' gestures in Figure 1.2 again. One might not notice if the three gestures and their corresponding sentences in (5) somehow got mixed up. For example, Gesture 3 shows television host Stephen Colbert sweeping his down-turned open-hand across the space in front of him in a diagonal upward and leftward trajectory. One could imagine this gesture occurring with *any* of the sentences in (5), as well as innumerable other sentences that may bear very little formal or functional resemblance to the utterance it actually occurred in. Indeed, I could imagine this gesture occurring with any of the sentences in (6).

(6) Possible utterances compatible with G3, Figure 1.2
   a. The road was quite steep for a portion of the drive.
   b. I told him to go away and bother someone else.
   c. The show was fabulous, by the way.

If this gesture is, in fact, felicitous in all of these contexts, then it would seem that it can convey meaning about *upward movement, incline, movement away from some origin, a region of space*, and some kind of *emphatic assessment*. Those are very different meanings, and we don't know which the gesture 'means' until we know the verbal utterance it occurs with. It is by this reasoning that gestures appear to lack 'pre-specified meaning'. However, by saying that a gesture can 'mean' very different things because it can occur in very different verbal contexts presupposes that the meaning of the gesture and its lexical affiliate are equivalent. I have argued against the validity of this in Section 1.3 and will continue to do so throughout this work – gestures *mean* independently.

Even if we admit that gestures convey meaning independently, one could argue that the meaning remains somehow under-specified or indeterminable out of context. This does seem to be the case. If gestures mean through the images they depict and events they enact, then it seems entirely possible that a single gesture could 'mean' very different things because

we can use the same movements to achieve very different real-life actions. For example, the formal features in G3 are compatible with, at least, tracing a particular incline, smoothing a slanted surface, and clearing a slanted surface of unwanted objects. Under such an analysis, the gesture does, in fact, lack a *single pre-specified meaning*. However, it is actually extremely difficult to find *any* linguistic unit, in any communicative mode, that can be said to have a clear *single pre-specified meaning*. To demonstrate this, we will consider the prevalence of *context-sensitive meaning* and *schematicity* in spoken language.

The requirement of 'pre-specified meaning' requires there to be a one-to-one mapping between a linguistic form and a meaning. In the lexicon, this would mean that every word form has exactly one meaning, and that meaning can be understood without considering additional context. Homonymy and polysemy are direct violations of this. Homonyms are words that share the same form but map to unrelated meanings (Lehrer 1990:207). For example, the linguistic unit "bat" in English can alternatively evoke the concept of a WOODEN IMPLEMENT or FLYING RODENT depending on the context in which it is presented (Falkum 2015:85). One cannot determine which meaning should be evoked by the form "bat" without sufficient context. This is demonstrated in (7).

(7) a. I brought my own bat to play baseball. → WOODEN IMPLEMENT
b. A bat flew out from the rafters. → FLYING RODENT
c. There is a bat in the garage. → ambiguous

Homonyms are typically analysed as distinct linguistic units that happen to share the same form through some historical accident. In other words homonyms do, in fact, maintain one-to-one mappings between form and meaning; it just so happens that the linguistic units representing the concepts of a WOODEN IMPLEMENT and a FLYING RODENT are pronounced the same in English.

One may be tempted to write homonyms off as exceptional given their relative infrequency. This would allow one to hold the claim that spoken linguistic units have *by default* pre-specified meanings, whereas pre-specified meaning in gesture is exceptional, even if we admit the possibility of homonymic gestures and acknowledge the existence of emblematic gestures. The problem is that pre-specified meaning is *not* the default in spoken language. *Polysemy* is the default, monosemy (i.e. true one-to-one form-meaning mappings) is the exception. As a somewhat trite demonstration of this, one may flip through a dictionary in an attempt to find an entry with only one 'sense'. Outside of very specific, technical, specialist vocabulary, such entries will not be found.[24]

Polysemy occurs when one linguistic form maps to multiple related meanings, typically called 'senses' (Falkum 2015). Determining which meaning should be evoked in a given

[24]Bavelas (1994), in her early argument for considering gesture as a part of language, makes a similar observation, using the "misbehavior" of the word *ground*, a seemingly common and unexceptional word, as an example. As she notes, this "misbehavior" of meaning does not seem to pose a problem for the linguistic status spoken forms. On the contrary, it is only through the careful study of this "misbehavior" that we have working theories of the interaction between word meaning and context. So why should analogous "misbehavior" pose a problem for the linguistic status of gestural forms?

instance once again requires consulting the context in which the form appears. To demonstrate this, let's reconsider the WOODEN IMPLEMENT meaning of "bat". Arguably, all of the sentences in (8) *relate* to this form-meaning mapping of "bat", and yet all but the first refer to the actual WOODEN IMPLEMENT directly.

(8) a. I brought my own bat to play baseball.
b. I'm in line to bat next.
c. The first player is at bat.
d. I'll go to bat for you if things get rough.
e. The movie was hilarious right off the bat.
f. Please bat that fly away from the cake.
g. I bat my eyelids whenever I'm nervous.

The senses of "bat" in (8b-c) relate to the WOODEN IMPLEMENT through metonymy; in (8b) "bat" refers to the action one takes with the WOODEN IMPLEMENT (in this case the particular type of wooden implement used in baseball), and in (8c) "bat" refers to the particular location that one occupies when doing the action with a WOODEN IMPLEMENT during a baseball game. The sense of "bat" in (8d) relates to the sense in (8b) via metaphor; one is expressing a willingness to defend another person by metaphorically attempting to win them points in a baseball game. The rest are much the same, often embedded in conventionalized phrases ("right off the bat" "bat away", "bat an eye") to further obfuscate meaning relations between senses.

The proliferation of senses via metonymy and metaphor is frequently discussed in terms of 'polysemic chains' and 'radial categories' (Brugman & Lakoff 1988; Lakoff 2008; Lewandowska-Tomaszczyk 2007), such that two senses may appear completely unrelated until the entire network is considered. We can apply the same frameworks in understanding the connections between use contexts in gesture.[25] Under such an analysis, using the same gesture for the sentences in (6) is not as surprising as it may have first seemed. Consider the full articulation of the gesture G3 in Figure 5 – the hand, with an open palm facing downward, moves from a point directly in front of Colbert's body leftward and upward to around his left shoulder. First, whether Colbert is *depicting* or *enacting* movement may be considered homonymic.[26] When *depicting* movement, the gesture may come to mean upward movement of an object, as it does in the context it was produced (5c), or metaphoric upward movement (5a-b). Through metonymy, it can then come to highlight certain aspects of the upward movement, such as its speed or steepness (6a). When *enacting* a movement, the gesture may come to mean the physical or metaphoric removal of objects from the speaker's immediate space, as

[25] Mittelberg (2018) offers a similar account of multifunctionality in gesture, though not in intentional relation to lexical polysemy.

[26] This distinction is often referred to as taking an 'observer viewpoint' or 'character viewpoint' in the literature (Parrill & Sweetser 2004).

in (6b). A metaphoric extension of this use may involve the removing of alternatives, as when describing things as the *only* or *best* option.[27]

The treatment of gestures as fundamentally *polysemous* is central to the understanding of gestural meaning in the present work. I take especially seriously pragmatic-based approaches to polysemy in lexical semantics (e.g. Carston 2002; Ruhl 1989; Sperber & Wilson 1998)[28] to account for the diversity of use contexts in gesture. These approaches allow the 'pre-specified' meaning of a lexical item to be abstract and schematic, much like what is required for considering the meaning of gesture on its own terms. Using insights from these approaches, I argue that there is a core meaning of a gesture that is enriched in context via integration with the verbal mode. A gesture's core meaning is the action or image schema it is associated with (e.g. Müller 2017) and is directly observable in the gesture's physical form. Methods for deriving an action schema from a gesture's formal features will be discussed at length in Chapter 2. As is the case with lexical items, the inherent polysemy of gesture does not undermine its status as an independent unit of meaning or as a linguistic unit.

1.4.4 Arbitrary form

The third diagnostic places a restriction on what types of meaning can be considered linguistic – *arbitrary* meaning is linguistic, whereas *iconic* meaning is not. Iconicity refers to a resemblance between the form and its meaning. In speech, iconic meaning is most obvious in cases of onomatopoeia, in which the sound of the word resembles the real-world sound it refers to, such as "meow" or "boom". Arbitrariness refers to a *lack of* resemblance between form and meaning. The majority of lexical forms in spoken language can be said to be arbitrary – there is no resemblance between the word "dog" and the four-legged mammal it refers to. Meaning in gesture, with the exception of emblems, relies heavily on iconicity – the gestures we have seen used to refer to upward motion do so by literally moving upwards. There are two major issues with using arbitrariness as a requirement for language: (i) if signed languages are to be accepted as fully linguistic, than iconicity must be accepted as a source of linguistic meaning as well; (ii) spoken and written language demonstrate a significant degree of iconicity when one looks beyond just the lexicon.

Iconicity plays a significant role in meaning making in sign languages, both at the level of individual signs and in larger grammatical constructions (e.g. Taub 2001). Though there are also arbitrary elements, the acceptance of sign languages as full languages and into the study of linguistics proper requires a recognition of iconicity's potential for conveying linguistic meaning. To argue otherwise would be to argue that sign languages are full languages that convey meaning through primarily non-linguistic means. In Hockett's (1978) defense of sign languages as full languages, he argues that communicative modalities take advantage of whatever tools and affordances they have at their disposal. The necessary linearization of speech reduces its capacity for iconicity, as our experience of the world is very much

[27]For further discussion of clearing gestures to express exhaustivity and maximality see Kendon (2004: Ch. 13).

[28]Also see Falkum (2015) for a more recent review of approaches to polysemy in lexical semantics.

multidimensional. Sign languages, as well as gesture, *can* take advantage, and thus *do* take advantage, of three-dimensional space in representing meaning, making a higher degree of iconicity both expected and efficient. As Hockett says, when we fetishize arbitrariness as a fundamental part of language, we are simply making "a virtue of necessity" (Hockett 1978:273-275). If we *do* accept iconicity in sign language as linguistic, then it is difficult to justify holding gesture to a different standard just because it co-occurs with speech.[29]

A preoccupation with making "a virtue" of arbitrariness may also result in an under-appreciation of the role of iconicity in speech. The central role of iconicity in speech is argued for extensively in work on ideophones (e.g. Dingemanse 2012; Perniss et al. 2010) and structural iconicity (e.g Givón 1985; Newmeyer 1992).[30] The study of ideophones includes prototypical examples of onomatopoeia, while also highlighting the relatively widespread grammaticalization of echoic units, phonesthemes, and sound-symbolism. For example, words like "bang" and "crash" have been extended from typical onomatopoeic uses via metonymy to refer to events related to, but not equivalent with, actual sounds (e.g. compounds like "car crash" and idioms like "crash and burn"). In the domain of 'structural iconicity', it has been argued that fundamental grammatical constructions, such as word order, argument structure and derivational morphology, are motivated by iconicity.[31] For example, the distinction between arguments and adjuncts and their position relative to verbs can be interpreted as iconic – the more closely related constituents are, the closer they tend to be syntactically. The role of iconicity in prosodic structure is even more apparent – we draw psychological attention to certain items (e.g. the focused constituent) by drawing perceptual attention to those items (e.g. through pitch accents).

I will discuss the role of iconicity in gesture meaning at length in Chapter 2, and will often return to it when analyzing particular gestural sequences. However, I will not return to the issue of arbitrariness. Given the central role of iconicity in gesture meaning, I do not believe that appealing to notions of arbitrariness will benefit the development of working model meaning in gesture. Going forward, I will take for granted that arbitrariness is not a prerequisite for linguistic status.

1.4.5 Cultural specificity

The final diagnostic requires that linguistic forms are subject to cross-cultural variation. The cultural specificity diagnostic is closely related to the arbitrariness diagnostic discussed above. Because iconic representations resemble their meaning in some way, it is argued that they are more likely to be shared across cultures. Moreover, they may be acquired without cultural transfer, i.e. without being learned directly from other members of the language community. Arbitrary linguistic forms, on the other hand, must be taught since

[29]Legitimizing this kind of double standard is even more difficult when one considers the occurrence of gesture in sign languages (e.g. Liddell & Metzger 1998; Liddell 2003).

[30]Also see Dingemanse et al. (2015) for a general overview of iconicity and non-arbitrariness in language.

[31]See Haspelmath (2006) for discussion of the relationship between iconicity and theories of grammatical 'markedness'.

they cannot be recognized or derived via resemblance to their meaning. Because of this, it is also expected that arbitrary forms will vary significantly between language communities. For example, the concept DOG is realized as "dog" in English, "hond" in Dutch, and "chien" in French. Despite evoking the same concept, the realizations bear no resemblance to one another. Someone who speaks French, but not English, would understand "chien" but would be unable to interpret the form "dog". Moreover, one may be able to spontaneously create a word for the sound dogs make that resembles a conventionalized linguistic form (e.g. "woof" or "arf"). One could not do the same for "dog".

Because the requirement for cultural specificity is largely a result of the requirement for arbitrariness, many of the reasons for questioning the legitimacy of this diagnostic are covered by discussion in the previous section. However, it is still worth considering some ways in which gestures exhibit cross-cultural variation *despite* their iconicity.[32]

Many of the interactive gestures discussed in this work have been observed across many language communities, which would at first seem to violate the requirement of cultural specificity. For example, the family of 'away' gestures, in which a gesturer's hand moves away from their body as if to remove an unwanted object, have been observed to be associated with negation and negative assessment in English (Germanic; Streeck 2009), German (Germanic; Bressem & Müller 2014, 2017), Israeli Hebrew (Semitic; Inbar & Shor 2019) Italian (Romance; Kendon 2004), Spanish (Romance; Teßendorf 2014), Savosavo (Papuan; Bressem et al. 2017), and Syuba (Tibeto-Burman; Gawne 2021). Still, cross-cultural variation in particular forms and functions occur. Gawne (2021), for example, observes that Syuba speakers' 'brushing away' gestures tend begin with a bunched hand shape, unlike those discussed for German and Spanish speakers, and seem to be more closely associated with 'absence' than with negative assessment.

Though iconicity does result in more recognizably shared form-meaning mappings across cultures, there is also variation. This suggests that gestures such as those considered in this work are learned through cultural transmission to at least some extent, and thus do satisfy an important aspect of the culturally specific diagnostic.

1.4.6 Discrete meaning and compositionality

The final argument against gesture-as-language that I will consider is the supposed non-compositionality of gestural meaning. Compositionality is core to models of linguistic meaning – units of meaning compose according to particular principles to form larger, more complex units of meaning. Because of this, dismissing the compositional capacities of gesture directly impacts the perceived relevance of gesture to linguistic study. I will briefly outline the basis of the debate here. The compositionality of gestural meaning will remain central to gesture analyses throughout this work.

In addition to the 'language-like' properties discussed above, McNeill's (2000; 2005) four continua for distinguishing 'gesticulation' from 'sign' have also been influential in determining

[32]For an overview of cross-cultural variation in gesture see Kita (2009).

the linguistic status of gesture. These have been used specifically to admit sign languages into our definition of 'language' proper without admitting gestures that co-occur with speech.[33] The first continuum is concerned with the relationship to and alignment with speech. The definition of gesture used in this work (see Section 1.3) draws no distinction between gesture used during or in the absence of speech. Given this, McNeill's first continuum can be considered orthogonal to the present work. The following two, 'relationship to linguistic properties' and 'relationship to conventions' are closely related to the diagnostics discussed in previous sections. The final proposed continuum, 'character of the semiosis', pertains directly to the question of how gestures convey meaning, a central question of the present work and thus worth further consideration here.

To distinguish the 'character of semiosis' of gesture and sign, McNeill (2000; 2005) introduces two dimensions of meaning: *global* vs. *segmented* and *synthetic* vs. *analytic*. The first distinction, *global* vs. *segmented*, relates to whether or not a given form is decomposable into meaningful parts. The second distinction, *synthetic* vs. *analytic* relates to the ability of a given form to compose with other adjacent forms. McNeill, and others since, argue that meaning in gesture is inherently *global & synthetic*, whereas meaning in speech (and sign) is *segmented & analytic*.

The claim that gestural meaning is *global* suggests that the individual parts of a gesture (e.g. movement, orientation, hand shape) are only meaningful as parts of the overall gesture and are not interpretable individually. Parrill (2008:198) points out that a gesture analyst may very well be able to assign meaning to individual features of a gesture, but holds that language-users in general cannot do the same. In speech, on the other hand, 'non-experts' can assign meaning to individual parts, such as to words in a sentence, or morphemes in a word.[34] The claim that gestural meaning is *synthetic*, suggests that a single gesture can convey complex meaning 'holistically', while a spoken utterance can only do so by putting a series of words into the proper linear order. This is in contrast with speech linguistic items that "convey information more selectively" Parrill (2008:198). As in the sections above, there are two angles from which to critique these claims. First, we must carefully consider whether or not gestural meaning is actually *global* and *synthetic*. I will argue here, and throughout this work, that it is not. Second, we must responsibly question our intuitions as to whether meaning in speech (as well as in writing and sign) is actually *segmented* and *analytic*. I will argue in this section that it is not, given established approaches to linguistic meaning, especially Frame Semantics (Fillmore 1976).

[33]For a review of this debate see Müller (2018).

[34]I do not know of any work that rigorously measures the ability of 'non-experts' to interpret aspects of a gesture's form. However, if McNeill and Parrill are right in their intuitions, literacy and Westernized education would prove serious confounding factors. In cultures that emphasize literacy, significant attention is paid to training students in the use, analysis, and composition of words from a very early age. The same attention is not paid in educational settings to gesture. We can thus say that all educated language users are, in a way, trained speech analysts, but not trained gesture analysts. In this framing, a claim that meaning in gesture is non-linguistic because a general audience cannot analyze it is dubious. Bavelas (1994:210) makes a similar point.

The claim that gestural meaning is 'global and synthetic' presupposes that we already know how gestures mean, which, as I have discussed above, is not necessarily the case. If one takes an action schema approach to gestural meaning, as I do in this work, then the ability for gesture to be both 'segmented' and 'analytic' is apparent. This is because *actions* are, by and large, segmented and analytic. Imagine the actions involved in taking a sip of your morning cup of coffee. This event can be decomposed into different steps (e.g. pick up coffee cup, bring coffee cup to mouth, sip coffee, return coffee cup to rest). Each step can be broken into particular actions and features of those actions. For example, in order to pick up your coffee cup you will need to extend your arm a certain distance and in a certain direction while shaping your hand in such a way that you will be able to securely hold the cup. Changing any feature in this action sequence may result in a very different event occurring. If you shape your hand correctly and extend your arm far enough, but reach in a different direction, you may pick up your partner's tea rather than your coffee. If you get the direction and extension right, but meet your coffee cup with a flat hand rather than curved fingers, you may end up pushing your coffee off the table. Each step of the event, and each part of each action involved in the event, is also *interpretable*. You shape your hand a certain way because you need to support an object of a particular size and shape. You reach your arm in a certain direction and to a certain length because the coffee cup is in a particular location relative to your body. If I asked you why you were doing one of these things, you would probably be confused as to why I was asking, but could easily answer nonetheless. This all suggests that the action of taking a sip of coffee is 'segmented', i.e. it can be decomposed into meaningful discrete units.

Gestures that mean through the enactment of action schema are much the same. Consider the 'palm-up open hand' gesture again. A gesturer may hold their upturned open hand out toward their interlocutor as they say "the thing is...", as if to hold up their argument as a metaphoric object for inspection. The ability of this gesture to metaphorically *present an object* is dependent on the composition of its parts – the hand must be open and upturned, as if to support an object, and the arm must be extended toward the addressee, as if to bring the object to them for inspection.

One may argue that actions, and by extension gestures, may be *technically* decomposable without being *practically* decomposable. When we are going to pick up a coffee cup, we don't think about all of the particular actions we need to take, we think about the event as a whole. We think about drinking our coffee, or perhaps, more abstractly, that we need to drink our coffee in order to start our work for the day. Perhaps this is the difference – action, and gesture by extension, is technically decomposable, while linguistic meaning is practically decomposable, i.e. we care about the parts. The problem with such reasoning is that the same can often be said for speech. In everyday conversation, we are usually not particularly interested in individual words, or choosing particular morphemes to compose those words.[35]

[35]Not to mention those cases in which too much interest in individual words would distract from the meaning of the whole, as is the case for idioms (e.g. "spill the beans") and collocations (e.g. "take a shower").

We care about events and stories. We care about getting our ideas across and completing tasks together.

So, sentences, gestures, and actions can all be 'segmented' into meaningful parts that we may or may not care about at the time of performance. This leaves the issue of 'synthetic' and 'analytic' meaning. Whether meaning is 'synthetic' or 'analytic' depends on what types of meaning we determine to be sufficiently "selective", to borrow from Parrill's (2008:198) phrasing. The problem is that this is impossible to determine in a principled way.[36]

We might begin by saying that words have sufficiently 'selective' meaning, but which words? To demonstrate the problem that arises here, consider the two sentences in (9).

(9) a. I had breakfast this morning.
b. I had coffee and an egg sandwich this morning.

It seems in this case that we have an event, BREAKFAST, that can be expressed either as a single word, "breakfast", or as a list of things that were consumed at the appropriate meal time. If you knew that I had coffee and an egg sandwich every day for breakfast, then these two sentences would convey identical information, despite having a different number and set of words. Does this mean that "breakfast" is in some way *less* selective?[37]

Very frequently, the articulation of a gesture will extend through the articulation of several words, as we saw with the "throwing" gestures in (3). This has been taken to mean that a single gesture expresses an event holistically, whereas individual words only convey individual parts of an event. However, the discussion of (9) above shows this is not a coherent distinction – the same event can be expressed by one ("breakfast") or many ("coffee and an egg sandwich") words. I argue that the claim that gestures convey more 'synthetic' meaning than speech can only be upheld if one equates the meaning of a gesture with the meaning of the speech it accompanies. I have argued against this stance, and have pointed to several cases in which it is demonstrably false (Section 1.3).

1.5 Multimodal linguistics as a stronger linguistics

The reliance on diagnostics as a means of determining what *is* and *is not* language can only work if we have some kind of neutral way to decouple the proposed diagnostics from a particular form of language. This is not the case for the diagnostics discussed above, or indeed any set of diagnostics I have seen, including Hockett & Hockett's (1960) original 'design features of communication'. Instead, these diagnostics rely on the following logic: because we know that speech is a form of language, we can derive our diagnostics for language

[36]In an unrelated, but surprisingly relevant debate, the status of *languages* as 'synthetic' or 'analytic' is contested on similar grounds (e.g. Haspelmath & Michaelis 2017). There seems to be no way to determine the boundaries of each category without choosing a language as a reference point. Needless to say, choosing a single language to build a general theory of language around has proven problematic.

[37]This dilemma underlies the search for 'semantic primitives' and the development of a 'Natural Semantic Metalanguage' (e.g. Goddard 2012; Wierzbicka 1972). It is worth noting that the result of such proposed metalanguages look nothing like naturally occurring languages.

from the features of speech. The inevitable result of this is a set of diagnostics for determining whether or not something is *speech*. Rather than discovering universal features of 'language', we have equated speech to language a priori. Gesture is, of course, *not speech*, and so will inevitably fail such diagnostics.

If linguistics is to embrace multimodality and a more holistic approach to the study of language, as the 'embodied turn' suggests, we need to do more than accept gesture as a part of language. We need to do more than incorporate gesture into the theories we've built for speech. We need to de-center speech and rethink our theories. If we do not make the effort to do so, we risk maintaining speech, even if only implicitly, as the *prototype* of linguistic communication. There is an insidiousness to this oversight, because what we maintain as the prototype of linguistic communication inevitably shapes our research questions, our methodologies, and our interpretations of the data. We know from seminal work on categorization and cognition (Rosch 1978, 1983) that prototypes can lead to both *asymmetric reasoning* and *stereotypes*, both of which can bias researchers toward asking certain questions and mask the potential for asking others.

Feature-based definitions of language expose some of the dangers of upholding speech as a prototype. I discussed the ways in which speech regularly fails the diagnostics that have been used to exclude gesture from our definition of language. This does not only reflect problems in the diagnostics themselves, but also shortcomings in our study of spoken language. In maintaining speech as a prototype, we undervalue in speech those aspects that it shares with gesture. We avoid considering polysemy and schematic meaning, dismiss iconicity, and miss opportunities to refine our notions of compositionality beyond its linear manifestations. Of course, much work has been done attempting to right these wrongs; new frameworks form around 'exceptions' in order to argue against their exceptionality.[38] However, the point remains that maintaining speech as the center of linguistic study is an unnecessary hindrance to theoretical and methodological innovation.[39]

De-centering speech in linguistic study is not only to the benefit of the gesture researcher, or the sign language researcher, it is also to the benefit of the speech researcher. By widening our understanding of language to include all conventionalized systems of communication, especially gesture, we expand our research focus, and in doing so discover new questions to ask, and new ways to ask old questions. It is thus not only in the interest of the gesture researcher to advocate for gesture as part of our language system, it is also in the interest of the syntactician, semanticist, and discourse analyst to advocate for gesture's integration into linguistic study.

[38] A particularly nice example of this may be Construction Grammar's embrace of idioms. Instead of dismissing idioms as exceptional, Construction Grammar uses idioms to illustrate a continuum between lexical and structural meaning that is underappreciated in other frameworks.

[39] In work on the 'Written Language Bias', Linell (2005) has argued that upholding writing as a prototype of language use has negatively impacted our theories of language. An embodied linguistics takes this critique a step further, calling for the remediation of a *Spoken* Language Bias.

Chapter 2

Describing gesture and multimodal data

2.1 Introduction

In this chapter I outline the methodologies I've used in this work during data collection, annotation, and analysis. I begin by reviewing the data collection process and how data is presented in this work (Section 2.2). I then discuss the strengths and weaknesses of the particular data set, focusing on the influence of particular spatial and social organizations on gesture performance (Section 2.3). After this, I move on to describe my multi-step procedue in analyzing gesture meaning independently (Section 2.4) and in context (Section 2.5). These two sections lay the groundwork for developing the model of interactive meaning in gesture presented in Chapter 4. Section 2.6 concludes with a brief discussion of how these methodologies are applied in subsequent chapters.

2.2 Data collection and presentation

All data comes from monologues and interviews on the American talk show *The Late Show with Stephen Colbert*. Data was collected through UCLA's Communication Studies Archive in collaboration with the Red Hen Lab using the "Edge" search interface depicted in Figure 2.1. This archive contains over two-hundred-thousand videos of news broadcasts from around the world, beginning in 2005 to the present.[1] The archive is not tagged for speech or gestural information, but is searchable through automatically generated transcripts of the videos. The Red Hen Lab is an international, distributed research organization focused on multimodal communication (Joo et al. 2017).[2] Access is granted to individual gesture researchers on a case-by-case basis through a sponsor system.

[1] More information available at https://tvnews.sscnet.ucla.edu/public/

[2] More information available at https://sites.google.com/site/distributedlittleredhen/

The Edge search interface allows users to search the corpus for specific phrases. Searches can be limited to particular time ranges, television networks, and individual shows. The search displayed in Figure 2.1 is for variants of the lexical discourse marker *anyway* ("anyway" and "anyways") occurring between January 1, 2015 and December 31, 2019 on the KCBS network show *The Late Show with Stephen Colbert*.

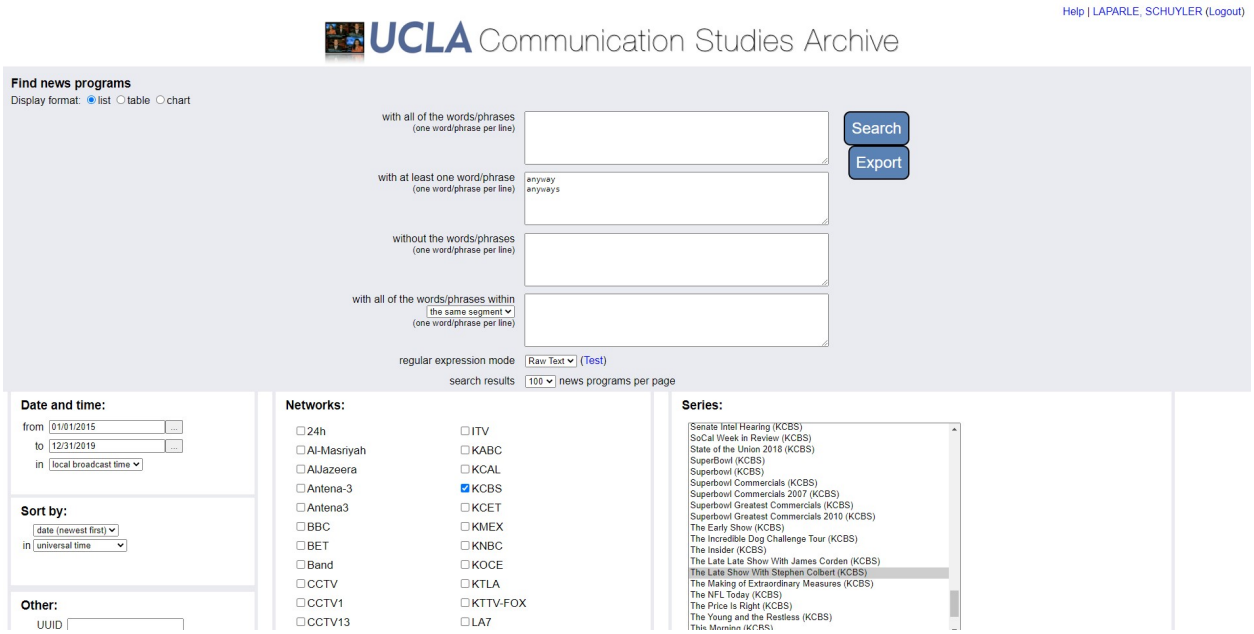


Figure 2.1: Search page in the UCLA News Archive

Search results are then displayed in reverse-chronological order, grouped by occurrences within a single video file. Each instance of the search term is listed with a segment of transcript and a timestamped link to the relevant video clip. The first listed instance of the search item in each video also provides links to that video's metadata and full transcript. The search results for the term "anyway" are depicted in Figure 2.2. Pressing the small clip image on the left side of the screen transfers the time-stamped segment into an enlarged video screen to the right. This view allows the researcher to re-watch the clip at various speeds (from 0.25x to 2x speed in increments of .25) as well as rewind and fast forward the video by intervals 10 seconds, 45 seconds, and 10 minutes. Finally, screenshots of individual frames can be taken and are immediately downloaded to the user's computer as individual .jpg files. This feature has been used in attaining the majority of screenshots presented in this work. Where more fine-grained image extraction was required (e.g. when looking at multiple parts of a single gesture), video clips were downloaded and opened in the ELAN video annotation software which allows for frame-by-frame analysis (Wittenburg et al. 2006).

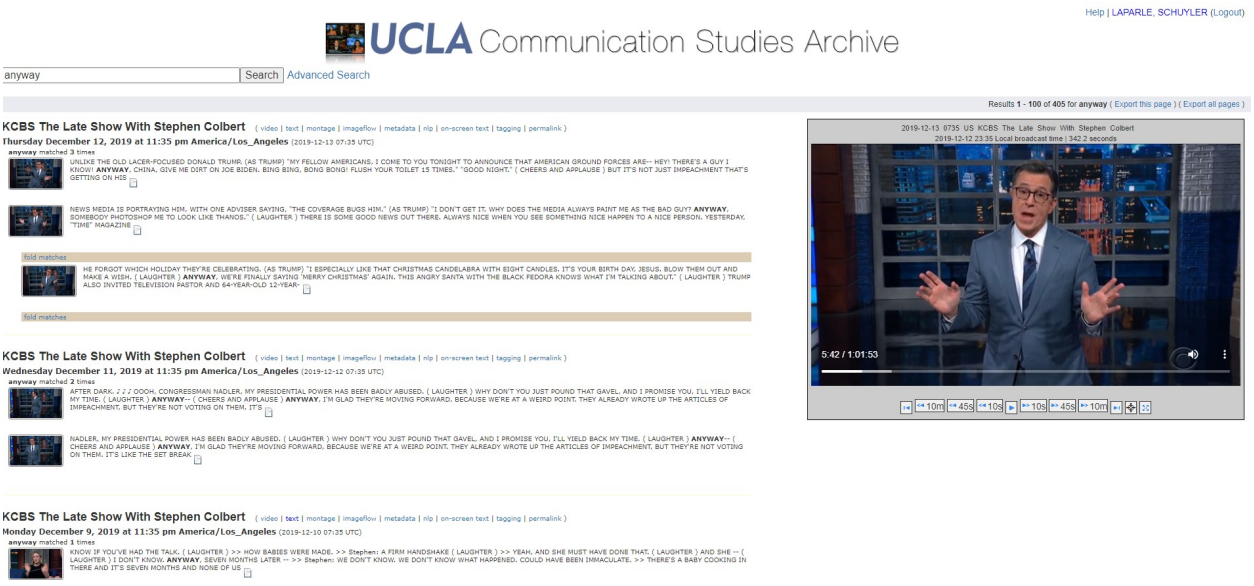
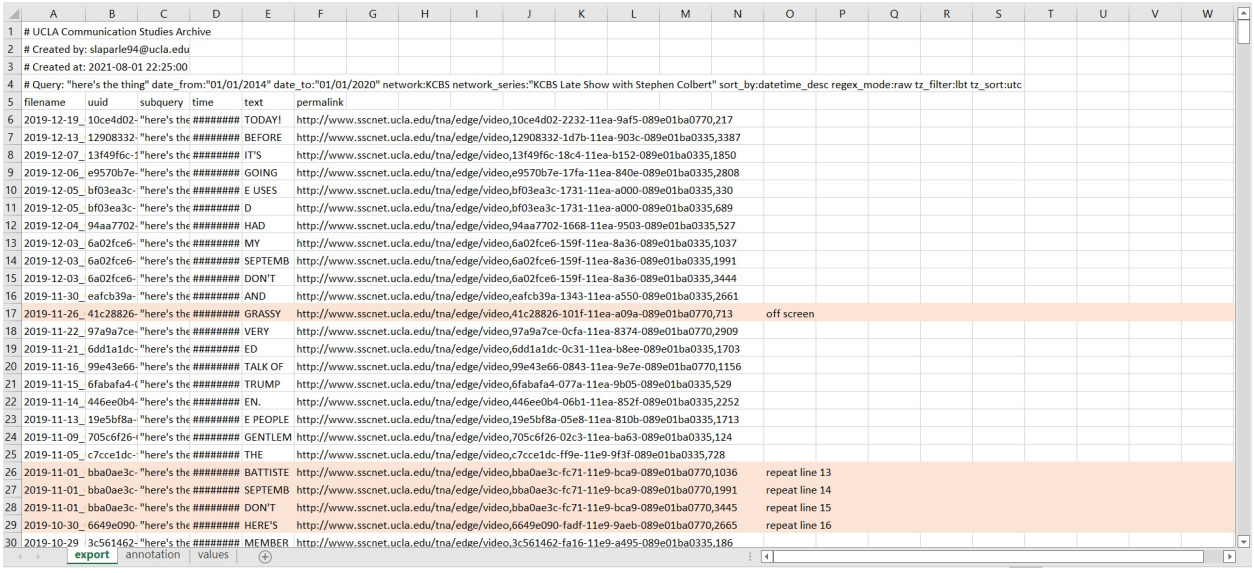


Figure 2.2: Result page in the UCLA News Archive



	A	B	C	D	E	F
1	# UCLA Communication Studies Archive					
2	# Created by: slaparle94@ucla.edu					
3	# Created at: 2021-08-01 22:25:00					
4	# Query: "here's the thing" date_from:"01/01/2014" date_to:"01/01/2020" network:KCBS network_series:"KCBS Late Show with Stephen Colbert" sort_by:datetime_desc regex_mode:raw tz_filter:lbt tz_sort:utc					
5	filename	uuid	subquery	time	text	permalink
6	2019-12-19_	10ce4d02-	"here's the	########	TODAY!	http://www.sscnet.ucla.edu/tna/edge/video,10ce4d02-2232-11ea-9af5-089e01ba0770,217
7	2019-12-13_	12908332-	"here's the	########	BEFORE	http://www.sscnet.ucla.edu/tna/edge/video,12908332-1d7b-11ea-903c-089e01ba0335,3387
8	2019-12-07_	13f49f6c-1	"here's the	########	IT'S	http://www.sscnet.ucla.edu/tna/edge/video,13f49f6c-18c4-11ea-b152-089e01ba0335,1850
9	2019-12-06_	e9570b7e-	"here's the	########	GOING	http://www.sscnet.ucla.edu/tna/edge/video,e9570b7e-17fa-11ea-840e-089e01ba0335,2808
10	2019-12-05_	bf03ea3c-	"here's the	########	E USES	http://www.sscnet.ucla.edu/tna/edge/video,bf03ea3c-1731-11ea-a000-089e01ba0335,330
11	2019-12-05_	bf03ea3c-	"here's the	########	D	http://www.sscnet.ucla.edu/tna/edge/video,bf03ea3c-1731-11ea-a000-089e01ba0335,689
12	2019-12-04_	94aa7702-	"here's the	########	HAD	http://www.sscnet.ucla.edu/tna/edge/video,94aa7702-1668-11ea-9503-089e01ba0335,527
13	2019-12-03_	6a02fce6-	"here's the	########	MY	http://www.sscnet.ucla.edu/tna/edge/video,6a02fce6-159f-11ea-8a36-089e01ba0335,1037
14	2019-12-03_	6a02fce6-	"here's the	########	SEPTEMB	http://www.sscnet.ucla.edu/tna/edge/video,6a02fce6-159f-11ea-8a36-089e01ba0335,1991
15	2019-12-03_	6a02fce6-	"here's the	########	DON'T	http://www.sscnet.ucla.edu/tna/edge/video,6a02fce6-159f-11ea-8a36-089e01ba0335,3444
16	2019-11-30_	eafcb39a-	"here's the	########	AND	http://www.sscnet.ucla.edu/tna/edge/video,eafcb39a-1343-11ea-a550-089e01ba0335,2661
17	2019-11-26_	41c28826-	"here's the	########	GRASSY	http://www.sscnet.ucla.edu/tna/edge/video,41c28826-101f-11ea-a09a-089e01ba0770,713
18	2019-11-22_	97a9a7ce-	"here's the	########	VERY	http://www.sscnet.ucla.edu/tna/edge/video,97a9a7ce-0cfa-11ea-8374-089e01ba0770,2909
19	2019-11-21_	6dd1a1dc-	"here's the	########	ED	http://www.sscnet.ucla.edu/tna/edge/video,6dd1a1dc-0c31-11ea-b8ee-089e01ba0335,1703
20	2019-11-16_	99e43e66-	"here's the	########	TALK OF	http://www.sscnet.ucla.edu/tna/edge/video,99e43e66-0843-11ea-9e7e-089e01ba0770,1156
21	2019-11-15_	6fabafa4-(	"here's the	########	TRUMP	http://www.sscnet.ucla.edu/tna/edge/video,6fabafa4-077a-11ea-9b05-089e01ba0335,529
22	2019-11-14_	446ee0b4-	"here's the	########	EN.	http://www.sscnet.ucla.edu/tna/edge/video,446ee0b4-06b1-11ea-852f-089e01ba0335,2252
23	2019-11-13_	19e5bf8a-	"here's the	########	E PEOPLE	http://www.sscnet.ucla.edu/tna/edge/video,19e5bf8a-05e8-11ea-810b-089e01ba0335,1713
24	2019-11-09_	705c6f26-(	"here's the	########	GENTLEM	http://www.sscnet.ucla.edu/tna/edge/video,705c6f26-02c3-11ea-ba63-089e01ba0335,124
25	2019-11-05_	c7cce1dc-	"here's the	########	THE	http://www.sscnet.ucla.edu/tna/edge/video,c7cce1dc-ff9e-11e9-9f3f-089e01ba0335,728
26	2019-11-01_	bba0ae3c-	"here's the	########	BATTISTE	http://www.sscnet.ucla.edu/tna/edge/video,bba0ae3c-fc71-11e9-bca9-089e01ba0770,1036
27	2019-11-01_	bba0ae3c-	"here's the	########	SEPTEMB	http://www.sscnet.ucla.edu/tna/edge/video,bba0ae3c-fc71-11e9-bca9-089e01ba0770,1991
28	2019-11-01_	bba0ae3c-	"here's the	########	DON'T	http://www.sscnet.ucla.edu/tna/edge/video,bba0ae3c-fc71-11e9-bca9-089e01ba0770,3445
29	2019-10-30_	6649e090-	"here's the	########	HERE'S	http://www.sscnet.ucla.edu/tna/edge/video,6649e090-fadf-11e9-9aeb-089e01ba0770,2665
30	2019-10-29_	3c561462-	"here's the	########	MEMBER	http://www.sscnet.ucla.edu/tna/edge/video,3c561462-fa16-11e9-a495-089e01ba0335,186

G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W
								off screen								
								repeat line 13								
								repeat line 14								
								repeat line 15								
								repeat line 16								

export | annotation | values

Figure 2.3: Export file contents for corpus search results

Search results are then exported to a .csv file that contains query information (the variables defined in the search), and a list of each result's file name, unique ID code, permanent time-stamped URL, and transcript segment. One such export file is displayed in Figure 2.3 for the search term "here's the thing". I then trimmed the data for duplicate and unsuitable examples. Duplicate videos appear due to rebroadcasting of the same show on different

dates. Only the most recent broadcast was kept for annotation. Unsuitable examples include any videos that (i) do not take place on the *Late Show with Stephen Colbert* set in one of three standard interaction formats (as listed in (10c)), or (ii) do not display the speaker's hands. Files that were not included in the final data set were kept in the extraction file and highlighted in red with a reason for exclusion.

2.2.1 Annotation formatting

All annotations were recorded into an Excel spreadsheet. For each case study there was a separate Excel workbook containing three sheets (i) the original export file and reasons for data exclusion, as displayed in Figure 2.3; (ii) the annotation file containing metadata, speaker-addressee information, discourse structural information, and gestural information, as displayed in Figure 2.4; and (iii) a 'values' file which was used for setting data validation criteria in the annotation file.

The annotation files for all three case studies (*anyway, by the way,* and *here's the thing*) include the columns summarized in (10).

(10) **Common annotation scheme**

a. Columns A-D: metadata copied from extraction file (corpus file name, unique identification number, time-stamped permalink, automatically generated transcript)

b. Column E: show; in this data set always *Colbert* for *The Late Show with Stephen Colbert*

c. Column F: interaction type – *monologue, interview, band interaction*

d. Column G: speaker – *Colbert, guest, correspondent*

e. Columns H-I: Guest name and demographic information (gender, race, nationality, profession); left blank for monologues

f. Column J: lexical unit – lexical discourse marker of interest plus adjacent lexical discourse markers[3]

The annotation formatting for discourse structural information and gestural information varied slightly across the three discourse markers. However, all annotation files included the information in (11).

(11) a. **Discourse structural information**: turn position, discourse relation

b. **Gestural information**: handedness, handshape, palm orientation, movement, gesture-speech alignment, gesture class

c. **Other**: "starred" (to mark examples for particular interest), "comments" (qualitative description of example)

[3]Where "adjacent" means (i) part of the same host clause, and (ii) with no intervening non-lexical discourse marker. This means that in the string "uh so anyway I uh", the lexical unit would be listed as "uh so anyway".

Additional information was annotated depending on other factors of interest for each discourse marker. For example, the *here's the thing* annotation sheet in Figure 2.4 includes a column (column K, "syntax") for noting variation in the lexical discourse marker's phrase structure (e.g. *independent* "here's the thing" vs *extended* "here's the thing that bothers me"). This type of variation was not observed for the other two lexical discourse markers, and so was not annotated.

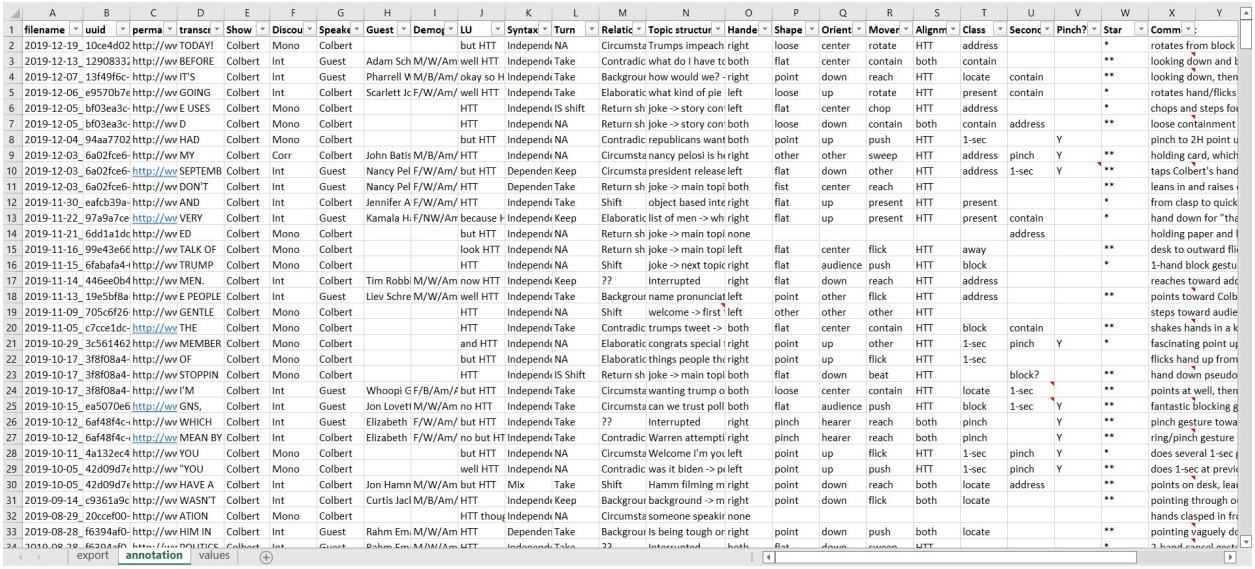


Figure 2.4: Excel annotation sheet for *here's the thing* data set

The most important aspects of each case study's annotation scheme will be summarized in the relevant chapter. The annotation of gestural features and determination of gesture 'class' will be described in detail in Section 2.4.

2.2.2 Data presentation

With the exception of this chapter, all examples discussed are presented with a transcript. Figure 2.5 shows the data presentation for examples in which only the verbal context is being considered. Where gesture is also considered, data is presented as an annotated transcript followed immediately by a co-indexed set of screenshots, one for each gesture in the annotated transcript. This is depicted in Figure 2.6. Annotated transcripts include either one or two gestural tiers (one for hand gestures, one for non-manual postural shifts), depending on what behavior is being considered at the time of presentation.[4]

[4]This presentation format is based on that proposed for Embodied Conversation Analysis in Mondada (2018).

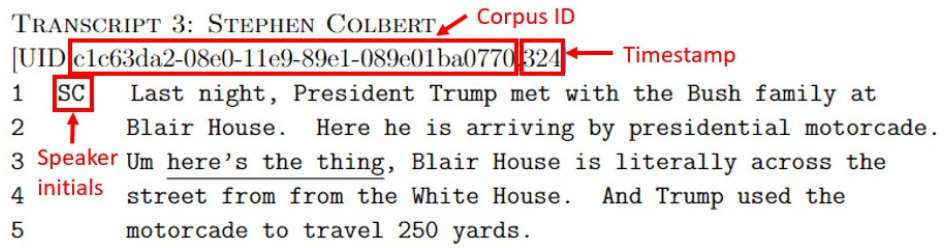


Figure 2.5: Data presentation for verbal context only (example taken from Ch.7)

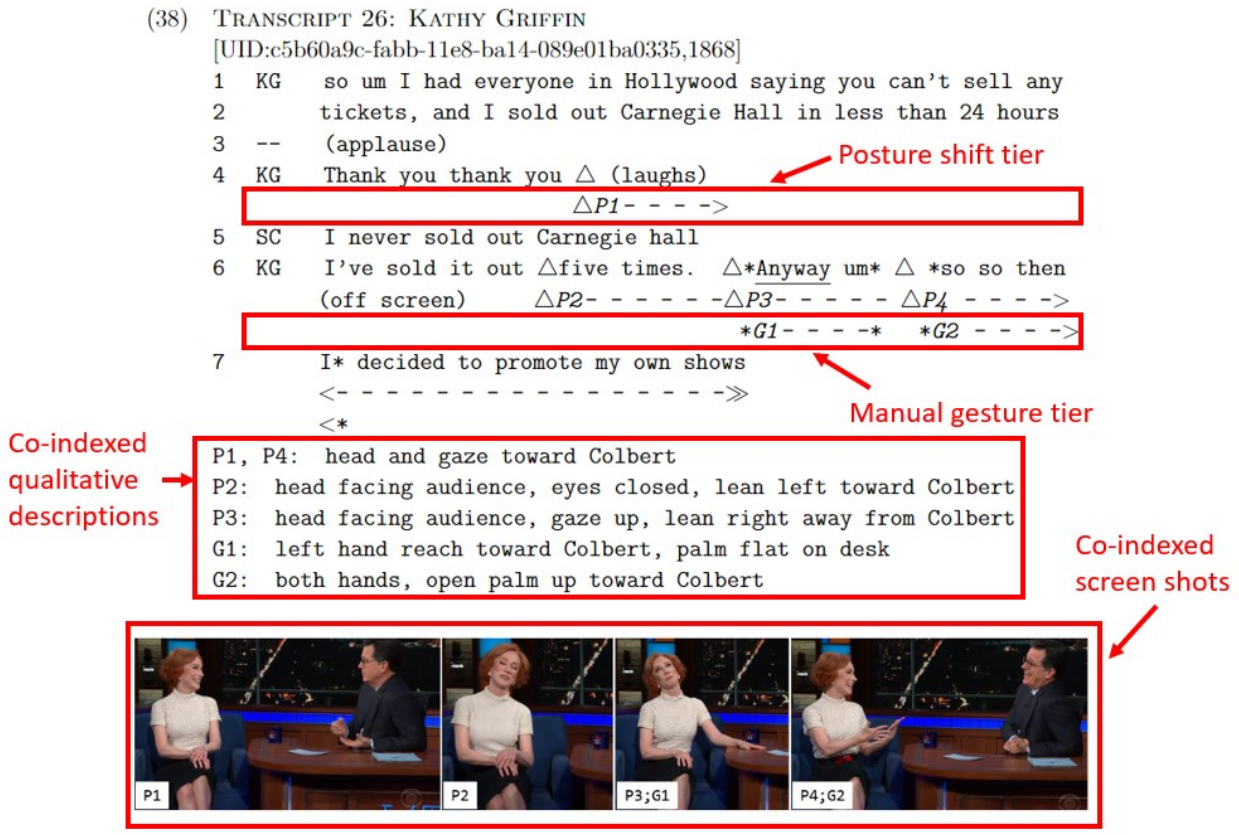


Figure 4.6: Gesture sequence corresponding to TRANSCRIPT 26: KATHY GRIFFIN, ex. 38.

Figure 2.6: Data presentation for manual and non-manual gesture in verbal context (example taken from Ch.6)

In the remainder of this chapter, I provide screenshots without accompanying transcripts. This is exceptional. I have chosen to do so because the primary goal of the remainder of this chapter is to demonstrate a methodology for describing gesture form and meaning *without* relying on accompanying speech.

2.3 Strengths and limitations of a one-show data set

Because all data comes from a single television show, *The Late Show with Stephen Colbert*, the present data set is relatively restricted and subject to certain idiosyncratic features. This simultaneously enhances and hinders aspects of the analysis. On the one hand, we are able to directly observe the effects of spatial organization and interaction type on gesture performance. Because there are a limited number of spatial and social arrangements, we can more easily explore the extent to which these factors contribute to variation in gesture form and contextualized gestural meaning. However, it is also worth acknowledging that limiting the diversity in data impacts the generalizability of any conclusions drawn. The idiosyncratic strengths and weaknesses of the present data set are described in turn.

2.3.1 Spatial organization and interaction types

All examples in the data set occur on the stage depicted in Figure 2.7.[5] Each number (1-5) overlaid on the image represents the positioning of particular interlocutors in particular roles. Stephen Colbert, the show's host, occupies Areas 1 and 2 in different segments of the show. Interviewees occupy Area 3. Colbert's house band, with which he occasionally interacts, occupies Area 4. The live audience is in front of the stage, indicated as Area 5. Each space has a particular set of physical and social affordances that can influence gesture performance.

Figure 2.7: The Late Show with Stephen Colbert in the Ed Sullivan Theater

[5]Image retrieved from https://www.jackmorton.com/work/the-late-show-with-stephen-colbert/

The data includes three distinct interaction types determined by the physical space and the position of cameras. The spatial arrangements of the interaction types are represented in Figure 2.8.

Figure 2.8: Differences in spatial organization across interaction types

The leftmost image shows the spatial organization of the show's opening monologue.[6] During this show segment, Colbert stands on the open stage (Area 1), directly in front of a camera and facing the live audience. During particular recurring segments, Colbert sits at his desk (Area 2), directly facing a camera and the live audience.[7] Interviews take place in the same region of the theater, with Colbert sitting at his desk (Area 2) and oriented toward his interviewees who sit across the desk (Area 3).[8] Both Colbert and his interviewees frequently shift their orientation between each other, the live audience, and the various cameras pointed toward them. This means that the interaction in interviews is highly dynamic as both primary interlocutors (Colbert and his guest) can engage independently in different spaces and with different addressees.

The spatial affordances of each interaction type impose different limitations on gesture performances. First, the differences effect the space in which gestures can occur relative to the speaker's body. Colbert's gesture space in the opening monologue is significantly larger than that in the other two interaction types. By standing, Colbert has access to a larger vertical space that is unimpeded by either his lap or furniture. Colbert can also freely move around the stage in ways that are not possible when seated. In the remaining two interaction types, the presence of the desk restricts the vertical dimensions of Colbert's gesture space – anything below chest level would not be visible to his interlocutors. Though the gesture space of the the interviewee is not impeded by the desk, their gesture space is limited by their seated position. Second, spatial differences also influence how gesturers position their hands when *not* gesturing. When Colbert's hands are 'at rest' in a monologue, his arms are extended loosely downward; his hands may be placed in his pockets or clasped in front of him. In the other two interaction types, Colbert's hands are at rest when laid on the desk. Interviewees' hands may be at rest in their lap, on the arms of the chair, or on the desk, depending on their bodily orientation. Finally, the spatial organization effects the relative

[6]Retrievable at UID:12908332-1d7b-11ea-903c-089e01ba0335,808
[7]Retrievable at UID:960813f4-11b1-11ea-bff4-089e01ba0335,911
[8]Retrievable at UID:6a02fce6-159f-11ea-8a36-089e01ba0335,1991

visibility of gestures. The visibility of Colbert's gestures is symmetrical in both monologue types, that is, movement of both hands is equally visible. This is not the case in interview settings, particularly when Colbert and his guest are facing each other. In this orientation, both interlocutor's gesture spaces are *asymmetrically* visible to the audience. Colbert's left arm is the most visible, while his right arm is partially obstructed by his body. The opposite holds for the interviewee.

Gesturers can also take advantage of the unique spatial affordances of each interaction type, as demonstrated in the gesture performances depicted in Figure 2.9.

Figure 2.9: Unique spatial affordances across interaction types

To the left, we see Colbert taking advantage of the space and his awareness of the camera's frame to "remove" an item from the show.[9] This is unique to the opening monologue in which the gesturer, Colbert, is standing and can move freely around an open space. In the center, we see Colbert pointing to a virtual screen to his upper right.[10] This screen does not exist in the physical space (compare to center screenshot in Figure 2.8). However, Colbert is aware of its virtual position and recurrently refers to it through deictic gestures. This is not possible in either of the other two interaction types because they lack equivalent screens (virtual or otherwise). To the right, we see British musician Thom Yorke taking advantage of the desk in front of him.[11] He performs a pushing gesture as he says "push him off the cliff". This is the only interaction type in which the edge of the desk is visible, and thus the only interaction type in which it can be taken advantage of in gesture production. The desk's surface can be used for spatial organization during both interviews and segment monologues.

In addition to differences in spatial affordances, each of the three interaction types differ in social affordances. Colbert has complete control over the discourse structure during monologues. The audience is the primary addressee and cannot contribute to the discourse structure directly (i.e. cannot 'take a turn' with an utterance that is added to the ongoing discourse). The audience can react to Colbert's performance and Colbert can attend to their reactions, though the coherence of the discourse does not depend on this interaction. During interviews, Colbert still asymmetrically controls the overall discourse structure by posing particular questions. However, the interviewee does take full turns and has some say

[9]Retrievable at UID:47198e46-d915-11e9-9fe6-089e01ba0335,227
[10]Retrievable at UID:c636d708-db70-11e9-ba03-089e01ba0335,1112
[11]Retrievable at UID:705c6f26-02c3-11ea-ba63-089e01ba0335,3360

over what is being talked about, how it is being talked about, and for how long it should be pursued. Social dynamics across interviews are also variable based on familiarity between Colbert and his guest, the relative social status of each participant, interviewee's gender, and age differentials. These social affordances must be considered in analyzing both verbal and gestural meaning.

2.3.2 Generalizability

All of the gesturers in my data set are professional performers in some way, appearing regularly under the scrutiny of the public eye as actors, comedians, journalists or politicians.[12] All of my data also occurs in a relatively 'unnatural' setting – on a stage with cameras and a live audience intruding upon the dialogue. Finally, all data involves interaction with the same person, the show's host Stephen Colbert. The reader may be concerned that these features of the data set pose serious problems for the generalizability of this work. This is a valid concern, and more work should be conducted on the differences between "performers" and non-performers, and between "performances" and unrehearsed dialogue. That being said, this concern should not distract from the contribution of this work and the framework it provides for analyzing multimodal interactive data. I offer two points to counter the potential risks to generalizability.

First, the overwhelming majority of gestures I observe and present are strikingly unexceptional. These are gestures that are well-documented as recurrent across speakers within and across cultures. My goal is not to identify 'new' gestural expressions. My goal is to reframe known gestural behaviors in light of a formal model of interactive meaning. For the few truly exceptional gestures in my data, I discuss them as such, and show the ways in which even these 'exceptional' idiosyncratic gestures resemble established patterns in both form and function. Second, this work does not seek to identify universals of expression. In fact, this work is in many ways adverse to such claims – optionality and variability of expression is *the point*. Instead, I am interested in developing a model that is useful in understanding gesture variation across communicative, social, and physical contexts.

There is no such thing as a perfect representative data set, regardless of whether it is gathered through naturalistic observations, experiments, or corpora. Naturalistic observations are limited in scope, focusing for practical and ethical reasons on only a few individuals. Experimental data is limited in both scope and validity. For practical and economic reasons, participants are generally few and from similar socioeconomic backgrounds.[13] Though the degree to which an experimental setting is "naturalistic" is highly variable, it always stands that experiment participants find themselves gesturing in a strange space and performing a task on demand with the knowledge of being observed. The limitations of corpus data, are,

[12]Given the availability of video data of "performers", such data is not unusual in established gesture literature (e.g. Calbris 2008; Casasanto & Jasmin 2010; Ferré 2012; Hart & Winter 2022; Hinnell 2019; Streeck 2008; Wehling 2017; Zima 2017).

[13]See Henrich et al. 2010 for a critique of how "WEIRD" (Western, educated, industrialized, rich and democratic) populations negatively impact the validity of experimental data in the social sciences.

of course, determined by the corpus itself. The challenge then is not to find a perfect data set, but to fully understand the limitations of the data set with which one works.

If anything, the data set used in this work may be considered unnatural in its relative 'cleanliness'. These particular gesturers are particularly good at what they are professionally trained to do, which is to efficiently and effectively communicate. Given the reputation of interactive gesture as particularly 'unruly' (Streeck 2009:181), such cleanliness is a strength, at least during a model's initial development. The strength of a theory does not depend directly on the particular data set used in its development. Instead, the strength of a theory lies in its capacity to account for different kinds of data. I hope that the proposed model of interactive meaning in gesture undergoes such assessment in future work, both by myself and others.

2.4 An integrated methodology for gesture description

In this work, I analyze gestural meaning in two stages. I frame these stages as distinct questions that require different methodologies to answer. The first question is simply *what does this gesture mean?*. When addressing this question, I consider gesture independent of accompanying verbal context, focusing on the gesture's physical form and how it relates to previous findings in the literature. This enables us to model gestural meaning in the terms and affordances of gesture itself. The second question is Bavelas' (1994) *"what does this gesture do and how?"*. Answering this question requires us to consider the gesture in context and model its integration with other semiotic systems, including the verbal and the social. This integrated two-step methodology is represented in Figure 2.10.

My process of gesture analysis begins by observing a gesture's physical form. This includes noting the gesture's "morphological features", the spatial arrangements of participants, and the kinetic context. This part of the analysis is arguably the only 'objective' part of the analysis – one is recording observable physical properties. In subsequent steps, the goal is maximal *intersubjectivity* – analysts should be able to reliably derive abstract gestural meaning from objective observable features.[14] In pursuit of intersubjectivity in my analyses, I borrow from Cognitive Linguistic theories of force dynamics (Talmy 1988) and image & action schema (Johnson 1987).[15] Once an action schema is hypothesized for a given gesture, I then consider previous work on related gestures, as well as the spatial and kinetic context in which the gesture is produced. As we will see, the gestural meaning derived in this way is highly schematic.

[14]The tension between objectivity and intersubjectivity in meaning-making is omnipresent in embodied approaches to cognition and, by extension, language. On the one hand, such approaches seek to ground subjective experience and social actions in bodily real-world phenomena. On the other, these experiences are only interpretable by the analyst through a view-pointed, and thus subjective, lens. See, for example, Violi (2008) for discussion.

[15]This is similar to the methodology advocated for in Mittelberg (2018).

The independent schematic meaning of the gesture is then integrated with the verbal, discourse structural, and social context to answer the second question – *what does this gesture do and how?*. To model this process I borrow from the Cognitive Linguistic theories of conceptual integration (Fauconnier & Turner 1998) and conceptual metaphor (Lakoff & Johnson 1980), as is advocated for in Parrill & Sweetser (2004).

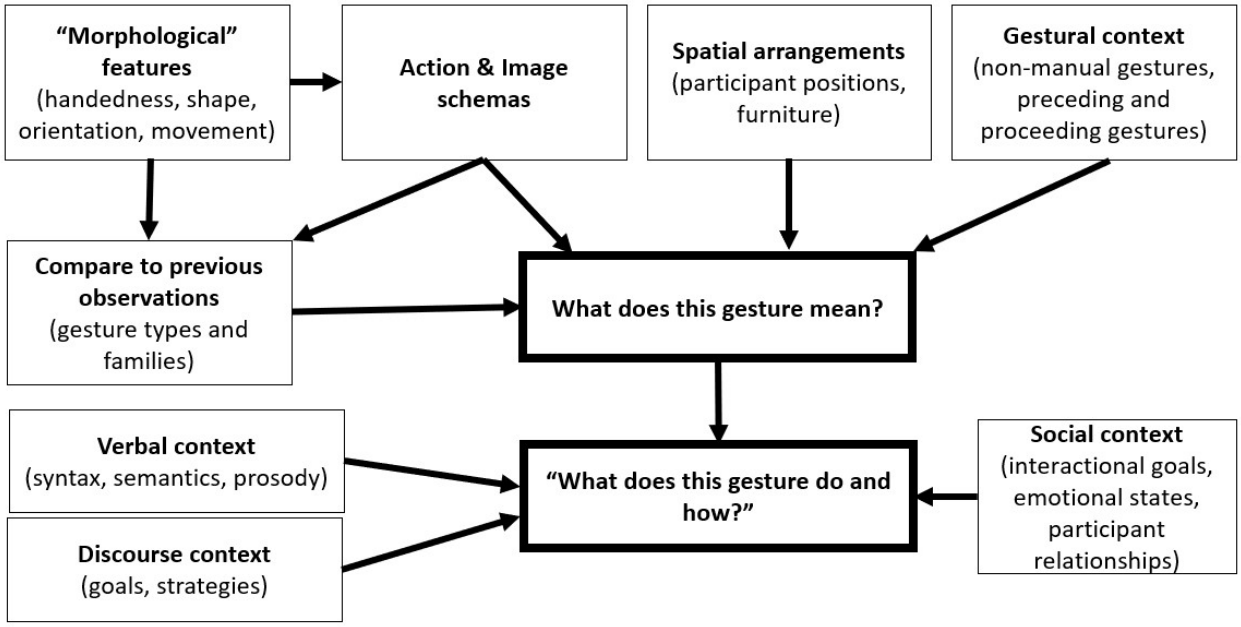


Figure 2.10: An integrated methodology for gesture description, inspired by proposed procedures in Bavelas (1994)

In the remainder of this section, I focus on the first question – *what does this gesture mean?*. The following section (Section 2.5) then outlines the strategies used for interpreting gesture meaning in context in order to answer the second question – *what does this gesture do and how?*.

2.4.1 Gesture segmentation

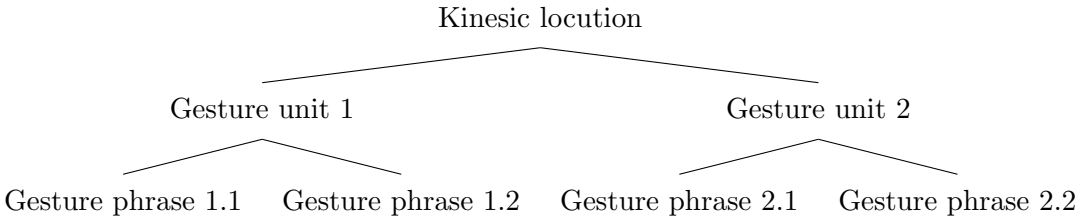
Just as words tend to flow together in speech, so too do gestures flow together in multimodal communication. Gestures, like words, are performed in a stream-like fashion such that the boundaries between one gesture and the next are sometimes difficult to discern. In order to model gesture meaning, we must first be able to divide this stream into meaningful, discrete units, i.e. individual gestures. I use Kendon's (1972; 1980) now standard notion of the "gesture phrase" to do so.[16]

Kendon (1972; 1980) proposes a hierarchical organization of gestural expression in face-to-face interaction. The largest and most abstract gestural unit is the 'kinesic locution', defined

[16]See Bressem & Ladewig (2011) for a frame-by-frame approach to gesutre phrase annotation.

as "stretches of body motion which share a distinctive movement or position" (McClave 2000:857). Within each kinesic locution is a set of 'gesture units' which are defined as 'excursions' from a resting position, i.e. all gestures that occur from when a speaker moves their hands from a resting position to when their hands return to a resting position. Each gesture unit consists of at least one 'gesture phrase', i.e. what we normally mean when we say *a gesture*. This is schematically represented in (12).

(12)



Each individual gesture also has an internal structure referred to as gesture 'phases'. An individual gesture minimally consists of a *stroke* phase which conveys the meaning of the gesture (McNeill 2005:32). Maximally, an individual gesture can involve articulatory onsets and offsets, called 'preparations' and 'retractions' respectively, and moments of stillness referred to as 'holds'.

(13) **Gesture phases**
(preparation) (pre-stroke hold) **stroke** (post-stroke hold) (retraction)

Initial annotation was conducted at the level of the gesture phrase for the stroke most closely aligned with the lexical discourse marker (*anyway, by the way*, or *here's the thing*). If no gesture stroke overlapped with the lexical discourse marker, gestures aligned with the attached clause were considered. Whether the annotated gesture was aligned with the lexical discourse marker, the clause it attached to, or both, was notated in the annotation file.

Quantitative findings in Chapters 5-7 are based on these gesture phrase annotations. However, much of my qualitative discussion is dedicated to assessing how informative these analysis are when the subject of study is gesture's contribution to maintaining coherence in ongoing discourse structure. These critical discussions look at gesture performance at the level of gesture units and kinesic locutions.

2.4.2 Gesture "morphology"

Once individual gestures are identified they are described for a set of physical features from which an action schema, and thus a meaning, can be derived. Many of the ways we describe the form of manual co-speech gestures have been borrowed from work on signed languages. Early work by Stokoe (1960) showed that signs contain sublexical structure, analogous to the phonemic structure of verbal words. This so-called phonology of (manual) signs consisted of handshape, location, movement, and, later on, orientation (Wilcox & Occhino 2017: 102-

104).[17] Because these formal features are meaningful in an action-schematic analysis, I will refer to them as *morphological* features rather than phonological features, following Fricke (2014).

(14) **Morphological features of signs and gestures**

a. *Orientation*: direction of the palm relative to the three axes of the gesturer's body (sagital, frontal, transverse)

b. *Hand shape*: configuration of the hand (e.g. curled fingers vs. flat palm)

c. *Movement*: trajectory and form of movement during stroke (e.g. lateral movement vs. repeated cyclic movements)

d. *Location*: position of the sign relative to personal (e.g. touching body vs. to side of body) or shared space

Palm orientation, hand shape, and *movement* were recorded for each annotated gesture, as well as *handedness* (i.e. whether the gesture is performed by one or two hands). Inventories of possible values for *movement* and *hand shape* features were developed with a team of undergraduate research assistants. Our goal was to develop a parsimonious and intuitive naming system that could account for the majority of recurrent gestures in our data sets. The development process was iterative, such that we would develop an inventory, use the inventory for annotation, and modify the inventory for clarity and breadth until we settled on those described in (16) and (17) below. *Location* was not recorded as a separate factor, but was described in the qualitative comments for each gesture.[18]

Each morphological feature is independent, such that any orientation can occur with any hand shape, with any kind of movement, and in any location. In this section, I provide examples of variation in each morphological feature. In Section 2.4.3, I discuss the ways in which action schemas are derived from particular constellations of morphological features. For palm orientation, there were six possible values, determined by the three axes of the body (sagital, frontal, transverse). A seventh "other" category was also used. These are described in (15).

(15) **Palm orientation**

a. *Palm up*: palm oriented upward (Figure 2.11, left)

b. *Palm down*: palm oriented downward (Figure 2.11, left-center)

[17] In sign language linguistics there are frequently well-defined inventories of hand shapes, often analogized to alphabets. Such sign 'alphabets' are occasionally used in discussions of gesture (e.g. McNeill 1992, 2005). This is not favored here as there is not a transparent connection between "alphabet" hand shapes and the action schemas they relate to.

[18] Unlike the other three morphological features, the *location* of a single gesture cannot be reduced to a single descriptor, especially where interactive meaning is under consideration. The location of a gesture conveys meaning not only relative to the speaker's body, but also the addressee's body, previous gesture locations, and objects (physical and metaphorical) in the shared space (e.g. Sweetser 2022; Sweetser & Sizemore 2008). The complexity and importance of gesture location to interactive meaning will be discussed at length in Chapter 4.

c. *Palm away*: palm oriented away from gesturer's body (Figure 2.11, right-center)
d. *Palm in*: palm oriented toward self (Figure 2.11, right)
e. *Palm center*: palm oriented toward midsaggital plane (Figure 2.14, G2)
f. *Palm periphery*: palm oriented away from midsaggital plane[19]
g. *Other*: no stable palm orientation or different orientations in two hands (Figure 2.12, G1)

Because palm orientation is a continuous variable, these are imperfect labels – the palm is rarely oriented perfectly in any of the six directions. Thus, for practical purposes, all *diagonal* orientations were binned as either *up* or *down*.

Figure 2.11 demonstrates the ways in which palm orientation can vary independently of hand shape.[20] Actor Jennifer Lawrence maintains roughly the same hand shape (palm exposed, fingers extended) while orienting her palm in different directions.

Figure 2.11: Palm orientations as performed by actor Jennifer Lawrence

Though palm orientation is most obvious in open-hand gestures where the palm is exposed, every gesture can be said to have an orientation. For example, the first two gestures in Figure 2.14 below have palm-center orientations.

Seven distinct hand shapes were used in annotation, described in (16) and depicted in Figure 2.12.[21]

(16) **Hand shape**

a. *Cup*: fingers partially extended and curved, palm partially exposed (Figure 2.12, G1)
b. *Fist*: fingers closed, palm not exposed (Figure 2.12, G2)
c. *Flat*: fingers fully extended, palm fully exposed (Figure 2.12, G3)

[19]No gestures were observed with this orientation.

[20]Figure 2.11 gestures retrievable at: 22e36ace-3709-11e8-9901-089e01ba0770,2058

[21]Figure 2.12 gestures retrievable at: G1: 47198e46-d915-11e9-9fe6-089e01ba0335,220; G2-5: 92fedfb8-1b1f-11ea-9518-089e01ba0335,2055; G6: 6af48f4c-ecba-11e9-bb9b-089e01ba0335,1945; G7: 6af48f4c-ecba-11e9-bb9b-089e01ba0335,2145

d. *Point*: one finger extended, others closed, palm not exposed (Figure 2.12, G4)[22]
e. *Loose*: fingers partially and variably extended, palm partially exposed (Figure 2.12, G5)
f. *Pinch*: fingers bunched, palm not exposed (Figure 2.12, G6)
g. *Ring*: index finger and thumb touching to form ring shape, remaining fingers partially or not extended, palm partially or not exposed (Figure 2.12, G7)

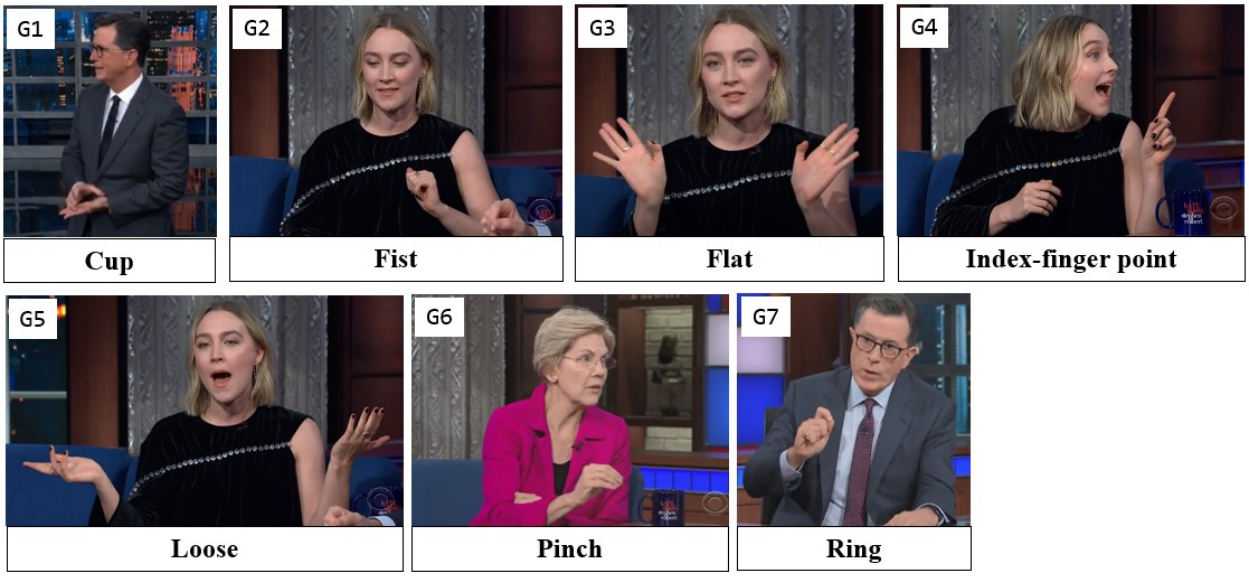


Figure 2.12: Annotated hand shapes with fixed forms

In addition to these seven fixed hand shapes, a gesture could be described as a *pantomime*, *rest*, or *other*.[23] *Pantomimes* are complex gestures depicting a particular activity such as using a touchscreen (Figure 2.13, G1) or texting (Figure 2.13, G2). *Rests* occur when the hands are at rest, such as in the gesturer's lap (Figure 2.13, G3) or on the desk (Figure 2.13, G4). *Other* was used as a catch all for behaviors that did not fit the other categories, such as self-adaptors like itching (Figure 2.13, G5) and object-oriented gestures (Figure 2.13, G6);

[22]Though the index-finger point was the most common in the data, one-finger points were also performed with the middle finger and the thumb

[23]Figure 2.13 gestures retrievable at: G1: fc6004ce-bb38-11e9-afb5-089e01ba0770,2189; G2: 13f49f6c-18c4-11ea-b152-089e01ba0335,1855; G3: c5b60a9c-fabb-11e8-ba14-089e01ba0335,1868; G4: f00485b2-051e-11ea-8f95-089e01ba0770,3559; G5: e5b2d83e-fd3a-11e9-aee3-089e01ba0335,2761; G6: a771d272-b00c-11e8-82c3-089e01ba0770,1630

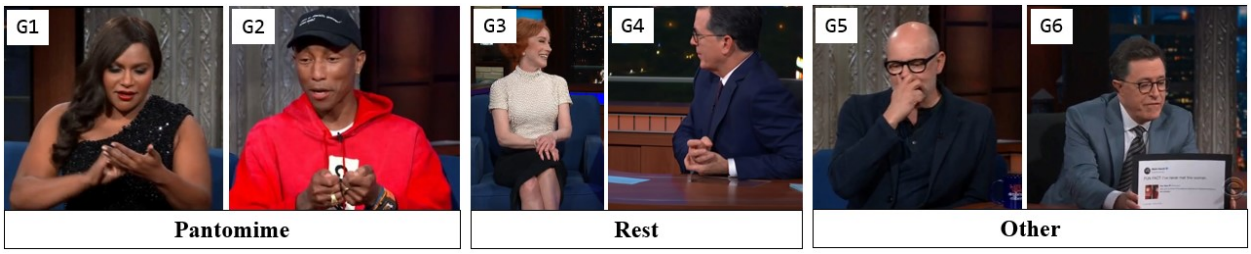


Figure 2.13: Annotated hand shapes without fixed forms

Ten movement types were described in the annotations, described in (17) and depicted in Figure 2.14.[24]

(17) **Movement types**

a. *Beat*: small repeated vertical movements (Figure 2.14, G1)

b. *Chop*: downward movement (Figure 2.14, G2)

c. *Clear*: lateral movement away from gesturer's body, flexion at elbow or shoulder (Figure 2.14, G3)

d. *Cylic*: repeated circular movements (Figure 2.14, G4)

e. *Flick*: lateral movement away from gesturer's body, flexion at wrist (Figure 2.14, G5)

f. *Push*: saggital movement away from gesturer's body (Figure 2.14, G6)

g. *Reach*: movement toward object or location (Figure 2.14, G7)

h. *Rise*: upward movement (Figure 2.14, G8)

i. *Rotate*: change in palm orientation (Figure 2.14, G9)

j. *Sweep*: lateral movement not categorized as "clear" (Figure 2.14, G10)

[24]Figure 2.14 gestures retrievable at: G1: c1d66448-e995-11e9-9e68-089e01ba0770,2771; G2: c7cce1dc-ff9e-11e9-9f3f-089e01ba0335,725; G3: 67f2c08a-f628-11e9-920c-089e01ba0335,3369; G4: 92fedfb8-1b1f-11ea-9518-089e01ba0335,2055; G5: 2d0e822a-a46f-11e9-80c4-089e01ba0335,1065; G6: 6a02fce6-159f-11ea-8a36-089e01ba0335,2870; G7: f1dc45d4-0067-11ea-ac64-089e01ba0335,1883; G8: a46f90b2-d40e-11e7-9367-089e01ba0335,2141; G9: a771d272-b00c-11e8-82c3-089e01ba0770,1635; G10: c1d66448-e995-11e9-9e68-089e01ba0770,2774

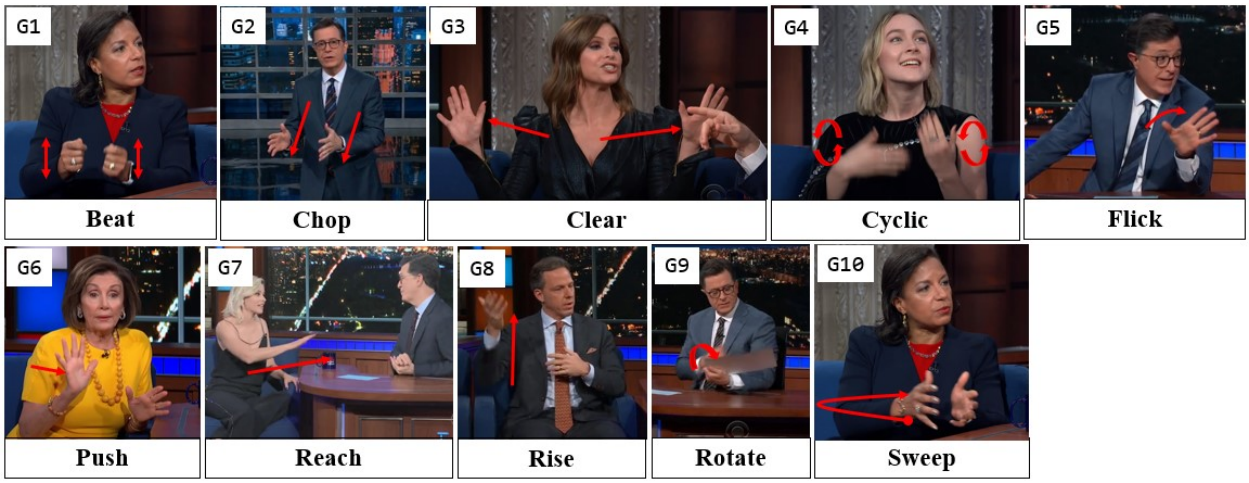


Figure 2.14: Annotated hand movements with fixed forms

Because gestural movement is extremely variable, many gestures were described simply as "other". Complex movements such as pantomimes (Figure 2.13, G1 & G2) or movements with multiple trajectories (Figure 2.15, G1) were categorized in this way, as well as emblematic gestures (Figure 2.15, G2) and social performances like clapping (Figure 2.15, G3).[25]

Figure 2.15: Examples of hand movements annotated as "other"

Describing the morphological features of a gesture does not give us the meaning of a gesture. Meaning is derived by connecting these features to particular actions and images. For example, the palm-away orientation does not mean 'stop'. Instead, the palm-away orientation is consistent with a stopping action in which one holds an outward-facing open hand to stop the incoming movement of an object. In other words, it is the action *indicated by* the palm-orientation that means 'stop', not the palm-orientation itself. As such, these features are not used directly in quantitative analysis in this work. However, recognizing these features is a necessary step in deriving gestural meaning. Only in doing so can we principally and reliably determine a gesture's action or image schema.

[25]Figure 2.15 gestures retrievable at: G1: 14480e4e-0f2a-11e9-bafc-089e01ba0335,2009; G2: 6af48f4c-ecba-11e9-bb9b-089e01ba0335,2140; G3: c5b60a9c-fabb-11e8-ba14-089e01ba0335,1854

2.4.3 Action and image schemas

As already discussed in Chapter 1, I consider gesture meaning to be primarily grounded in iconicity – a gesture's formal features evoke particular action and image schema by *resembling* particular everyday actions and images (e.g. Cienki 2013; Mittelberg 2018, 2019; Müller 2017). The schema evoked constitutes the gesture's meaning, independent of accompanying speech. This core meaning is *iconic*, based on a literal resemblance between the physical properties of the hand gesture and the physical properties of an action. The gesture's *function* is (usually) metaphoric, derived by mapping the core action schema to a metaphoric target domain. This section focuses on identifying the *iconic core meaning* of the gesture. I will turn to identifying *metaphoric functional meaning* in the following section.

In this section, I give an overview of the five classes of action schema that play significant roles in this work: PRESENTATION, REMOVAL, STOPPING, REFERRING and REORIENTATION. Basic descriptions of these five classes and their respective sub-types are summarized in Table 2.1. For each class, I discuss the relationship between the action schema and morphological features, as well as connections to recurrent gestures in existing literature (Fricke et al. 2014; Ladewig 2014b; Müller 2017). Action schematic meaning is represented as semantic *frames* composed of the participants, entities, and events involved in taking the relevant action (Fillmore 1976).

Type	Description	Section
Presentation	hand gestures that enact the presentation of an object	*2.4.3.1*
Containment	two-handed gesture with palm-center orientation, as if to hold a rectangular object	
Precision Grip	one-handed gesture with fingers bunched, as if to hold a small delicate object	
Palm-Up Open-Hand	one- or two-handed gesture with palm-up orientation and palm exposed, as if to hold an object up for inspection	
Removal	hand gestures that enact the removal of an object from gesturer's immediate body space	*2.4.3.2*
Clearing	one- or two-handed gesture with fingers extended and palm exposed, moved laterally as if to clear a surface	
Throwing	one-handed gesture with fingers loosely extended, movement away from body, as if to throw an object	
Flicking	one-handed gesture, loose hand shape, small quick movement away from body, as if remove a small object	
Stopping	hand gestures that enact the stopping of some object from from entering the gesturer's immediate body space	*2.4.3.3*
Blocking	one- or two- handed gesture, palms exposed and oriented away, as if to hold an object away from the body	
Pausing	one- or two- handed gestures with index finger upward, as if to indicate the numeral "1"	
Referring	hand gestures used to direct joint attention, and lacking the physical affordances of presentation	*2.4.3.4*
Addressing	deictic gesture toward addressee	
Locating	deictic gesture toward location other than addressee	
Reorientation	change in orientation of gaze, head, or body	*2.4.3.5*
Engagement	orientation toward the speaker-hearer line	
Disengagement	orientation away from speaker-hearer line	

Table 2.1: Primary action schemas

2.4.3.1 Presentational gestures

Presentational gestures are those gestures that exhibit the physical affordances of presenting an object for inspection. These gestures are both iconic in that they enact the presentation of an object, and deictic in that the object is introduced into a particular location relative to the position and orientation of both interlocutors.[26] Both the iconic and deictic aspects are necessary – holding an object into a space where it cannot be inspected fails to evoke the PRESENTATION action schema.[27]

In this data set, I identify three types of presentational gesture, *palm-up open-hand* (PUOH), *containment*, and *precision grip*, as depicted in Figure 2.16.[28] Though all three evoke the PRESENTATION action schema, their morphological features indicate different physical properties of the presented object.

Figure 2.16: Forms of presentational gestures

The one-handed PUOH gesture is compatible with the presentation of a small to medium-sized, light-weight object that can be supported by an open hand. Two-handed variants are compatible with supporting a larger, heavier object for inspection. Containment gestures are compatible with the presentation of a medium sized, rectangular object, such as a box, that may contain additional objects. The precision grip gesture is compatible with the presentation of a small, delicate object that requires careful inspection. This interpretation of the action schema is supported by the non-manual gestures that frequently accompany it, such as the lean in depicted in Figure 2.16, G3 – the gesturer brings their head closer to the presented object, emphasizing the need for careful inspection.

The physical affordances of each type of presentational gesture are represented by the the semantic frames in Table 2.2. Much of the frames are the same – there is always a presenter

[26]See Kok et al. (2016), McNeill (2005), and Sweetser (2022) for discussion of multifunctionality in gesture.

[27]As is typical for Cognitive Linguistic frameworks, I use SMALL CAPS to denote abstract concepts. This includes the terms for action schemas, image schemas, semantic frames, and metaphoric source and target domains.

[28]Figure 2.16 gestures retrievable at: G1: 97a9a7ce-0cfa-11ea-8374-089e01ba0770,2909, G2: c7cce1dc-ff9e-11e9-9f3f-089e01ba0335,728, G3: 6a02fce6-159f-11ea-8a36-089e01ba0335,1037

and an observer, always an event of presenting, and always the result of an object being perceived. These aspects of the frame are profiled primarily by the movement and trajectory of the gesture – something is brought into the shared space between interlocutors for mutual inspection. Properties of the presented object are primarily profiled by the gestures' hand shape.

PALM UP OPEN HAND	CONTAINMENT	PRECISION GRIP
Participant: presenter	Participant: presenter	Participant: presenter
Participant: observer	Participant: observer	Participant: observer
Entity: presented object	Entity: presented object	Entity: presented object
Properties: medium sized	Properties: variably sized	Properties: small
indeterminate shape	rectangular shape	possibly delicate
	flat edges	difficult to see
	can contain other objects	
Event: presentation of object	Event: presentation of object	Event: presentation of object
Result: object is perceived	Result: object is perceived	Result: object is perceived

Table 2.2: Frame structure of presentational gestures

The PUOH gesture is well-known for its pervasiveness across use-contexts and cultures (Cooperrider et al. 2018; Ferré 2012; Kendon 2004; McKee & Wallingford 2011; Müller 2004). The titular orientation and hand shape has also been shown to compose with other gestural features to evoke different action and image schema, such as in CYCLIC gestures (Ladewig 2011, 2014a), shrugs (Debras 2017; Marrese et al. 2021), and gestures used to express contrast (Hinnell 2019). The extent to which the PUOH gesture is used and its diversity of use contexts may be related to the fact that it evokes a relatively generic PRESENTATION schema – unlike the other two types, it does not convey information about the presented objects shape or delicacy. The gesture is also fundamentally *ambiguous* because its formal features are consistent with two distinct action schemas – presenting and requesting an object (Müller 2004).

The precision grip gesture has received some attention for its use in emphasizing important points, especially as a recurrent gesture in Italian discourse (Kendon 1995, 2004) and political debates (Streeck 2008).[29] Though containment gestures do appear in work on interactive gesture (e.g. Mittelberg 2017), I am not aware of work that explores the communicative capacities of containment gestures in detail.

The PRESENTATION action schema appears in all three case studies in this work (Chapters 5-7). However, the importance of distinguishing the three types is most apparent in Chapter 7 on specification.

[29]See also Müller (2014) for a review of the use of the ring hand shape cross-culturally.

2.4.3.2 Removal gestures

Removal gestures are those gestures that exhibit the physical affordances of removing an object from the speaker's immediate space. As with presentational gestures, these gestures are both metaphoric iconic in that they enact the moving of a metaphoric object, and deictic in that the trajectory is always *away* from the speaker. Also as before, both the iconic and deictic components are necessary for evoking the REMOVAL action schema – there must be an object, and that object must be moved in a particular direction relative to the interaction.

There is significant formal variation in the removal gestures identified in the present data set, especially regarding palm orientation. In Figure 2.17, this formal variation is mapped to three types of removal gesture – *clearing, throwing* and *flicking*.[30] Though all three evoke a REMOVAL action schema, they differ in the types of objects that are removed and manner of removal.

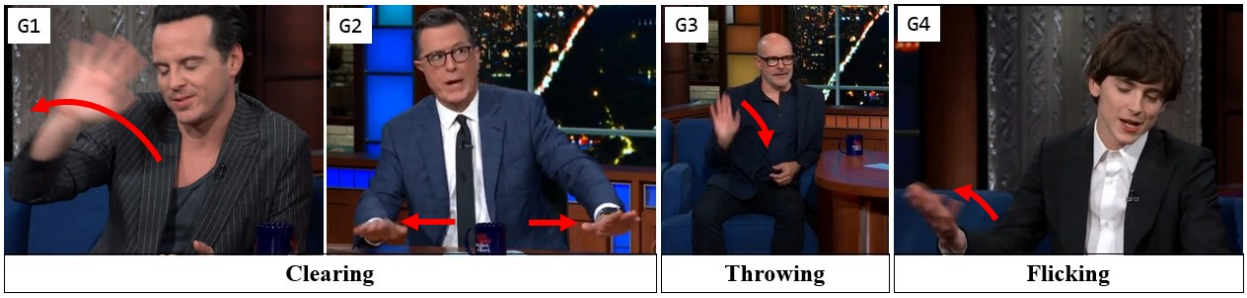


Figure 2.17: Forms of removal gestures

Clearing gestures are compatible with the removal of objects from a flat surface. The flat hand shape and outward lateral movement profile the clearing action itself, whereas the orientation of the hand profiles the orientation of the metaphoric surface being cleared. Palm-away clearing gestures enact the clearing of a vertical surface, as if to wipe steam off a window. Palm-down clearing gestures enact the clearing of a horizontal surface, as if to clear a table of crumbs. Throwing gestures are compatible with the removal of a small object, as if to throw something into a waste bin. The loose hand shape profiles the holding of a metaphoric object, and the outward trajectory, with downward flexion at the wrist, profiles the act of throwing. Flicking gestures involve small, quick lateral movements away from the gesturer, with flexion at the wrist, as if to swat away a fly. Hand shape is variable, but generally loose.

The physical affordances of the three types of removal gestures are represented by the semantic frames in Table 2.3. As with presentational gestures, much of the frame is the same across types – there is always a remover and an observer, the event of clearing objects from

[30]Gestures retrievable at: G1: f00485b2-051e-11ea-8f95-089e01ba0770,3559, G2: 135ad8d4-f496-11e9-9c8d-089e01ba0770,1230, G3: e5b2d83e-fd3a-11e9-aee3-089e01ba0335,2766, G4: 4b3223f4-ea3b-11e8-985b-089e01ba0770,1858

the immediate space, and the result of those objects no longer being accessible to either participant. The trajectory of movement is the most important morphological feature in evoking the REMOVAL action schema – for objects to be "removed", the trajectory of the gesture must be away from the speakers body. The palm orientation and hand shape profile information about the object removed and the manner of removal.

CLEAR	THROW	FLICK
Participant: remover	Participant: remover	Participant: remover
Participant: observer	Participant: observer	Participant: observer
Location: flat surface		
Entity: removed object(s)	Entity: removed object	Entity: removed object
Properties: possibly numerous indeterminate size	Properties: medium sized	Properties: small light-weight
Event: clearing of surface	Event: removal of object	Event: removal of object
Result: surface is cleared of objects	Result: object is no longer accessible	Result: object is no longer accessible

Table 2.3: Frame structure of removal gestures

REMOVAL gestures are well documented in the literature across cultures and use-contexts, most notably in work on the so-called 'away' gesture family (Bressem & Müller 2014, 2017, and have been associated with negative assessment. Clearing gestures, especially those with a palm-down orientation, have been associated specifically with negation and cancellation of inferences (Harrison 2010; Kendon 2004). I favor the term 'removal' here over 'away' because 'removal' conveys the action schema that is evoked, and therefore the meaning of the gesture, more transparently than 'away' does.

Removal gestures are observed in the case studies of topic-shifting (Chapter 5) and digression (Chapter 6). However, the full range of Removal gestures was only observed for topic-shifting. Removal gestures in digressions were limited almost entirely to palm-away clearing gestures.

2.4.3.3 Stopping gestures

Stopping gestures are gestures that exhibit the physical affordances of stopping an object from moving closer to the speaker's body. As before, stopping gestures are both iconic and deictic. The defining features, a palm-away orientation without lateral movement, are compatible with stopping a medium or large object from getting closer to the gesturer's body, evoking the STOP action schema. Forward sagittal movement is also sometimes observed, which is compatible with pushing an object further away from the gesturer's body.

I identify two types of stopping gestures in the data, *blocking* and *pausing*, depicted in Figure 2.18.[31]

[31]Figure 2.18 gestures retrievable at: G1: d129e820-ba6f-11e9-a245-089e01ba0335,2886, G2: afbec190-

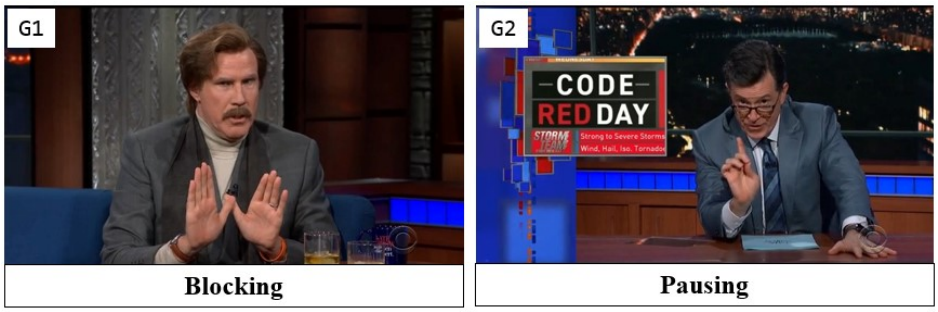


Figure 2.18: Forms of stopping gestures

Blocking gestures can be performed with either one or two hands and enact the stopping of a medium or large object. Pausing gestures combine the palm-away orientation with an upward index-finger point, as if to ask for someone to "wait one second". This form is also compatible with stopping a very small object, one that can be stopped with just one finger, from getting closer to gesturer's body. Non-manual gestures, including eyebrow furrowing and leans forward, frequently accompany pausing gestures. Blocking gestures more frequently occur with leans backward, as if to move even further away from the stopped object.

The two variants of the STOP action schema expressed by *blocking* and *pausing* gestures are represented as semantic frames in Table 2.4. In both, the participants, event, and result remain the same and constitute the more generic STOP schema. The differences in hand shape profile differences in the physical properties of the stopped object.

BLOCK	PAUSE
Participant: stopper	Participant: stopper
Participant: observer	Participant: observer
Entity: stopped object	Entity: stopped object
Properties: medium or large	Properties: singular and small
possibly heavy	
Event: stopping object	Event: stopping object or action
Result: object cannot move closer	Result: object cannot move closer

Table 2.4: Frame structure of stopping gestures

These gestures have been observed in political argumentation (Streeck 2008; Wehling 2017) and in discussions of the 'away' gesture family, in which they are typically referred to as 'holding away' gestures (Bressem & Müller 2014, 2017; Bressem & Wegener 2021). I consider stopping gestures to be fundamentally different in meaning from other 'away' gestures simply because they evoke different action schemas. Stopping gestures do not enact

9d5c-11e9-9e5f-089e01ba0770,1190

the removal of an object, and so the result of the action is very different – the metaphoric object is still present and accessible to interlocutors.

Stopping gestures are observed recurrently in cases of digression (Chapter 6) and specification (Chapter 7), though pausing gestures are significantly more frequent in specifications. Stopping gestures are also observed infrequently in cases of topic-shifting but are limited to particular use-contexts (Chapter 5).

2.4.3.4 Referring gestures

Referring gestures are those gestures that deictically direct attention toward a location in shared space but lack the physical affordances of presentation. These gestures are primarily deictic in that they are oriented toward a particular region of space. They do not enact an action, aside from directing jointing attention, or depict an image. However, they do serve to reorganize the conceptualization of space around the pointer, who serves as the deictic center, or 'origo' (Bühler 1982).

I identify two types of referring gestures in the data: *addressing* gestures which direct attention to the addressee, and *locating* gestures which direct attention toward some other region of space. These two variants are depicted in Figure 2.19.[32] Referring gestures in general do not have the physical affordances that are compatible with the manipulation of objects. However, there are particular instances in which the morphological features of a gesture may evoke an object-oriented action schema. For example, the locating gesture in Figure 2.19 is compatible with holding an object down on the desk. Hand shape is variable, with both one-finger points and open-hand forms occurring frequently in the data.[33]

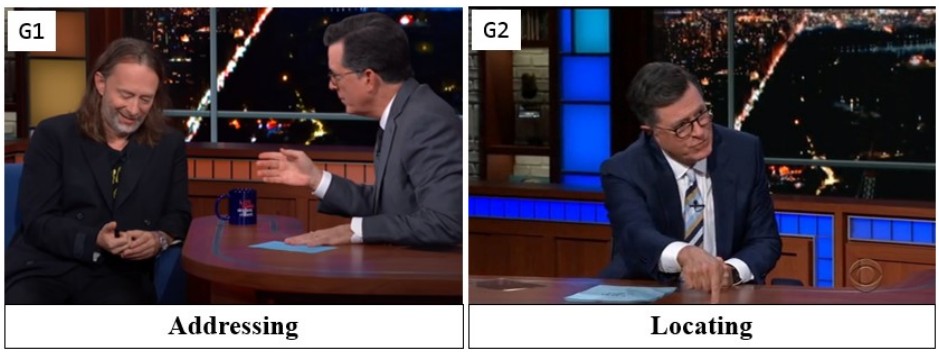


Figure 2.19: Forms of referring gestures

The two classes of referring gestures are distinguished by the trajectory of the deictic gesture. Both classes can occur with open-hand or one-finger hand shapes, though the open-hand shape seems to be preferred in addressing gestures. In some contexts, there is

[32]Figure 2.19 gestures retrievable at: G1: 705c6f26-02c3-11ea-ba63-089e01ba0335,3329, G2: d6941018-a793-11e9-902d-089e01ba0335,1273

[33]See Kita (2003) for a comprehensive review of variation in pointing gestures.

ambiguity between locating and addressing gestures, such as when the topic being referred to pertains to information about the interlocutor. However, this is a functional rather than formal distinction and will be discussed as such in Chapter 4.

The two variants of the REFER action schema, as expressed by *addressing* and *locating* gestures, are represented as semantic frames in Table 2.5. In both, the participants, event, and result remain the same and constitute the more generic REFER schema. Differences in trajectory determine the the target of joint attention and the presence of an object.

ADDRESS	LOCATE
Participant: referee	Participant: referee
Participant: addressee	Participant: observer
	Entity: object referred to
Event: direct attention toward addressee	Event: direct attention toward object
Result: attention is given to addressee	Result: attention is given to object

Table 2.5: Frame structure of stopping gestures

Addressing gestures are well-documented in interactive gesture and have been observed to perform a range of discursive functions, including citing addressee contributions, eliciting backchannels, and signalling turn-transitions (Bavelas et al. 1992; Bavelas 1994; Mondada 2007; Trujillo & Holler 2021). Locating gestures have also been discussed in the interactive gesture literature, specifically in terms of 'abstract deixis' (McNeill et al. 1993; McNeill 2003). In this framing, pointing gestures are used to organize and negotiate the location of topics and referents as metaphoric objects in space.

Referring gestures occur in all three case studies (Chapters 5-7). However, locating gestures are restricted to particular use-contexts in topic-shifting (Chapter 5).

2.4.3.5 Action schemas in non-manual gestures

Though the majority of my attention is paid to the hand gestures described above, non-manual shifts in orientation also play an important role in orchestrating conversation and often reinforce the actions performed by concurrent hand gestures. In particular, I focus on the enactment of *engagement* – the orientation of gaze, head, and body relative to the space shared by co-present interlocutors. These non-manual gestures evoke very basic action schemas of PARTICIPATION. In order to participate in an activity, one must orient themselves to the space in which the activity is taking place. I refer to this as an ENGAGE action schema. Orienting away from the space of an activity evokes a DISENGAGE action schema, signalling a voluntary exclusion from participation.

Three sets of ENGAGE and DISENGAGE actions are depicted in Figure 2.20.[34] In all three cases, we see the interview guest reorient their gaze, head, and body from *engaging* with

[34]Figure 2.20 gestures retrievable at: G1-G2: c5b60a9c-fabb-11e8-ba14-089e01ba0335,1854, G3-G4: bf03ea3c-1731-11ea-a000-089e01ba0335,2088, G5-G6: fc6004ce-bb38-11e9-afb5-089e01ba0770,2182

Colbert (G1, G3, G5) to *disengaging* from Colbert (G2, G4, G6). The G5-G6 pairing is particularly interesting as Colbert increases his engagement with his guest, leaning toward her across the table, while she turns away to engage with the audience.

Figure 2.20: Forms of reorientation

The two variants of reorientation actions, *engagement* and *disengagement*, are represented as semantic frames in Table 2.6. In both, the participants and location remain the same and determine the space of engagement. The events and results are directly opposed. The ENGAGE schema involves engagement in the shared space, resulting in participation in the interaction, whereas the DISENGAGE schema involves disengagement from the shared space, resulting in the cessation of participation.

ENGAGE	DISENGAGE
Participant: agent Participant: observer Location: space between agent & observer Event: engage in shared space Result: agent participates in the interaction	Participant: agent Participant: observer Location: space between agent & observer Event: disengage from shared space Result: agent ceases participation in the interaction

Table 2.6: Frame structure of engagement actions

The functions of bodily orientation in face-to-face interactions is well-studied in the disciple of proxemics (e.g. Gill et al. 2000; Hagemann 2014; Hall 1995; Kendon 1967, 1990, 2010). Head orientation in particular is argued to be derived from the basic action schema of turning one's head away from unwanted sensory input (Harrison 2014:1498).

The functions of engagement and disengagement actions are discussed for particular examples in all three case studies (Chapters 5-7). I also discuss them in detail in Chapter 4, in which I develop a model for interactive meaning and multimodal discourse management.

2.5 Interpretation in context

The previous section was dedicated to deriving gesture meaning independent of the context in which the gesture occurs. However, similar to many lexical items, the independent meaning of a gesture is highly schematic. To understand what a given gesture is *doing*, i.e. what it is actually helping to communicate, verbal, discursive, and social context must be considered.

To derive contextualized meaning from the action-schematic meaning of gesture itself, I rely largely on the cognitive linguistic theories of frame semantics (Fillmore 1976) and conceptual metaphor (Lakoff & Johnson 1980), both of which relate abstract meaning to our embodied real-world experiences. This approach is similar to that advocated for in Parrill & Sweetser (2004).[35]

The action schema evoked by the gesture serves as the SOURCE frame which is mapped to a TARGET frame that must be identified in context. Because I am concerned with interactive meaning and the use of gesture in contributing to discourse structure, the TARGET frame in all cases can be considered aspects of the discourse itself (Wehling 2017). The basic composition of the DISCOURSE frame includes the discourse participants, the topics being discussed, and the processes of introducing, managing, and ending discourse topics. The source frame in all cases is PHYSICAL INTERACTION, specifically joint physical interactions involving object-oriented actions. The basic metaphoric mappings between these two frames are given in Table 2.7.

[35]See also Stickles (2016) for a similar approach specifically within a Construction Grammar framework.

DISCOURSE		PHYSICAL INTERACTION
Participant: interlocutor	$\Longleftarrow$	Participant: agent
Participant: interlocutor	$\Longleftarrow$	Participant: agent
Entity: topic	$\Longleftarrow$	Entity: object
Event: topic management	$\Longleftarrow$	Event: object management
Event: contributing to discourse	$\Longleftarrow$	Event: manipulating object in shared space

Table 2.7: Basic metaphoric mapping of physical interaction to discourse

This metaphor relies on the primary metaphor IDEAS ARE OBJECTS in which all abstract concepts can be conceptualized as physical manipulable objects (Lakoff & Johnson 1980). In the DISCOURSE frame, the relevant "ideas" being conceptualized as metaphoric objects are discourse topics.

Each of the five classes of action schema discussed in the previous section can be mapped to the DISCOURSE target frame to profile different types of discourse management actions. I will demonstrate this by considering the metaphoric mappings evoked by the PRESENTATION, REFERRING, and REMOVAL action schemas.

In order to participate in a coherent and cooperative discourse, discourse participants must agree on what they are talking about. The topic of conversation can change throughout the discourse, but for coherence to be maintained, each topic must be established and accepted by both participants. The process of introducing a topic for discussion can be metaphorically understood as presenting an object for inspection. This is represented by the frame-based metaphor mappings in Table 2.8.

TOPIC INTRODUCTION		PRESENTATION
Participant: speaker	$\Longleftarrow$	Participant: presenter
Participant: addressee	$\Longleftarrow$	Participant: observer
Entity: topic	$\Longleftarrow$	Entity: presented object
Event: introduction of topic	$\Longleftarrow$	Event: presentation of object
Result: topic is accepted	$\Longleftarrow$	Result: object is perceived

Table 2.8: Metaphoric mapping of presentational gesture to discourse

The topic itself is conceptualized as a metaphoric object that can be perceived and manipulated by both participants once it is present in the shared space.[36] The form of presentational gesture can provide more specific mappings as to the properties of the topic

[36]The mapping between TOPIC and OBJECT is an extension of the primary metaphor IDEAS ARE OBJECTS (Lakoff & Johnson 1980). This primary metaphor plays an important and pervasive role in our understanding of discourse structure, as demonstrated in work on the so-called 'Conduit Metaphor' (Reddy 1979). I discuss the use of this metaphor in discourse analysis in Chapter 3.

as a metaphoric object. For example, when the CONTAINMENT action schema is evoked, the topic can be conceptualized as a container containing arguments that requires unpacking.

As a discourse progresses, participants make contributions to established discourse topics. The topics, as metaphoric objects, can be located in particular regions of the shared space and can be deictically REFERRED to via *locating* and *addressing* gestures. Doing so draws attention to the particular object in order to, for example, specify or request additional information. The metaphoric mapping between the REFER action schema and the discourse process of TOPIC CONTRIBUTION is represented in Table 2.9.

TOPIC CONTRIBUTION		REFER
Participant: speaker	$\Longleftarrow$	Participant: referrer
Participant: addressee	$\Longleftarrow$	Participant: observer
Entity: topic	$\Longleftarrow$	Entity: object referred to
Event: attend to topic	$\Longleftarrow$	Event: direct attention toward object
Result: topic is contributed to	$\Longleftarrow$	Result: object is attended to

Table 2.9: Metaphoric mapping of referring gesture to discourse

More complex mappings can be evoked when multiple topics are metaphorically present in the shared space. For example, topics can be contrasted by locating them in different regions of the shared space.[37]

Participants can also decide to stop talking about a topic in favor of a new topic or in order to end the interaction altogether. For coherence to be maintained, participants must mutually agree to end the topic. The process of dismissing a discourse topic can be metaphorically understood as removing the topic as a metaphoric object from the shared space, thus preventing further interaction with it. The metaphoric mapping between the REMOVAL action schema and the process of TOPIC DISMISSAL is represented in Table 2.10.

TOPIC DISMISSAL		REMOVAL
Participant: speaker	$\Longleftarrow$	Participant: remover
Participant: addressee	$\Longleftarrow$	Participant: observer
Entity: topic	$\Longleftarrow$	Entity: removed object
Event: ending of topic	$\Longleftarrow$	Event: removal of object
Result: topic is no longer pursued	$\Longleftarrow$	Result: object is no longer accessible

Table 2.10: Metaphoric mapping of removal gesture to discourse

More specific metaphoric mappings addressing the topic's dismissal and speakers' attitudes about the dismissed topic can be evoked through particular formal features of the

[37]See Hinnell (2019) and Laparle (2022) for discussion.

removal gesture. For example, when the CLEAR action schema is evoked by a clearing gesture, the shared space is metaphorically cleared of all objects – a completely new topic can begin. This can be contrasted with the FLICK action schema which evokes the removal of a relatively small insignificant object. This can evoke a mapping to the dismissal of a particular detail or subtopic, rather than an ending of all ongoing topics.

Determining what metaphor is evoked by a given interactive gesture requires us to consider what is happening in the discourse at the time of the gesture's performance. Though the gesture evokes an action schema independently of this context, we need to consider the context when analyzing exactly how the evoked action schema is being used to express discourse management.

2.6 Applying the methodologies

After providing an overview of my approach to discourse structure in Chapter 3, I use the methodologies described above to develop a formal model of interactive meaning in gesture (Chapter 4) and to analyze the use of interactive gesture in particular discourse structural contexts (Chapters 5-7). The formal model I propose centers around the conception of a physical shared space, which I call the *Interaction Space*, as a space in which to construct a discourse structure through the management of metaphoric objects. The action schemas described in this chapter are used as an inventory of discourse actions that act upon this space. I then use this model to analyze the regularity with which interactive gestures are used to perform particular kinds of discourse management, namely *topic-shifting, digression* and *specification*.

Chapter 3

Constructing a discourse

3.1 Introduction

In understanding the role of gesture in managing discourse structure, we must act as both gesture analysts and discourse structure analysts. The previous chapter outlined my approach as a gesture analyst. This chapter outlines the assumptions I make and framework I use for interpreting discourse structure. The purpose of this chapter is not to provide an overview of discourse analysis as such. Instead, its purpose is to articulate how and why I arrive at particular discourse structure analyses by highlighting particularly relevant work on cooperative discourse and hierarchical discourse structure.

I begin by distinguishing two important aspects of a discourse's structure, *information* structure and *interaction* structure, and how they relate to different communicative pressures in conversation (Section 3.2). Section 3.3 outlines the question-based approach to information structure used in this work. Section 3.4 reflects on the pervasiveness of metaphor in discussions of discourse structure, and how these metaphors relate to the action schemas discussed in Chapter 2. Section 3.5 concludes.

3.2 Communicative pressures in face-to-face conversation

In conversation, interlocutors are under pressure to meet both informative and social demands. In order to effectively convey information about our experiences in the world, we have to remain engaged in the conversation. This means that there are two main concerns that a discourse analyst must contend with when developing a theory of discourse structure. First is the structure of *information*, i.e. how information is communicated in such a way that it is understood by discourse participants. Second is the structure of the *interaction*, that is how participants orchestrate their contributions throughout a conversation. A comprehensive model of discourse structure is one that accounts for both sets of demands and

provides a framework for determining how they are negotiated as a discourse unfolds.

As Sacks et al. (1978) observe, interlocutors rarely enter into conversation with a shared plan, even when general communicative goals (e.g. giving directions) are known from the start. We don't know exactly who will say what, or when they will say it. This means that the negotiation between the informative and social demands of an interaction happen in real time, as the conversation is taking place. An analyst's job is to understand what makes this possible. What strategies do we have at our disposal to ensure effective communication? What rules of interaction do we follow, if any? To address these two questions, I will consider the 'grammatical' dimensions of interaction (3.2.1) and principles of cooperation and politeness (3.2.2).

3.2.1 Interaction-based grammar

Traditional generative approaches to grammar consider language structure from a rule-based perspective – language users must abide by certain rules to accomplish communicative goals (Matthiessen & Halliday 2009). Functional grammars flip this around, taking a resource-based perspective – language users have a set of tools to help achieve communicative goals (*ibid*). We need both, but let's consider the resource-based perspective first.

Systemic Functional Grammar (Halliday 1985) posits three metafunctions of an utterance: *experiential, textual* and *interpersonal*.[1] The *experiential* (or *ideational*) metafunction relates to the use of language to express propositions about the world, what we typically associate with semantic meaning. The *textual* metafunction relates to the use of language as a contextualized system, as something we do in response to the communicative circumstances in which we find ourselves. The *interpersonal* metafunction relates to the use of language as a social act, as something we do to maintain social relations. All utterances function at all three levels, and we have different tools at our disposal to emphasize each. The *experiential* metafunction foregrounds the need to be informative, whereas the *interpersonal* metafunction foregrounds the desire to maintain the interaction and social relation. The *textual* incorporates both communicative priorities by conveying information about why something is being said in a certain way and at a certain time.

To demonstrate the layering of the three metafunctions, consider the example in (18).

(18) In the corner of the bar, obscured by shadow, sat a stranger hoping not to be noticed.

First, this utterance communicates a proposition conveying truth-conditional information. The utterance asserts the existence of a stranger in a particular mental state and a particular physical location. This is the utterance's *experiential* function – by reading it we gain new knowledge about the state of the (fictional) world. Second, we have a sense while reading it that we are in the middle of a story. We know the bar, and we're interested in what is happening in it. This is the utterance's *textual* function – by reading it we gain knowledge about the setting of a story and what is likely to be the next plot point (i.e. discovering the

[1] As summarized and updated in Thompson (2013).

stranger's identity). Finally, there is a feeling of suspense in reading the sentence – we really want to know what is happening in the shadowy corner of the bar. This is the utterance's *interpersonal* function – in reading it we become curious and engaged with the information being conveyed.

To achieve each of these functions, I am using a set of communicative conventions. To convey the propositional information I am abiding by the syntactic conventions of English, putting the location in a prepositional phrase headed by "in", expressing the stranger's mental state with the gerund clause "hoping not to be noticed", and so on. To convey contextual information I am taking advantage of topic-comment structure and a non-canonical word order, placing the known location at the beginning of the sentence and the new information at the end. To successfully build suspense, I am taking advantage of attentional structure, making you wait just long enough to hear the new information I would like to frame as particularly exciting and relevant to the unfolding story. This demonstrates a negotiation of informative and social pressures – I am using different strategies to ensure that I am conveying the information I want to convey while doing my best to keep you engaged in the interaction.

Each of the strategies described can be framed as taking advantage of resources or abiding by rules. In constructing a model of interactive meaning in gesture, both perspectives must be considered. This necessity is integrated into the *principles of multimodal discourse management* proposed in Chapter 1 (and repeated below).

(19) **Principles of Multimodal Discourse Management**

a. *Multiple strategies*: There are multiple strategies, within both the verbal and gestural modes, for expressing discourse management.

b. *Optionality of expression*: The use of each expressive strategy is optional and subject to contextual variation.

c. *Independence of contribution*: Strategies may be employed independently and each independently employed strategy can profile a different aspect of discourse management.

d. *Compositional management*: The strategies employed in both modes are integrated systematically and predictably into a single coherent multimodal message.

Principles A (*multiple strategies*) and D (*compositional management*) foreground a grammar-as-rules perspective. To achieve communicative goals, we have a particular conventionalized inventory of strategies that combine with each other in a systematic and predictable way. Chapter 4 develops a framework to model these conventions. Principles B (*optionality of expression*) and C (*independence of contribution*) foreground a grammar-as-resource perspective. In achieving discourse goals there is flexibility. We can take advantage of this flexibility to reflect our priorities as the interaction unfolds and as we negotiate informative and social pressures.

3.2.2 Being cooperative and polite

Grice's (1975) *Cooperative Principle* and Lakoff's (1973) *Rules of Politeness* set a foundation for understanding discourse as a cooperative activity shaped by particular conventions.[2] These two sets of principles are especially helpful in the present work for considering strategic communicative trade-offs as a discourse unfolds. In prioritizing informative imperatives, we may make social sacrifices. In prioritizing social imperatives, we may settle for a less informative message. These principles provide a rule-based perspective on what these trade-offs are and what is at stake.

Grice's Cooperative Principle posits that discourse participants act according to a shared rationality and assume the rationality of one another. For example, as you read this dissertation, you assume that I am writing with the intention of communicating something coherent and worthwhile. If you did not make this assumption, you would have little reason to engage with the text. The Cooperative Principle is composed of four maxims that rational interlocutors are expected to abide by. These are paraphrased in (20).

(20) **Grice's (1975) maxims**

a. *Quality*: tell the truth

b. *Quantity*: convey all necessary information and nothing more

c. *Relation*: only make relevant contributions

d. *Manner*: convey information in a clear and orderly way

The Cooperative Principle is not meant to imply that rational participants always abide by all maxims. Indeed, there are many communicative situations in which it is expected that interlocutors do not abide by this information (e.g. when telling a fable or joke). Instead, the principle and its constituent maxims are meant to provide violable constraints on how interactions are expected to unfold – in the model, interlocutors expect each other to, by default, honor the maxims, but allow for violation within reason. This principle, as a set of violable rules, makes predictions as to how interlocutors will choose to express themselves. Flouting a maxim puts the perception of one as a rational actor at risk. If in a particular circumstance an interlocutor decides to flout a maxim, there should be a reason, and if that reason is not obvious, it should be expressed to maintain the perception of their rationality. The forms of interactive meaning I consider in this work often signal the flouting of a maxim. The interlocutor is doing something unexpected, and uses some conventionalized strategy to warn their addressee and justify (or apologize for) their uncooperativeness. To demonstrate this, consider the discourse excerpt in (21).[3]

In this example, Colbert has asked his interviewee, Irish actor Andrew Scott, what his favorite love story is. In response to this question, Scott may be expected to provide an

[2]Despite their apparent age, these principles remain incredibly influential and continue to generate insights and debate in work on cooperative discourse. See, for example, Hadi et al. (2013) for a critical review of Grice's continued influence, and Dynel (2009) for discussion of the relationship between 'cooperation' and 'politeness' in the refinement of theories of cooperation in discourse.

[3]This discourse excerpt is also discussed in Chapter 5 on multimodal topic-shifting.

honest answer and a sufficient explanation for his answer. The excerpt in (21) is part of his expected explanation for why *Hamlet* is his favorite love story. The gist of his answer is that he finds it impressive that a four-hundred year old text can still be so effective and relevant. However, he expresses this information in a very long and somewhat circuitous way, potentially flouting the maxims of *quantity* and *manner*. Some of the information may even be *irrelevant* (e.g. that people are willing to binge watch television shows; lines 13-14). Scott, as a rational agent, is aware that his answer might be perceived as uncooperative, and signals this twice (lines 7-8 & 16, relevant segments underlined).

(21) TRANSCRIPT 1: ANDREW SCOTT
[UID:f00485b2-051e-11ea-8f95-089e01ba0770,3510]

```
1  AS  that story, which always just blew my mind, it's a four
2      hundred year old story, and the idea that it's about a young
3      man who's got mental health issues, has been told by his
4      family "don't feel, don't feel, c'mon, be a man, be a man, be
5      a man".  Um, it's a-, we did it a hundred and fifteen times,
6      and it's nearly a four hour long play.  We had incredible
7      audiences.  I'm a bit of a Shakespeare nerd, so- and I don't
8      want to bore you.  But wh-what I find fascinating about it is
9      that, y'know in order to for Shakespeare to stay relevant, I
10     don't think y- you need to cut it down.  I think you might
11     need to make it as uh- exciting and as thrilling for an
12     audience uh- of today as it would be four-hundred years ago
13     because, y'know, we binge watch TV. Y'know, we watch five
14     hours of television if it's exciting.  So the idea is don't
15     cut it down, just make it four hours of really exciting um
16     play- plays.  Anyway, I'll stop talking Shakespeare (laughs)
17 SC  No not at all, not at all.  That's amazing.
```

First, he provides a warning as to his digression (lines 7-8); he doesn't want to "bore" Colbert (i.e. flout the quantity maxim by providing too much information for too long) and provides a justification for risking this (because he's a "Shakespeare nerd"). After continuing for several more sentences, and providing potentially irrelevant information, Scott signals that he knows he is being uncooperative and is prepared to stop (line 16).

Grice's maxims help us to understand why Scott perceives the *information* he is providing to be uncooperative, and thus why he might choose to offer an explanation or apology. However, they fail to appreciate the underlying *social* motivations. For example, the maxim of quantity predicts that Scott's overly long answer is uncooperative, but it does not overtly relate this to Scott's concern about 'boring' Colbert. It is also not clear from these maxims why Colbert would comfort Scott after he was self-admittedly uncooperative (line 17). This is where Lakoff's (1973) *Rules of Politeness*, as paraphrased in (22), prove useful.[4]

[4]See Dynel (2009) for a critical review of the relationship between these two sets of principles.

(22) **Lakoff's (1973) Rules of Politeness**

a. *Don't impose*: don't require more time and effort than is necessary

b. *Give options*: respect your interlocutor's agency by allowing them to choose their actions

c. *Be friendly*: make your interlocutor feel valued

Scott's expressed concern about boring Colbert can be related directly to Lakoff's *don't impose* rule. Providing more information than necessary violates the *quantity* maxim, which is uncooperative *because* it imposes unnecessarily upon Colbert's time. This, in turn, helps to explain why Colbert offers comfort, denying that Scott has imposed. In doing so, Colbert is *friendly*, a particularly important rule to follow in his role as host of the show.

As with Grice's maxims, there are plenty of circumstances in which the rules of politeness are violated. We might not care to be friendly in a heated argument, and we are unlikely to give options when providing directions to the store. In certain types of interaction, politeness is simply not the goal, and does not play a significant role in determining strategies for expression. As before, the purpose is not to identify principles that are never violated. The purpose is to identify principles that are honored 'by default' and require explanation when violated.

These two sets of principles are especially important for Principle B (*optionality of expression*) and Principle C (*independence of contribution*) in the proposed multimodal discourse management. If strategies for expressing discourse management are both optional and independent, then we need a systematic way to reason about why strategies are used at all, and what each is meant to accomplish. Now we can say that strategies are used to signal cooperation and politeness, especially when they are perceived to be at risk. The independence of strategies can be said to contribute to a negotiation between communicative imperatives – where one strategy addresses informativeness concerns, another may address social concerns.

3.3 Modeling discourse goals and contributions

In the previous section, I discussed principles of cooperation in discourse and how they relate to different dimensions of meaning and communicative imperatives. Though this is helpful in discussing interlocutor behavior in general, we do not yet have a clear way to predictably model a discourse's structure, contribution by contribution, as it unfolds. We know what interlocutors should *not* do in order to meet informative and social imperatives. We don't really know what interlocutors *should* do, except in very general terms. To make predictions as to what cooperative and polite interlocutors should do at any given time, I use a question-based approach to model discourse structure (Roberts 1996, 2012; van Kuppevelt 1995).

In this section, I begin by providing a sketch of the question-based approach I use (3.3.1) and discuss its relationship to naturalistic data (3.3.2). I then discuss how question-based structures are co-constructed in conversation as interlocutors make particular discourse

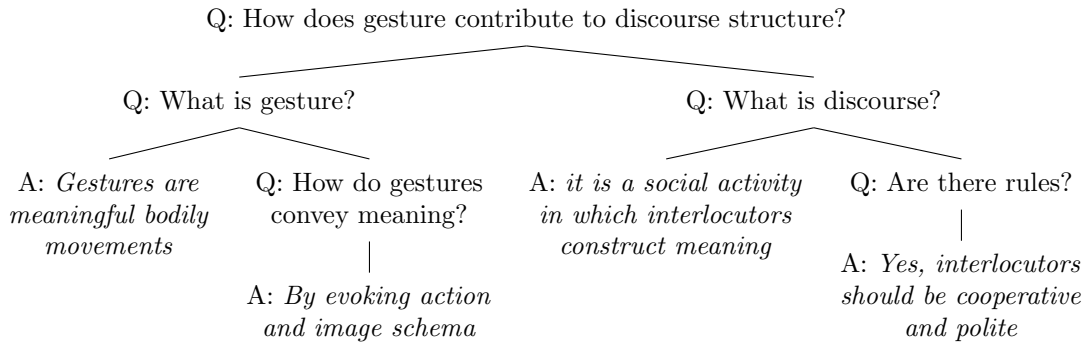
moves (Section 3.3.3) and mark their moves to maintain coherence and signal cooperation (Section 3.3.4).

3.3.1 Discourse as a pursuit of answers

In a question-based approach to discourse structure, a discourse is conceptualized as a set of nested question-answer pairs. A discourse begins by posing a question that interlocutors cooperatively answer as the discourse progresses. Each utterance either offers a (partial) answer to the question, or sets a strategy for answering a complex question by posing a sub-question that is easier to answer. This understanding of discourse is fundamentally goal-oriented – interlocutors establish a question as the discourse's goal, and move toward achieving that goal with each contribution. In the absence of a clearly articulated goal, as in casual conversation, the question being addressed is a very general one – *what are the way things are?* (Roberts 1996, 2012). The goal set by pursuing such a general question is simply to better align interlocutors' understanding of the world. To demonstrate, let us consider the structure of this dissertation so far.

The primary goal of this dissertation is to contribute to a better understanding of how gesture contributes to a discourse's structure. We can phrase this as the "big question" being pursued – *how does gesture contribute to discourse structure?* Of course, this is a very complicated question, and to achieve anything close to a sufficient answer we have to break it down into more manageable parts. First, we have to make sure that we agree on what *gesture* and *discourse* are. Chapters 1 & 2 offered answers to the first preliminary question (*what is gesture?*). I began by offering a general definition (*gestures are meaningful bodily movements*), and then addressed questions that arose from it (e.g. *if gestures are meaningful bodily movements, then how do gestures convey meaning?*). This process was repeated until, hopefully, enough information was provided to comfortably agree that we know *what gesture is*. This chapter offers answers to the second question (*what is discourse?*), and follows the same procedure. The resulting discourse can be represented as a branching hierarchical structure, as in (23).

(23)



Each contribution to a discourse (i.e. each utterance) constitutes a discourse *move*, ideally one that help us achieve our set goals. A question is considered *open* until it is sufficiently

answered or deemed unanswerable. When a question is *closed*, i.e. sufficiently answered or deemed unanswerable, we set a new strategy for answering a more general, still open question. For example, after Chapters 1 & 2, the question *what is gesture* was closed, and I moved on to the next step in answering the more general, still open question *how does gesture contribute to discourse structure*. The strategy I set for doing so was to pose the question *what is discourse* as the next immediate goal. If we have sufficiently answered all open questions we can either set a new discourse goal or end the discourse.

3.3.2 Unanswered and unasked questions

If my dissertation were an ideal discourse, each sentence would move us incrementally closer to an answer to the big question. By the end, the reader and I would leave without any lingering questions, fully satisfied with the answers that I've carefully organized and offered, and sure that we've met our goals. Of course, this dissertation is not an ideal discourse. It is likely that I won't answer all of the questions that you pose as you read. It is also likely that I will offer answers to questions that you don't find particularly helpful.

Face to face conversation is more troublesome still. We digress, backtrack, and get interrupted. We get distracted by unexpected details while haphazardly attending to other tasks. We lose interest in the questions we've asked and abruptly pursue others. We begin to say something only to realize that we can't quite explain it, or don't want to, or we notice that our addressee is annoyed. If actual conversation so often fails to organize around nicely ordered question-answer pairs, then how could this possibly be a good model of discourse?[5]

As is the case with Grice's maxims and Lakoff's principles of politeness, the usefulness of this model, as I understand it, is not necessarily in modelling what we *actually* do. It's usefulness arises from the simplicity of its predictions. Question-based discourse models predict, very simply, that each utterance contributes to the same goal as the previous utterance (to answer a particular question) until that goal is achieved. This prediction then invites an equally simple empirical question – what happens when an utterance *doesn't* contribute to a discourse goal? I use this question as a strategy for answering the "big question" of this dissertation.

3.3.3 Making moves

As stated, every contribution to a discourse, every utterance conveyed by discourse participants, constitutes a discourse *move*. A move can consist of an answer to an open question, or can pose a new question to be addressed. Given this, I consider three types of discourse move in this work: *specification, digression,* and *shift*. Each of these moves contributes to a discourse's structure in a different way.

A *specification* move offers an answer to the most immediate open question. Consider, for example, my explanation of the discourse structure framework I use in this work. If I

[5]See recent work by Ozerov (2022) for a recent and particularly detailed critique of Question Under Discussion as failing to account for naturalistic data.

were to say I take 'a goal-oriented approach', a reasonable question to ask would be 'what constitute discourse goals?'. Answering this question constitutes a specification move. This is exemplified in (24) and schematically represented in (27). I consider specification to be the maximally cooperative move – given that the goal of a discourse is to answer questions in the order that they arise, and given the cooperation of both interlocutors, a specification is expected as the default contribution.

(24) I take a goal-oriented approach to cooperative discourse. [In particular, I assume a question-answer structure in which discourse goals are equated to answering open questions]$_{\textsc{Specification}}$

A *digression* move pursues a question embedded under an answer rather than an open question (van Kuppevelt 1995; Riester 2019). This kind of contribution is thematically related to a topic under discussion, but does not contribute to achieving discourse goals. Crucially, it also doesn't abandon discourse goals. Once a digressive question is answered, the discourse's still open questions are returned to. An example of a digression is given in (25) and schematically represented in (28). I consider this move type to be minimally disruptive – an interlocutor is only momentarily uncooperative.

(25) In a question-based approach to discourse structure, cooperative interlocutors are expected to address open questions until they are sufficiently answered. [By the way, some analysts argue passionately against question-based approaches.]$_{\textsc{Digression}}$ Anyway, if a question cannot be sufficiently answered, participants can agree to close the question as 'unanswerable' and move on.

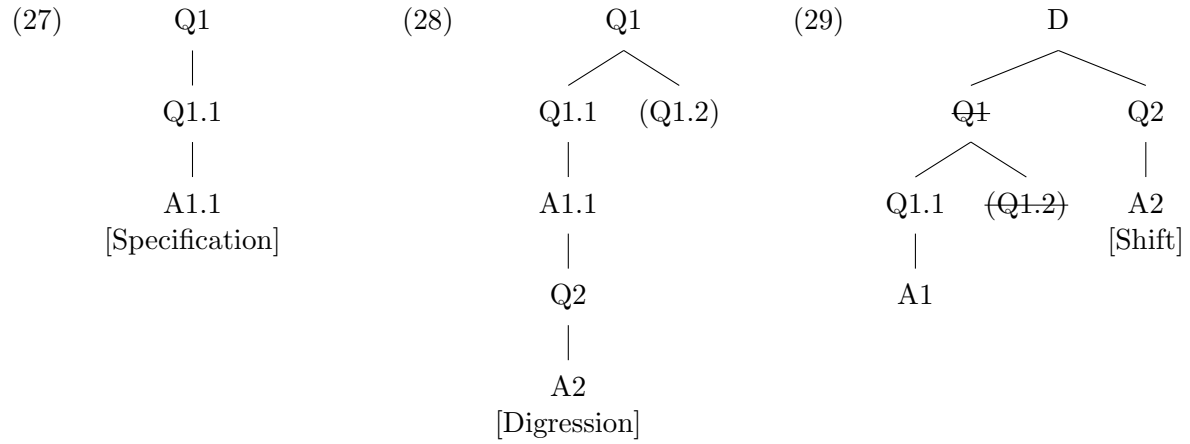
A *shift* move pursues a question to a newly posed question before the previous question is fully answered. An example of a shift move is given in (26) and schematically represented in (29). This shift is particularly jarring because I begin the first sentence with "on the one hand", which sets an expectation that I will present a second contrasting condition, likely prefaced by "on the other hand" (Scholman et al. 2017).

(26) On the one hand, speakers might have good reasons to perform a shift in certain contexts. [But what I really want to talk about is how gestures express these different moves.]$_{\textsc{Shift}}$

I consider this move type maximally disruptive – an interlocutor has, without negotiation, abandoned a discourse goal and established a new one. Though I consider all shifts more disruptive to a cooperative discourse than either of the other two move types, not all shift moves are equally disruptive. The more general the abandoned question is, the more disruptive the shift is to established discourse goals.

These three moves are schematically represented in (27)-(29). Questions indexed with a simple number (Q1, Q2) are considered primary discourse goals. Strategic questions for answering primary questions are indexed as increasingly complex embedded numbers (Q1.1, Q1.1.1, and so on). For example, if two specific questions are used in pursuit of a more general question, the more general question (Q1) will branch to two immediate subordinate

sub-questions (Q1.1, Q1.2). If one of those subordinate questions is too complex and requires multiple partial answers, it also branches into immediately subordinate sub questions (Q1.1.1, Q1.1.2).[6] When an answer does not sufficiently answer an open question, it is expected that the next move will pursue another sub-question that will contribute to a sufficient answer. *Expected* questions are represented within parentheses (e.g. "(Q1.2)"). These 'expected' questions are important for understanding the disruption caused by digressions and shifts. To qualify as a digression, the expectation that an open question will still be pursued is maintained. A shift cancels expectations. The cancellation of an expectation is represented as the expected question and its super-ordinate open question being crossed out (e.g. "~~(Q1.2)~~").



(27) is a representation of a specification move in which an answer (A1.1) is offered for the most recent open question (Q1.1). This question partially answers the main question (Q1). The default cooperative next move would be to offer another specification to a second sub-question of Q1 (Q1.2).

(28) is a representation of a digression move in which an answer (A2) is offered to a digressive question embedded under A1.1. This means that some aspect of A1.1 is questioned, not as an answer to a discourse question, but on some other grounds. In the example provided in (25), the digressive question is something like *do all discourse structure analysts use question-based approaches?* This question does not help to answer the open discourse question regarding the framework I use for discourse analysis. However, it is directly related to the immediately preceding contribution (i.e. that I use a question-based approach). The cooperative next move given this discourse structure would be to close the digressive question and return to address another sub-question of Q1 (Q1.2).

Finally, (29) is a representation of a topic-shift move in which an answer (A2) is offered to a newly posed discourse question (Q2). This question is not embedded under a previ-

[6]Most question-answer structures I discuss are binary-branching, but this is not a constraint on the model. A question can consist of any number of immediately subordinate sub-questions. For example, a question regarding a sequence of events (e.g. *what happened this weekend?*) may consist of an indefinitely long set of *what happened next?* sub-questions.

ously established question, and does not constitute a strategy for addressing an established discourse goal. Importantly, this move occurs before Q1 is considered sufficiently answered. The cooperative next move would be to accept the premature closure of Q1, and offer a contribution to address Q2. As we will see in Chapter 5, this is an oversimplification of topic-shifting which is a structurally and functionally heterogeneous class of discourse move.

In Chapters 5-7, I consider these moves in reverse order, beginning with the most disruptive and working back to the most cooperative. In doing so, I am working under the assumption that the discourses in my data set are cooperative, and that cooperative interlocutors have a responsibility to justify any disruptions. Given this, we expect the most disruptive move type, *shifts*, to be the most obviously marked.[7] Topic-shifts, as conceptualized in this work, directly violate cooperative expectations, and a cooperative interlocutor should warn their addressee of this violation. Specification moves should be the least obviously marked, because they are what a cooperative interlocutor is assumed to do anyway. Digressions are somewhere in between, violating expectations, but only temporarily. In other words, my exploration of interactive meaning in gesture begins where gestural expressions should be most obvious, and becomes more nuanced as subtler cues are considered.

3.3.4 Marking moves with words and gestures

When contributing to a discourse, interlocutors may provide textual information about what their contribution is and how it relates to discourse goals. The most well-studied of these cues are lexical discourse markers – words and phrases that signal something about the structure of the discourse and interaction (e.g. Fraser 1999; Redeker 2006; Schiffrin 1987).[8] Consider, for example, the bolded phrases in (30), all taken from this text.

(30) a. **First**, we have to make sure that we agree on what gesture and discourse are.
b. If we have sufficiently answered all open questions, we can **either** set a new discourse goal **or** end the discourse.
c. Consider, **for example**, the bolded phrases in (30).

In each case, the bolded words offer some information about the intended purpose of the contribution and set expectations for what the reader will encounter next. "First" signals that I am offering a partial answer to a complex questions and that more partial answers are forthcoming. "Either...or" signals that my contribution contains two contrasting points. "For example" signals that the contribution I'm about to make provides support for something I've already said.

[7]The intuition that more disruptive or unexpected discourse moves require some kind of warning is supported by work on lexical discourse markers. This work has shown that the less 'expected' the move is, the more likely it is that a discourse marker will be used (e.g. Asr & Demberg 2012; Murray 1997; Sanders 2005).

[8]These cues go by many names in the literature, including *discourse particles* (e.g. Aijmer 2002; Fischer 2000), *pragmatic markers* (e.g. Aijmer 2013; Fraser 1996), and *coherence markers* (e.g. Kamalski et al. 2008; Sanders & Spooren 2009), among others. I favor the term *discourse markers* in this work.

Very broadly speaking, discourse markers can perform two functions (Fraser 1999) – they can convey information about what type of discourse *move* is taking place, or they can convey information about the *relationship* that holds between consecutive segments. Often they provide both types of information. In a question-based framework, the first of these functions relates to the attachment of a contribution to the overall discourse structure. The second function relates to the identity of the question. For example, an "either...or" construction provides both kinds of textual information. It tells us that there is an open question composed of exactly two subordinate questions. It also tells us that those two subordinate questions stand in contrast to one another (e.g. *what about X?* and *If not X, what about Y?*).

I am primarily concerned with the first function – when and how discourse moves are signaled. In particular, I am interested in how interlocutors use gesture to signal that their contribution is expected and cooperative (i.e. a *specification*) or unexpected and disruptive (i.e. a *digression* or *topic-shift*). I take advantage of previous work on lexical discourse markers to address these questions through corpus analysis. The markers *here's the thing*, *by the way*, and *anyway* are used as proxies for gathering instances of the discourse moves they are expected to signal (*specification, digression* and *topic-shift* respectively). For each case study I then look at variation in the gestures that co-occur with the lexical discourse markers, focusing on the five action schemas discussed in Chapter 2 – Presentation, Removal, Stopping, Referring and Reorientation.

This methodology also leads to new insights into the function and distribution of the search terms (*here's the thing*, *by the way*, and *anyway*). Variation in the gestural mode can direct us toward finding polysemy and underspecification in the verbal mode. For example, if different action schematic gestures are used recurrently with *anyway*, this may indicate that *anyway* itself can be used differently, or that 'topic-shifting' is not a single monolithic phenomenon.[9] Throughout the three case studies, I will argue that this is indeed the case.

3.4 Metaphor in discourse theory

The previous sections in this chapter outlined my approach to discourse structure as an abstract system and discussed the ways in which informative and social pressures in communication underlie the model used. In this section, I discuss the ways in which Conceptual Metaphor Theory (Lakoff & Johnson 1980) can be used to assess the consistency of our discourse structural framework across the verbal and gestural modes. This step can be considered a form of *meta*-analysis – we have theories of gesture and discourse structure and we are critically assessing to what extent they are compatible. To illustrate how this fits in with the overall process of analysis and reflection in this work, consider Figure 3.1.

[9]This is reminiscent of contrastive approaches to lexical discourse markers in which cross-linguistic translations are used to identify polysemy networks (e.g. Aijmer & Simon-Vandenbergen 2003; Cuenca 2008; Fischer 2000; Takahara 1998).

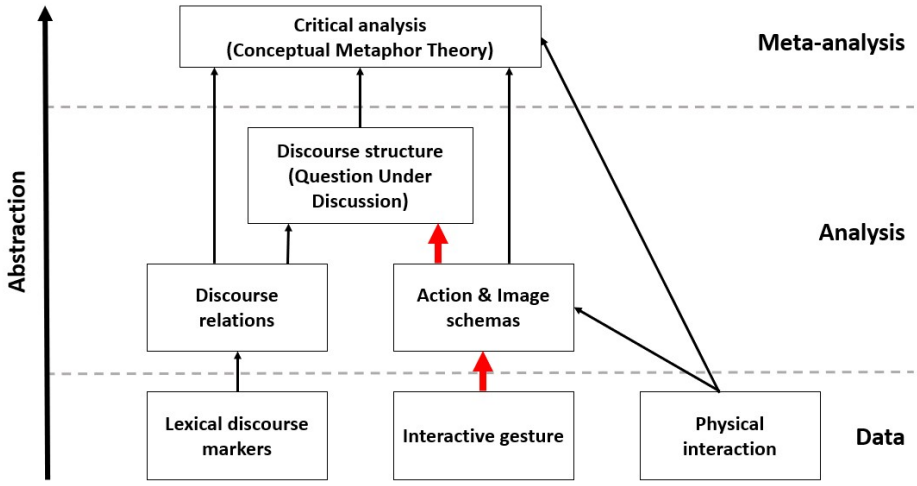


Figure 3.1: Levels of analysis

I am primarily concerned with better understanding how gesture contributes to managing discourse structure, the steps of analysis indicated by red arrows in the diagram above. However, to do so, I must rely on pre-existing notions of *discourse structure* and theories of how verbal cues, like lexical discourse markers, contribute to its management. Because these are themselves abstract models developed by analysts in pursuit of their own research questions, they are inevitably imperfect representations of what discourse actually is and how verbal cues actually work. The action schematic model I use for modeling gesture meaning is also, inevitably, an imperfect representation of how gestures actually work. When integrating abstract models, we need to ensure that these imperfections, the artifacts left by reducing infinitely rich data to an abstract workable model, are, at the very least, *compatible* with each other. If we do not take this step, we may misinterpret how and why our findings confirm or dispute the predictions made by a given theory.

Conceptual Metaphor Theory is a particularly powerful tool for assessing the compatibility of our models of gesture meaning and discourse structure for two reasons. First, since Lakoff & Johnson's (1980) seminal work on primary metaphors, it has become widely accepted that metaphor plays a fundamental role in both cognition and communication. Metaphor is *everywhere* in our use of language, and, as a result, is *everywhere* in the discussion and development of theoretical models.[10] This is especially apparent when discussing and cognizing about abstract concepts, such as the structure of discourse (e.g. Jakobs & Hüning 2022; Sweetser 1992). Our theories of discourse structure are inherently metaphoric. There are no literal branching trees of question-answer pairs. Tree diagrams, as in QUD (e.g. Riester 2019) or Minimalism (e.g. Hornstein et al. 2005), and embedded box structures, as in Segmented Discourse Representation Theory (e.g. Asher 1993) or Construction

[10]See reviews in, for example, Gibbs (2014); Landau et al. (2010); Ottati & Renstrom (2010), and Thibodeau et al. (2017).

Grammar (e.g. Fried & Östman 2004), are not discourses. They are representations of discourse. Hopefully, they are helpful representations that enable us to learn something about an ephemeral, intangible system. The pervasiveness of metaphor ensures that we have something to compare across our models – what metaphors are we using to theorize about gesture, discourse markers, and discourse structure? Are they the same metaphors, or are there irreconcilable differences?

The second strength of Conceptual Metaphor Theory for meta-analytic purposes is that it is explicitly motivated by an *embodied* understanding of thought and language. Under this theory, the metaphors we use in everyday speech and, eventually, in developing complex models of abstract systems, are motivated by observable bodily experiences. We use our basic everyday interactions with the world, such as walking around the house and organizing our belongings, to reason about everything else. Using Conceptual Metaphor theory in meta-analysis is thus consistent with contributing to the 'embodied turn' in linguistics (Mondada 2016; Nevile 2015), an underlying motivation of this work. In using Conceptual Metaphor Theory to critically analyze the compatibility of discourse structure models with models of gesture meaning, I am explicitly drawing on notions of embodiment at two levels of analysis – first in determining a gesture's action schema, and second in determining whether my interpretation of that action schema in context is consistent with my interpretation of, say, a particular lexical discourse marker in context.

In this section, I consider three basic metaphors that are at play in our discussions of communication and discourse structure. I discuss the ways in which these metaphors ground our theories in every day experience, and the ways in which they are reflected in action schematic gestures.

3.4.1 Communication as Object manipulation

The 'Conduit Metaphor', as proposed by Reddy (1979), is one of the most influential analyses of metaphor for communication. The Conduit Metaphor holds that COMMUNICATION is conceptualized as OBJECT EXCHANGE, such that interlocutors "package" meaning into linguistic expressions which are then "sent to" and "received by" their addressee. As Grady's (1998:205) revisiting of this proposal notes, the Conduit Metaphor has directly influenced influential analyses of speech acts (Johnson 1987), the psychological reality of metaphor (Gibbs 1994), and semantic change (Sweetser 1990).

The Conduit Metaphor, in its original formulation as COMMUNICATION AS OBJECT EXCHANGE certainly does appear in spoken and written language, as exemplified in (31). However, Grady (1998) points out that this metaphor, rather than being the primary metaphor for communication, can be broken down into more general metaphors that are grounded more clearly in embodied experience.

(31) a. His phrasing **sent** the wrong message.
b. Slogans **carry** little meaning.
c. You **gave** me a great idea.

d. Let's **exchange** information.

Though I agree with Grady's general intuitions, I argue that the gesture data discussed here, as well as a cooperative understanding of discourse structure, necessitates the more general metaphor COMMUNICATION AS OBJECT MANIPULATION. The Conduit Metaphor is then a subcase of this more general, more embodied conceptualization.

Table 3.1 outlines the metaphoric mappings for the general COMMUNICATION AS OBJECT MANIPULATION metaphor. These mappings presuppose, or are compatible with, primary metaphors suggested by Grady (1998) and Lakoff & Johnson (1980), especially IDEAS ARE OBJECTS, CONSTITUENTS ARE CONTAINERS, and KNOWING IS SEEING.

COMMUNICATION		OBJECT MANIPULATION
Participants: interlocutors	$\Longleftarrow$	Participants: agents
Entities: topics	$\Longleftarrow$	Entity: manipulated objects
Event: introduction of topic	$\Longleftarrow$	Event: presentation of object
Event: topic management	$\Longleftarrow$	Event: manipulation of object
Event: ending of topic	$\Longleftarrow$	Event: removal of object

Table 3.1: COMMUNICATION AS OBJECT MANIPULATION metaphoric mappings

This metaphor is omnipresent in our casual and theoretical discussion of discourse. We *pose* questions as metaphoric objects to be considered, and *offer* answers to be accepted or not. The discourse itself has *structure* that we *build* together, evoking CONVERSATION AS CONSTRUCTION, a more elaborate sub-case of COMMUNICATION AS OBJECT MANIPULATION. The crucial point in using this metaphor rather than the traditional Conduit Metaphor is that the discourse topics as metaphoric objects are not simply passed from one participant to another, but are instead located in a shared, mutually accessible space where careful inspection and manipulation is possible for both participants. To say otherwise would be to limit the interlocutors' ability to contribute to an ongoing cooperative discourse.

The COMMUNICATION AS OBJECT MANIPULATION metaphor is also very apparent in the interactive gestures discussed. Metaphoric objects are PRESENTED, REFERRED to and REMOVED from a shared space. As long as a topic as a metaphoric object remains in the shared space, both interlocutors can interact with it. Conceptualizing COMMUNICATION AS OBJECT EXCHANGE cannot account for the fact that these movements *mean* relative to a shared space. The presentation of a topic must involve movement into the shared space, and the removal of a topic must involve movement out of that shared space. The understanding of interactive gesture as manipulating mutually accessible metaphoric objects provides the foundation for the model of interactive meaning proposed in Chapter 4. This means that the primary metaphor underlying my proposed model of interactive meaning in gesture is consistent with that in our theory of discourse structure more generally.

3.4.2 Discourse as a Journey

We can also frame DISCOURSE as a shared JOURNEY, in which we move toward discourse goals together (e.g. Sweetser 1992). This metaphor is, of course, a subcase of the primary metaphor PROCESSES ARE JOURNEYS (Lakoff & Johnson 1980), which allows us to frame all activities as motion-based. This metaphor is frequently evoked in natural conversation as we comment on what is happening in the discourse and what we would like to happen. I can encourage you to keep talking by saying *keep going* or *go on*, then we can *move on* from a topic when we're ready. Hopefully we can *get past* any misunderstandings and *reach* an agreement. You can *leap to* a conclusion by not providing enough reasoning. If I am not clear enough, you may not *follow* my reasoning and we may get *stuck in* a misunderstanding.

Table 3.2 outlines the metaphoric mappings entailed by the DISCOURSE AS A JOURNEY metaphor.

DISCOURSE		JOURNEY
Participants: interlocutors	$\Longleftarrow$	Participants: travellers
Entity: communicative task	$\Longleftarrow$	Entity: journey
Goal: align knowledge states	$\Longleftarrow$	Goal: reach destination
Locations: states of a discourse	$\Longleftarrow$	Locations: points along a path
Process: introducing, managing and ending topics	$\Longleftarrow$	Process: travelling along a path

Table 3.2: DISCOURSE AS A JOURNEY metaphoric mappings

This metaphor has already appeared in our discussion of discourse structure. Most centrally, it underlies the notion of discourse *move*, in which making a contribution to the discourse is conceptualized as moving along a path. It also appears in certain recurrent phrases, such as when I discuss interlocutors as *moving toward* discourse goals, *moving* the discourse *forward*, *pursuing* answers, or *navigating* communicative pressures. Creative uses also appear in academic discourse, such as when Glynn et al. (1982:196) argue that discourse markers "pave the way for important items of information" (Glynn et al. 1982:196)[11].

This metaphor is not as evident in the interactive gestures discussed in this work, nor does it play a significant role in the model of interactive meaning proposed in Chapter 4. However, it is arguably evoked by the STOPPING action schema, if the stopped motion is interpreted as the forward motion of the interlocutor (e.g. Wehling 2017). It also appears exceptionally in discussions of specific gesture performances, such as when a speaker moves their hands in repeated circular motions as they begin speaking, as if to propel themselves forward (e.g. Ladewig 2011, 2014a).

[11] As cited in Gaddy et al. (2001:89).

3.4.3 Discourse as Physical Space

The final metaphor is an even more primary, more embodied metaphor that underlies the others discussed – Wehling's (2017) Discourse Space is Physical Space. In saying that Communication is Object manipulation or Discourse is a Journey, we presuppose a shared physical space in which the object manipulation or journey take place. If communication is indeed a cooperative activity, it is not enough to say that communication involves agents manipulating metaphoric objects. We have to say that the space in which the objects are manipulated is shared, and that the manipulation of objects directly affects the accessibility of those objects for discourse participants. This metaphor is particularly embodied because in face-to-face conversation the discourse *literally* takes place in a shared physical space.

As with the other metaphors discussed, this metaphor appears regularly in casual discourse. We can be *in the middle* of a story or *at* its conclusion. We take certain *positions* and *stances* on topics under discussion, and we *view* those topics from certain *angles*. All of these uses frame discourse as a location in which we as participants, as well as the metaphoric objects under discussion, are located.

Table 3.3 outlines the basic metaphoric mappings entailed by the Discourse as a Physical space metaphor.

Discourse		Physical Space
Participants: interlocutors	$\Longleftarrow$	Participants: agents
Location: (cooperative) discourse	$\Longleftarrow$	Location: shared physical space
Entity: topics in discourse	$\Longleftarrow$	Entity: objects in shared space
Event: participation in discourse	$\Longleftarrow$	Event: interaction in shared space

Table 3.3: Discourse as a Physical space metaphoric mappings

This metaphor appears in the framing of very basic theoretical notions of discourse structure. For example "common ground", the knowledge shared and accumulated by participants throughout a conversation (e.g. Stalnaker 1978), frames discourse as a Location in which interlocutors communicate information and align their understanding of the world. In discussion of turn-structure we frame beginning and ending a turn as contributor as taking and ceding "the floor" (e.g. Sacks et al. 1978). In interactive gesture, this metaphor is perhaps most apparent in Reorientation actions where the gesturer repositions themselves relative to a shared space in order to signal participation or lack thereof in the ongoing discourse.

This metaphor is central to the model of interactive meaning proposed in Chapter 4. The proposed Interaction Space is simultaneously the physical space in which an interaction takes place (the Source domain of this metaphor) and the metaphoric space in which participants co-construct a discourse (the Target domain of this metaphor). It is also evoked by the three lexical discourse markers I have chosen, *anyway, by the way* and *here's the thing*, all of which are transparently derived from spatial terms (e.g. Traugott 1985, 2020, 2022).

3.5 An embodied discourse structure

One of the central goals of this dissertation is to demonstrate the ways in which the incorporation of gesture into linguistics can benefit our theories of discourse. In this chapter, I first outlined the approach to discourse that I use in this work, an approach that was developed without gesture in mind. I then turned to the prevalence of metaphor in our existing theories and how recognizing these metaphoric framings is a crucial step in assessing the compatibility of our theories. In the remaining chapters, I explore the ways in which interactive gesture physically manifests the metaphors we already implicitly rely on, and argue that this provides an unparalleled opportunity for empirically studying discourse management as an embodied practice.

Chapter 4

The Interaction Space

This chapter introduces the proposed Interaction Space, quite literally the space between interlocutors in which an interaction takes place. The development of the Interaction Space as it pertains to interactive gesture and discourse management is one of the main theoretical contributions of this work. The Interaction Space is meant to provide a predictive framework for analyzing interactive meaning in gesture. Though functionally similar spaces have been described in previous literature on interactive gesture, this is, to the author's knowledge, the first significant attempt at formalizing such a space. In addition to building an explicitly predictive model of interactive meaning in gesture, I hope to show the ways in which it can be integrated into existing theories of discourse structure. In defining the Interaction Space and its functions, I use tools from a range of cognitive linguistic theories, especially Cognitive Metaphor Theory (Lakoff & Johnson 1980) and Mental Spaces Theory (Fauconnier 1994), and integrate the insights of previous observational work on interactive gesture.

4.1 Introduction

If the reader, upon reading the title of this dissertation, insists on asking "well, what is the shape of discourse?", my answer is this – the shape of discourse is the shape of the interaction itself. It is the space in which interlocutors make meaning and progress toward discourse goals together. In my proposed terms, this is the Interaction Space. The Interaction Space is the space between interlocutors in a face-to-face conversation against which interactive meaning in gesture is assessed. The Interaction Space is used as both the physical and metaphoric space in which participants work to build a coherent discourse. It is physical in that participants attend to each other and the communicative task at hand by maintaining the space with their bodies. The space is metaphoric in that the actions participants take in relation to the space are interpreted as actions taken upon the discourse itself, which is an abstract rather than concrete entity (see Chapter 3, Section 4). The ability to act upon metaphoric objects via concrete observable actions is a unique strength of the gestural mode. A theory of gesture meaning that is concerned with *how* gestures mean, not just *what*

they mean, should center this ability. I use the Interaction Space, as well as insights from cognitive linguistics and observational gesture studies, to build such a theory.

The Interaction Space as a formal model for interactive meaning in gesture is grounded in the simple observation that interlocutors must engage in a shared space to maintain a face-to-face interaction (e.g. Gill et al. 2000; Kendon 2010). This is perhaps most obvious when a social task involves mutual interaction with an object. For example, when playing a board game, the game must be placed centrally in a shared space such that all participants have access. In order for the board game to proceed, participants must orient toward this space and interact with the objects therein. The Interaction Space extends these observations of cooperative object-oriented interactions to conversation. Where the primary social task is conversation, it is the discourse itself that becomes the metaphoric object towards which participants orient. I call this orientation toward the Interaction Space and the discourse structure therein **engagement management**. Discourse topics and referents are the metaphoric objects within the Interaction Space with which participants interact. I call manipulation of these metaphoric objects **content management**. Engagement and content management are used in tandem to accomplish communicative goals, just as they would in playing a board game.

The Interaction Space is meant to specifically serve as the foundation for a predictive theory of interactive meaning in gesture. The analogy to board games is helpful in understanding what 'predictive' means in this context, and what needs to happen to ensure the Interaction Space does, in fact, make useful testable predictions about the interface between gesture and discourse structure. So, as already stated, in order to play a board game, participants must physically engage with a shared space and act upon the game's physical pieces. In addition to these forms of game management, participants must follow a set of rules. If they want to "win", they also must employ a set of well-reasoned strategies. Given the rules of a game and the strategies that are helpful for winning, players' actions become at least partially predictable. In a traditional turn-based game, we know who is going to interact with the game's pieces and when. We also have a sense of how each participant will move their pieces during their turn; they are likely to act in response to the previous participant's actions and in a way that will increase their chances for success in the game. In the context of the Interaction Space, we are also interested in predicting what a participant will do during their (conversational) turn. In particular, we are interested in what types of interactive gesture appear where, and how they are integrated into participants' overall multimodal discursive strategies. To make these predictions, we have to know the rules for acting upon the discourse's metaphoric objects and we have to have a sense for which types of actions would most benefit participants in achieving discourse goals.

In determining the 'rules' of the Interaction Space, I center the communicative capacities particular to the gestural mode and draw upon cognitive linguistic approaches, specifically Cognitive Metaphor Theory (Lakoff & Johnson 1980) and Mental Spaces Theory (Fauconnier 1994), as they have been applied to gesture analysis. In considering what types of actions would most benefit the discourse, I rely on basic principles of information structure (Krifka 2008) and question-based approaches to discourse structure (van Kuppevelt 1995; Roberts

1996, 2012), as described in Chapter 3. In making predictions about particular gesture forms, I use previous observational work on interactive gesture (e.g. Bavelas et al. 1992; Bavelas 1994; Bressem & Müller 2014; Kendon 1995; McNeill et al. 1993; Müller 2004; Streeck 2009; Wehling 2017).

Following general principles of cognitive linguistics (Croft & Cruse 2004), I hope for the Interaction Space and its associated model for interactive meaning to be grounded in actual language use and to be cognitively *plausible*. However, I make no pretense at its cognitive *reality*. I do not mean to argue that the Interaction Space is in some way how interactive meaning is represented 'in the brain'. I mean only to articulate a theory that will prove helpful in understanding how gesture systematically and predictably contributes to discourse structure and coherence. In particular, I argue that understanding the Interaction Space as an embodied discourse structure allows one to make predictions as to the types of gesture participants will use in particular discourse structural contexts. The ability to make such predictions can allow gesture analysts to test their interpretations of particular gestures, and discourse analysts to test their understanding of particular discourse structures. In knowing what to expect, we are able to more clearly identify the unexpected and exceptional, and can then adjust our theories accordingly.[1]

After giving a brief overview of previous work on gesture position (Section 4.2), I will define the Interaction Space in detail, discussing its delineation, affordances, and motivations (Section 4.3). I then discuss examples from talk show interviews that demonstrate engagement management (Section 4.4) and content management (Section 4.5). Section 4.6 concludes with a summary of predictions as to how the Interaction Space is systematically used in constructing a coherent discourse.

4.2 Traditional approaches to gesture position

The position of a gesture is most often described in relation to the speaker's body, as demonstrated by the predominance of McNeill's (1992; 2005) 'gesture space' in gesture analysis and annotation (e.g. Ladewig 2011; Kipp et al. 2007). This traditional space is roughly a square in front of the speaker's body approximately the size of their chest, and gesture position is described as being central or peripheral relative to this space. Though McNeill's 'gesture space' has proven helpful in gesture analysis and in developing tractable annotation schemes, there are three immediate shortcomings: (i) the sagittal plane (from the speaker's body outward) is left underspecified, (ii) there is no consideration of the particular affordances of a given space (e.g. the presence of a table), (iii) there is no consideration of addressee location.

[1]Though I am primarily interested in the Interaction Space as a *model* for interactive meaning, rather than its cognitive reality, we can look to work on the 'peripersonal space' for evidence of its cognitive plausibility. This body of literature suggests that the space around the body that is immediately relevant for interaction with the world (i.e. space that is *within reach*) is cognitively privileged shaping our perception of the availability, usefulness, and potential dangers of objects. See de Vignemont & Iannetti (2015) and Hunley & Lourenco (2018) for review.

More simply put, McNeill's gesture space fails to recognize the importance of *mutual* space in face-to-face interaction.

McNeill's gesture space has been used to argue for the contribution of position to a gesture's meaning. For example, Ladewig's (2011) analysis of the cyclic gesture uses McNeill's gesture space to differentiate use contexts; 'word-search' uses are associated with the central gesture space, whereas descriptive uses are associated with the peripheral gesture space. Despite this recognition of gesture position as meaningful, meaning relative to a *mutual* space is only regularly recognized for a subset of gestures, namely deictic gestures (e.g. Kita 2003) and interactive gestures (e.g. Bavelas et al. 1992).

Deictic gestures are used to incite joint attention toward a mutually perceivable location or object. This means that the gesturer must consider not only their physical space and visual field, but also that of their addressee. The consideration of both participants has been shown to hold for deictic expressions in both the gestural (pointing) and verbal (demonstratives) modes (see Peeters et al. 2021 for a recent overview). Moreover, there is evidence that speakers also take their addressee's mental states into consideration in their performance of deictic gestures. Deictic gestures that are supplementary and mutually known tend to be smaller and less well-articulated, whereas deictic gestures that provide new information are generally larger, longer, and more clearly articulated (Enfield et al. 2007). That both spatial and social contexts are necessary for interpreting a deictic gesture suggests that a single speaker-centric gesture space is not sufficient for functional analyses.

In interactive gestures, the importance of mutual space is at least as apparent as in deictic gestures. These gestures relate to the interaction itself, the discourse structure, and other forms of pragmatic meaning as it is cooperatively built in the interaction (e.g. Bavelas et al. 1992; Bavelas 1994; Kendon 1995). As such, interactive gestures are necessarily interpreted in the mutual space that is being used to maintain the interaction, and many of the gestures, such as those requesting and offering information (Cooperrider et al. 2018; Müller 2004), explicitly refer deictically to the addressee. As with deictic gestures, the meaning of an interactive gesture cannot be fully interpreted without taking into account the shared space and, in particular, the location of the addressee.

There is also increasing evidence that gestures are fundamentally multifuncitonal, *always* containing deictic and interactive components (Kok et al. 2016; Shattuck-Hufnagel & Prieto 2019; Sweetser 2022). This multifunctionality is demonstrated by the ways in which speakers change general gesturing behavior according to the presence and position of an addressee. First, studies manipulating addressee visibility have repeatedly shown that speakers gesture more frequently and more visibly (e.g. higher and larger) when their addressee is visible (as opposed to, say, on the phone or behind a screen) and attentive (e.g. Alibali et al. 2001; Bavelas et al. 1992; Cohen & Harrison 1973; Holler et al. 2011; Jacobs & Garnham 2007). Second, speakers will change a gesture's trajectory relative to their addressee's location. For example, Özyürek (2002) shows that when producing a primarily representational (i.e. semantic) gesture depicting throwing something 'away', the speaker will ensure that it is away from both their own body and their addressee's. These findings suggest that gestures function via their performance in a mutual space regardless of the 'type' of gesture that is

being performed.

Given the recognition of gesture position as meaningful, and given the accepted relevance of mutual space to the interpretation of deictic and interactive meaning, we are in need of a formalized space other than McNeill's personal gesture space. Though McNeill's personal gesture space is by far the most common and the most formalized, proposals for other spaces are not unheard of. For example, Kendon (2010: 2-3) observes that when describing any embodied experience a much more general 'use-space' must be considered, the space which is immediately relevant to and involved in any given interaction with the world. The use-space in conversation, Kendon (1990; 1992; 2010) argues, is defined and maintained through specific formations of participants used to delineate a *shared-transactional space*, or 'o-space'. However, the use of the 'o-space', and similar proposals in the gesture literature[2], is largely restricted to macro-level interaction. That is to say, this space is discussed when analyzing inclusion and exclusion from an interaction or overall interaction type rather than how individual gesture sequences exploit the space to convey meaning.

Sweetser & Sizemore (2008), in noticing the importance of gesture location to individual gesture function, introduce two mutual spaces, the *interpersonal space* and the *unclaimed space*. Interpersonal space is defined as the space between the individuals' personal gesture spaces. They argue that the interpersonal space is often used in floor-management gestures. For instance, they observe instances where an interlocutor reaches forward, out of their personal space, in order to signal that they want to either take or maintain the floor (*ibid*: 26). They further argue that use of the interpersonal space is limited to interactive gestures, whereas the personal gesture space may be used for both interactive and representational gestures (*ibid: 27*). The unclaimed space is any space (i) not between the two interlocutors, and (ii) outside of the interlocutor's personal space. This may be above, behind, or to the side of the speaker. This space is used primarily for depictions of imagined spaces and actions. Sweetser & Sizemore's proposal is a helpful step in formalizing a mutual space that can be used in gesture analysis. However, maintaining a distinction between the 'personal space' and 'interpersonal space' introduces unnecessary ambiguity in describing a gesture's position – one must always decide whether a gesture is in the 'personal' or 'interpersonal' space, and it is difficult to argue for a clear delineation.

The Interaction Space that I propose combines the function of Sweetser & Sizemore's (2008) 'interpersonal space' with the formal simplicity of Kendon's (2010) 'o-space'. I opt for the term 'Interaction Space' in hopes of making the connection between the space and interactive meaning apparent.

4.3 Defining the Interaction Space

In this section, I clarify the definition of the Interaction Space and discuss its physical and discursive affordances. The physical affordances arise from the realization of the Interaction Space as a literal physical space between participants. The discursive affordances of the

[2]These include Gill et al's (2000) *space of engagement* and Battersby's (2011) *interaction space*.

Interaction Space, that is how participants manage and organize discourse topics, are derived from the space's physical affordances.

The Interaction Space is all space between a pair of participants, subsuming their personal gesture spaces. The distinction between an individual's personal gesture space and the portion of the Interaction Space that overlaps with it is one of function. Representational, or semantic, meaning can be assessed relative to the traditional gesture space. Interactive meaning, on the other hand, is assessed relative to Interaction Space that participants actively maintain and negotiate throughout a discourse. The space is centered along the speaker-hearer line, a virtual line that connects the center of each participants' bodies (Sweetser & Stec 2016). This is schematically represented in Figure 4.1. This space must be actively and cooperatively maintained by the participants. This means remaining in roughly the same location, and orienting toward each other. Doing anything else – turning away, leaning away, walking away – necessarily signals some break in the interaction.

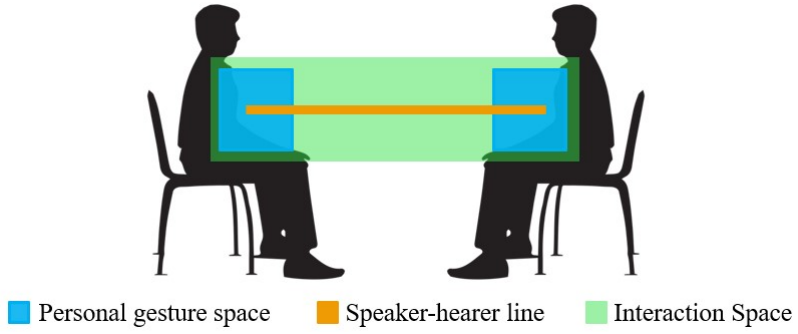


Figure 4.1: The delineation of the Interaction Space

The physical dimensions and manipulation of the Interaction Space are also affected by the affordances of the particular space in which the interaction takes place. For example, an Interaction Space will likely be larger in a park than in a crowded bar. If there is a table between participants, they will likely take advantage of its physical surface when organizing the discourse in ways that wouldn't be possible if they were standing with nothing between them. Across all physical contexts, the Interaction Space is delineated and maintained by the the position of participants. As physical circumstances change, as they are wont to do in our embodied experiences, so too does the optimal positioning of partcipants' bodies in interaction.[3] These changes in circumstance are directly reflected in the physical dimensions and affordances of a given Interaction Space.

Though the entirety of the Interaction Space is considered mutually accessible to both participants, the space immediately in front of each participant can be considered more closely *associated* with that participant. To concretize this, consider the orientation of water

[3]See Kendon (2010) for a nice overview of the ways in which participants' orientations change given different physical and interactive circumstances.

glasses on a table. I would likely place a pitcher of water in the center and then place my water glass quite close to me, and yours quite close to you. We would each drink from the glass closest to us – indeed, it would be very strange and relatively inconvenient to do anything else. However, we may continue to interact with each other's glasses in various ways. For example, I may point to your glass to ask if you want more water. If you say yes, I may then move to pour more water into your glass, perhaps even picking up your glass in order to do so. However, these meaningful regions of the Interaction Space should not be confused with the speakers' personal gesture spaces. Though there very well may be physical overlap, there is not functional overlap. The association of a particular region of the Interaction Space with a particular participant is relative – the space is more associated with me by being further from you. Personal gesture spaces, on the other hand, are absolute, and depend only on the position of a single participant.

In the case of the Interaction Space, participants may be more or less associated with particular topics and referents. We expect these associations to be reflected in where these metaphoric objects are placed and manipulated. This is based in the conceptualization of our attributes, experiences, and ideas as physical possessions (e.g. Dancygier & Sweetser 2014; Lakoff & Johnson 1980), as reflected in our ability to say things like "I *had* a crazy day / cool idea" and "I *lost* my motivation". As such, we expect these metaphoric possessions to be located closest to their metaphoric possessor. For example, if I am contrasting my childhood in the country with your childhood in the city, I may begin by gesturing closer to myself, pointing to various locations in space as I describe the farmhouse and vegetable garden. I am likely to gesture further out, closer to you, as I ask about your apartment and the city park you grew up next to. In this way, the space itself becomes meaningful. To understand the interactive meaning of a gesture it becomes necessary to understand the meaning embedded in the layout of the Interaction Space.

The centrality of the speaker-hearer line also lends inherent meaning to regions of the Interaction Space. It is this line that that participants use to directly attend to each other during an interaction. The most obvious way participants do this is through mutual gaze, i.e. moments when the participants' eyes meet. Mutual gaze has been shown to elicit listener backchannels such as head nods and short vocalizations like "yeah" and "mmhmm" (Bavelas et al. 2002). It also serves as a cue for turn-transitions (Beattie 1981; Chen & Harper 2009; Jokinen et al. 2013). The virtual line maintained in these moments overlaps directly with the speaker-hearer line. Speakers also make manual gestures along the speaker-hearer line to elicit and offer responses (Bavelas et al. 1992). These gestures include extending a palm-up open hand toward the interlocutor as if to offer an object for inspection (Cooperrider et al. 2018; Müller 2004), and reaching toward the interlocutor in a request to take the floor (Sweetser & Sizemore 2008; Wehling 2017).

The salience of the speaker-hearer line in interactive gesture contributes to 'centrality' as a meaningful component of the Interaction Space. There is a meaningful, and thus functional, distinction between the space's physical center and its peripheries.[4] The center is used in

[4]This is also the case in McNeill's personal gesture space.

negotiating the structure of the interaction, and also to signal the importance of particular topics and referents to discourse goals. Topics that are of immediate and continuing concern are located centrally, whereas asides are located in the periphery. The analogy to board games is once again helpful in understanding the embodied motivation for this distinction. In setting up a board game, the board is situated in the center of the table, as mutually accessible to all participants as possible. This positioning is functional – the board is central to the game's goals, and it is the object that all participants must regularly interact with. Other game objects, such as cards and tokens, are located along the peripheries. This positioning is once again functional – these are objects that participants interact with as individual players, and may only interact with at certain points in the game. If there are game objects that require even less interaction, say ones that are only used once or only in calculating points at the end of the game, these are located even further in the periphery and may not be physically accessible to all participants. In the case of the Interaction Space, we expect topics that are central to discourse goals and that are being recurrently referred to by both participants to be located, as metaphoric objects, in the center of the Interaction Space. As in board games, this is because it is these main topics that are regularly interacted with by both participants. Topics that do not require continued attention by participants, such as asides and clarifications, are expected to be located, as metaphoric objects, in the Interaction Space's peripheries.

Granting inherent discourse structural meaning to the center and periphery of the Interaction Space aligns directly with discussions of discourse structure, which often implicitly rely on spatial metaphors (e.g. Wehling 2017). In studies of discourse and information structure, people frequently discuss the mutual *accessibility* of topics and referents which determine certain linguistic expressions including pronominal reference, definiteness, and choice of demonstrative (e.g. Clark & Marshall 1981; Givón 1989; Grosz & Sidner 1986; Kamp & Reyle 1993; Lambrecht 1996; Prince 1981). Accessibility is determined by different discourse structural, social, and psychological factors such as recency of last mention, salience, and relevance. The more 'accessible', the more likely pronouns, definite articles, and proximal demonstratives will be used. Of course, in all of these discussions on discourse and information structure, the term 'accessible' is metaphoric, framing topics as physical objects that can be physically perceived and manipulated. In a recent distillation of this work, Evans et al. (2018) relate 'accessibility' to what they call a system of 'engagement', i.e. "a grammatical system for encoding the relative accessibility of an entity or state of affairs to the speaker and addressee" (*ibid*:118). The Interaction Space can be thought of as a physical manifestation of this grammatical system.

The correspondences between the Interaction Space as a physical space on the one hand and as a metaphoric resource on the other can be understood as the result of blending the Physical space in which interlocutors are located with the abstract notion of a Discourse Space (e.g. Wehling 2017). This is a conceptual *blend* (Fauconnier & Turner 1998) rather than a conceptual *metaphor* (Lakoff & Johnson 1980), because (i) some of the mappings are literal correspondences (e.g. interlocutors are literally, not metaphorically, physical agents), and (ii) there are emergent properties that are not reducible to a one-to-one correspondence

between frames. This is represented in Figure 4.2.

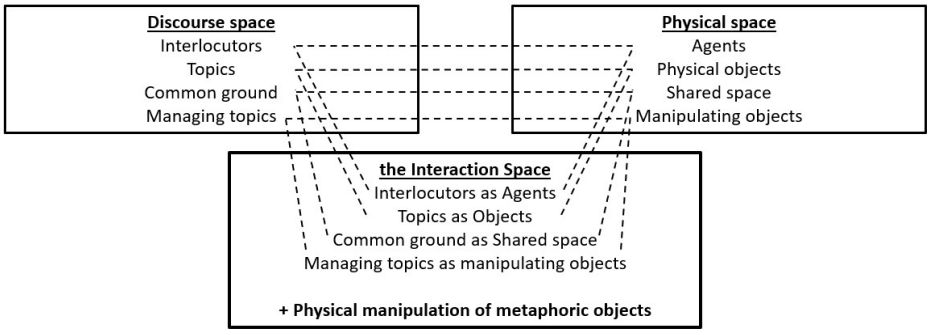


Figure 4.2: The Interaction Space as a blended concept

The emergent property that is central to the use of the Interaction Space as proposed in this work is the "physical manipulation of metaphoric objects". The interactive gestures I look at are *physical* actions. However, the objects those actions act upon are (mostly) *metaphoric*.[5]

In the remainder of this section, I demonstrate the ways in which participants manage interaction and topic accessibility by physically managing the Interaction Space. I then outline a formal representation of the Interaction Space that can be used to systematically particular Interaction Space states and forms of management.

4.3.1 Types of management

Participants make meaning using the Interaction Space in two ways, by managing their **engagement** and managing the space's **contents**. Engagement management relates to participants' roles and conveys primarily social meaning. For example, participants can manage their engagement in the Interaction Space to signal when they intend to contribute, and whether or not they approve of the discourse's direction. This is achieved by changing orientation (e.g. turning away), changing posture (e.g. leaning in or leaning back), and particular hand gestures that are used in requests, such as blocking gestures that metaphorically stop the discourse from proceeding (Wehling 2017; Sweetser 2022). Content management relates to the discourse's structure and conveys information about what is being talked about, when, and how. This is achieved through hand gestures that refer to and manipulate metaphoric objects in the Interaction Space.

Speakers manage their *engagement* with the Interaction through body positioning. By orienting their heads, gaze, torso, and manual gestures *toward* an interlocutor, speakers *engage* with the Interaction Space, and signal the start or continuation of a discourse. By

[5]We see one exception to this, in which the gesture performs a physical action upon a *physical* object, in Section 4.5.1.

moving or turning away from an interlocutor, speakers *disengage* from the Interaction Space, signalling a desire to discontinue the present discourse. The example given in Figure 4.3 demonstrates a process of engagement, disengagement, and subsequent re-engagement with the Interaction Space. In the clip, late night comedian Stephen Colbert is interviewing American actor Alicia Silverstone.[6] The first panel shows Colbert and Silverstone maintaining the speaker-hearer line by orienting their gaze, head, and torso directly toward each other. Both participants signal heightened engagement by also leaning into the Interaction Space, decreasing the space between them and making any actions more easily accessible to the other. In the second panel, Silverstone disengages from the Interaction Space by leaning back, and breaking the speaker-hearer line by turning her head and gaze away from Colbert. In the final panel, Silverstone signals re-engagement by once again leaning into the Interaction Space and reestablishing the speaker-hearer line by returning her gaze to Colbert.

Figure 4.3: Disengagement with the Interaction Space and breaking of the speaker-hearer line.

This form of management via the Interaction Space is further discussed and exemplified in Section 4.4. As I will show, managing engagement can convey a range of interactive meaning, allowing speakers to co-ordinate turns, indicate heightened attention, signal attitudes toward the main discourse, and attend to secondary interactions.

Speakers manage the *contents* of the Interaction Space by introducing, locating, and removing physical and virtual objects that are under discussion. Any discourse topics established within the space are available to all interlocutors, including those established within an individual's personal gesture space. Figure 4.4 exemplifies a particular type of content management which I call Refer. In this clip, Colbert is interviewing American journalist John Dickerson about strategies in the 2020 Democratic primary race.[7] Colbert asks a question about why democrats seem to avoid talking about billionaires when there are so few of them, seeming to underline a figure on a script in front of him. The position of the script is shown in the first panel, and Colbert's underlining gesture is shown in the second. Dickerson then refers the same location when contributing details about the identity of the billionaires, as depicted in the third panel. In this case there is a physical object in the Interaction Space (the script) metonymically representing the ongoing discourse topic (billionaires in Amer-

[6]UID:b7d27c2c-7c2f-11e8-938d-089e01ba0770,2799

[7]UID:465f22fa-01fa-11ea-8c13-089e01ba0335,1987

ica). The deictic gesture used by Dickerson manages the content of the Interaction Space by inciting joint attention toward the topic to which information is being added.

Figure 4.4: Locating topics within the Interaction Space

This particular form of content management is a clear example of 'abstract deixis' in which participants establish a location for a topic and repeatedly refer to the location through explicitly deictic gestures (McNeill 2003; McNeill et al. 1993). Other forms of content management in which topics are *presented, separated* and *removed from* the Interaction Space also contain deictic elements that locate the given topic relative to the Interaction Space. However, many of these content management gestures also foreground non-deictic meaning, such as managing discourse goals, establishing relationships between topics, and expressing speaker attitude toward particular topics. These variants of content management are further discussed and exemplified in Section 4.5.

4.3.2 Formal representations

The Interaction Space can be modeled as a tuple that is defined by participant pairs, contains active discourse topics, and is managed through gestural actions. This is represented in (32). Much like in representations of Common Ground (e.g. Murray 2014) which track the shared knowledge states of discourse participants, every discourse move relates to a distinct state of the Interaction Space (IS_t). The participants are the two people responsible for physically maintaining the Interaction Space and are indexed for their role as speaker (S) or addressee (A) in each IS state.[8] The content is a list of topics and referents that are immediately under discussion.[9] Management is the list of actions taken by a participant in the Interaction Space during a particular discourse move. Each action is specified for the gesture that expresses it, and, where relevant, the particular topic that the gesture is acting upon. The presence of an arrow ("→") signals that a given gesture (specified to the arrow's left) is acting upon a particular discourse topic (specified to the arrow's right).

[8]Because the Interaction Space is anchored around the speaker-hearer line, it is always between exactly two participants. This means that in multiparty interactions an Interaction Space exists for each pair of participants. See Section 4.4.3 for further discussion.

[9]When integrating this model with Question Under Discussion the contents could be listed as QUDs. In this chapter, I will list topics as 1-3 word paraphrases.

$$(32) \quad IS1_{\mathrm{t}} = \begin{cases} \text{Participants} : Participant_{\mathrm{S/A}},\ Participant_{\mathrm{S/A}} \\ \text{Content} : Topics \\ \text{Management} : Action : (Gesture \rightarrow Topic) \end{cases}$$

Changes in participant roles are represented by "+/-" symbols. For example, when an action enables an addressee to become a speaker, this change in the Interaction Space is represented as "+S" in the participant list. Where an action signals that a speaker is intentionally ceding the floor, this is represented as "-S" in the participant list. When content is added to the Interaction Space, a "+" appears in front of the introduced topic. When content is removed, the relevant topic appears crossed out (e.g. ~~Topic~~).

Unlike representations of the Common Ground which keep an ongoing record of the entire discourse, the contents of the Interaction Space are limited to those topics that are under immediate discussion. It is only active discourse topics that are accessible to participants; closed topics must be reintroduced. When topics are introduced to the discourse, they are immediately considered as a part of the contents of the Interaction Space, regardless of whether or not they were explicitly introduced in the gestural mode. When discourse topics are closed, they are considered no longer present in the Interaction Space, and thus are not accessible for further reference or management. Again, this is the case regardless of whether the topic was removed via gesture. Actions that participants take can either relate to the discourse as a whole, in which case they are managing their engagement as participants, or to particular topics, in which case they are managing content. Participants often employ multiple strategies at once. I will argue that this allows for composite actions in which multiple concurrent gestural strategies combine compositionally to achieve a form of management that is distinct from the management achieved by its individual components.

Using formal representations of the Interaction Space also help in systematically identifying predictions as to participant actions. Each state of the Interaction Space restricts the actions available to interlocutors in the subsequent state. For example, if a topic is removed from the Interaction Space at t_1, then an action involving that topic cannot be performed at t_2, unless the action serves to reintroduce the topic to the space. There are also a particular set of actions called REQUESTS that are meant to elicit particular actions by the interlocutor. These REQUEST actions are modeled as requesting a particular IS state (IS_{t+1}) that must be enacted by the other participant. The prediction in these cases is that a cooperative participant will, in fact, enact the desired state.

In the following section, I propose two actions related to managing participant roles: ENGAGE, DISENGAGE. In Section 4.5, I then focus on four actions used to manage Interaction Space contents: PRESENT, REMOVE, REFER, SEPARATE. Each action is expressed by at least one gesture. REQUEST actions are discussed in both sections, as they are composed of gestures that enact both types of management. The connections between particular gestures and particular actions are motivated by recurrent observations in the gesture literature, cognitive linguistic theories of primary metaphors, and, where relevant, mental space organization and navigation. In the following two sections, an example of each action type will be given with a transcript, set of accompanying screenshots, proposed formal representation,

and a discussion regarding interpretation and motivation. Each gesture sequence is listed as a sequence of two letters and a number. The first letter corresponds to the gesturer's last name; the second letter is either a "G" for manual gestures, or "P" for non-manual changes in posture and orientation. The number corresponds to where the gesture occurs in the discourse excerpt. For example, in an interaction between Stephen Colbert and politician Elizabeth Warren, the second manual gesture Warren performs would be listed as "WG2", whereas the third head orientation Colbert assumes would be listed as "CP3".

4.4 Managing engagement

In the proposed model, managing engagement with the Interaction Space has two primary functions – signaling gesturer intentions regarding turn-structure, and signalling gesturer attitude toward the state of the discourse and the actions of its participants, including their own. As we will see, these two functions often overlap. The proposed functions can be formalized as three distinct management actions: ENGAGE, DISENGAGE, REQUEST. The ENGAGE and DISENGAGE actions are speaker-directed, meaning that they relate directly to the actions the speaker takes in relation to the Interaction Space, and thus the discourse. These actions do not directly act upon the addressee, though the addressee is expected to take these actions into consideration when planning their next move.[10] Example representations of ENGAGE and DISENGAGE actions are given in (33).

(33) a. **Engage**

$$IS1_{\text{t}} = \begin{cases} \text{Participants: P1}_{+\mathbf{S}}\text{, P2}_{\text{A}} \\ \text{Content : } \textit{Topics} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{P1G1}) \end{cases}$$

b. **Disengage**

$$IS1_{\text{t}} = \begin{cases} \text{Participants: P1}_{\text{A}}\text{, P2}_{-\mathbf{S}} \\ \text{Content : } \textit{Topics} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{P2G1}) \end{cases}$$

The representations above relate to gestural expressions of turn-transitions. In (33a), the ENGAGE action serves to signal that the gesturer (P1) is taking the floor. This action is expressed by a particular gesture (P1G1). The floor-taking itself is also represented as a change in participant role ($\text{P1}_{+\text{S}}$). This particular representation assumes a turn-transition point, meaning that P2 has already signaled the end of their turn and thus assumed the role of addressee, represented as P2_{A}. To represent a turn-transition in which the P1's addressee has not yet ceded the floor, the only difference in representation would be that P2 would still have a speaker role (P2_{S}). In such a situation, a cooperative P2 would then perform a DISENGAGE action, such as that represented in (33b). This representation is the inverse of the previous. The DISENGAGE action serves to signal that P2 is ceding the floor, and does so through a particular gesture (P2G1). The floor ceding itself is represented by the change

[10] This expectation is based on basic principles of cooperation in conversation (e.g. Clark & Schaefer 1989; Grice 1975).

in speaker role ($P2_{\text{-S}}$). Again, in the case of an overlapping turn in which P1 has already moved to take the floor, the only difference would be that P1 would already be indexed as a speaker (S).

ENGAGE and DISENGAGE actions can be performed using a number of well-documented gestural strategies. In proxemics, the study of how body position mediates social interaction, these strategies include leaning in and out and reorienting one's body toward or away from an addressee (e.g. Gill et al. 2000; Hagemann 2014; Hall 1995; Kendon 1967, 1990, 2010). In work on turn-taking behaviors, strategies include attaining and breaking mutual gaze, as well as opening the mouth as if to speak (e.g. Chen & Harper 2009; Jokinen et al. 2013; Streeck & Hartge 1992). Engagement management can also include manual gesture cues, including the onset of gesture from a rest position or the return to a rest position (e.g. Kendon 1972, 1980; Streeck & Hartge 1992).

Engagement management gestures can also combine compositionally with another interactive gesture to perform a REQUEST action.[11] REQUEST actions are addressee-directed, meaning that the gesturer's desire for a subsequent Interaction Space state must be brought about by their addressee. For example, when asking for a response to a specific question, a speaker may lean back while reaching toward the addressee with a palm up open hand as if to request an the answer, as a metaphoric object, to be placed in their hand. The DISENGAGE action informs the addressee that the gesturer intends to cede their turn, while the manual gesture presents a particular topic to be attended to. In the proposed terms, this manual gesture would constitute a content management action called PRESENT in which a topic is introduced into the Interaction Space as a metaphoric object accessible to both interlocutors. This discourse move is represented in (34).

In this representation, the DISENGAGE action is expressed by one gesture (G1) and signals that the speaker is ceding their turn ($P1_{\text{-S}}$). The PRESENT action is expressed by another gesture (G2), which acts upon a particular topic in the Interaction Space ($\rightarrow \text{Topic}_1$). In this case, the action is to introduce the topic to the space which is represented as adding the topic to the space's contents ($+\text{Topic}_1$). The combination of these two gestures (G1+G2) requests that the addressee bring about a particular Interaction Space state (IS_{t+1}). The "t+1" subscript signals that the requested state should be the immediately proceeding state, and that the state has not yet been realized. In this case, it is asking them to take the floor (i.e. perform an ENGAGE action), and make reference to the particular topic introduced in IS_1.

[11] Though all of the REQUEST actions discussed in this work are compositional, usually combining engagement and content management strategies, one could argue that mutual gaze could serve as an independent request for the addressee to take the floor. This is based on previously mentioned studies suggesting that mutual gaze is a recurrent cue for turn-transitions (Kendon 1967; Jokinen et al. 2013). However, this previous work does not discuss to what extent these moments of mutual gaze co-occur with other gestural cues. Determining whether REQUEST actions can be performed using a single gestural strategy, like attaining mutual gaze, is worth addressing in future research.

(34) **Request**

$$IS1_{t} = \begin{cases} \text{Participants: P1}_{\textbf{-S}}\text{, P2}_{A} \\ \text{Content: +Topic}_{1} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{G1}) \\ \quad \textsc{Present}(\text{G2} \rightarrow \text{Topic}_{1}) \\ \quad \textsc{Request}(\text{G1+G2})\text{: IS1}_{t+1} \end{cases} \qquad IS1_{t+1} = \begin{cases} \text{Participants: P1}_{A}\text{, P2}_{\textbf{+S}} \\ \text{Content: Topic}_{1} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{G3}) \\ \quad \textsc{Refer}(\text{G4} \rightarrow \text{Topic}_{1}) \end{cases}$$

Hand gestures associated with Request actions include cross-linguistically common interactive gestures such as the palm-up open-hand gesture associated with requesting and offering information (Bavelas et al. 1992; Cooperrider et al. 2018; Ferré 2012; Kendon 2004; McKee & Wallingford 2011; Müller 2004), as well as abstract deictic gestures used to refer to topics as metaphoric objects in space (e.g. Azar & Özyürek 2015; McNeill 2003; McNeill et al. 1993) Request actions can also be used by an addressee to request that the current speaker cede the floor. In previous literature, this often involves reaching toward the current speaker (Sweetser & Sizemore 2008) or holding up an out-turned open palm, in a sense 'blocking' the current speaker from further action (Wehling 2017).

In the remainder of this section, I present examples of engagement management actions as observed in television interviews. I include engagement management actions performed within a single Interaction Space (Sections 4.4.1 & 4.4.2) and across Interaction Spaces with different co-present and imagined participants (4.4.3).

4.4.1 Engagement at turn transitions

The first example shows a sequence of turn-transitions between interviewer (Stephen Colbert) and interviewee (Elizabeth Warren). The dynamic of this interaction limits the interviewer largely to Request actions, asking the interviewee to provide particular information. Conversely, the interviewee is largely limited to self-directed engagement actions, offering information as answers to the interviewer's questions. The participants use different strategies in managing their engagement, but converge when managing the Interaction Space's contents.

In this clip, Colbert is asking American politician Elizabeth Warren a series of questions about how she would handle situations as president. In this particular portion of the discourse, Colbert asks about a sensitive foreign policy issue considering violence in the Middle East (line 1-2). Instead of directly answering the question Warren focuses on "where we are right now" (lines 4-5), making reference to the inappropriate response then-president Donald Trump had made (lines 6-9). She begins the turn by looking at Colbert and pointing to the center of the Interaction Space (WP1,WG1). As she provides her answer, she then also engages with the audience, directing her head and gaze to the front of the stage. As she finishes the palm-down gesture correlated with her emphatic negation[12] (line 9, WG2),

[12]This gesture in which a flat open hand with the palm facing down is swept across the speaker's gesture space has been observed recurrently with expressions of negation in Iberian Spanish speakers (Harrison 2010)

Colbert raises his hand sightly from the desk to take the floor and ask a question about her past political party affiliations. As Colbert provides necessary background information (line 10), he leans in sharply (CP2). When actually asking the question and ceding the floor, he leans back into a neutral position. Warren begins her next turn by first pointing into the Interaction Space (line 12, WG3) and directing her gaze toward her gesture (line 12, WP4). She then looks back up toward Colbert (WP5) and continues her explanation.

(35) Transcript 1: Elizabeth Warren
[UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335,2143]

```
1   SC   Okay, Iran versus the Saudis, what would convince President
2        Warren that the Iranians did this, and if so, what would the
3        response be?
4   EW   Okay, so look, uh let's be really clear, we don't have the
5        kind of evidence we need right now, let's just talk about
6        where we are right now.  And no president gets to declare war
7        on their own.  In this case what we need- (.)  He wants to
8        talk about bombing somebody, you gotta come to congress and
9        get an authorization for the use of military force.  △That's
                                                              △WP1->
10       what the constitution says.  *Nobody gets to drop those*
         <- - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
                                        *WG1- - - - - - - - - - -*
11       △bombs on their own.  *No.* △
         △WP2- - - - - - - - - - - -△WP3>
                               *WG2-*
12 SC    △*You were a republican for a long time, you switched △in
         △CP2- - - - - - - - - - - - - - - - - - - - - - - - - - △CP3>
          *CG1- - - - - - - - - - - - - - - - - - - - - - - - - - ->
13       △'96.* *Did you become an independent or a democrat at that
         △CP4- - - - - - - - - - - - - - - - - - - - - - - - - - - ->
         <- -* *CG2- - - - - - - - - - - - - - - - - - - - - - - - ≫
14       time?
15 EW    *No, △I've been actually a republican△ and an independent*
         <- -△WP4- - - - - - - - - - - - - - - △WP5- - - - - - - - >
         *WG3- - - - - - - - - - - - - - - - - - - - - - - - - - - *
16       *because I just wasn't that political*, △it's how I
         <- - - - - - - - - - - - - - - - - - - - - - - - - - - - △
         *WG4- - - - - - - - - - - - - - - - - - - - - - - - - -*
17       registered to vote.
```

and Italian speakers (Kendon 2004, ch. 13).

```
WP1, WP3, WP5:  head and gaze toward Colbert
WP2:  head and gaze toward audience
WP4:  head toward Colbert, gaze down toward gesture
WG1:  right hand, index-finger point to central IS
WG2:  right hand, open palm down, lateral sweep across body space
WG3:  right hand, index-finger point down to desk
WG4:  both hands, palm up open hand shrug
CP1, CP4:  head and gaze toward Warren, leaned in slightly
CP2:  head toward Warren, gaze down toward gesture, leaned far in
CP3:  head and gaze toward Warren, back straight
CG1:  right hand, loose index-finger point toward Warren, beats
      throughout
CG2:  right hand, loose index-finger point to desk, beat right then
      left
```

Figure 4.5: Gesture sequence corresponding to Transcript 1: Elizabeth Warren, ex. 35.

In this example we see both manual and non-manual forms of engagement management. Colbert moderates his posture, leaning in and away, as he takes his turn and subsequently passes the floor to Warren. Colbert's pointing gestures signal both the start of his turn (CG1) and contribute to content management by referring to particular locations within

the Interaction Space (CG2). Similarly, Warren uses pointing gestures to refer to a central location in the Interaction Space in both of her turns (WG1, WG3). Her non-manual forms of engagement management are more complex than Colbert's as she alternates between engaging with Colbert and the audience.

The Interaction Space states provided below reflect the turn-transition at lines 11-12. Colbert cedes the floor while performing two distinct engagement actions. First, he disengages by leaning away, signalling that the most important portion of his turn is complete and he is ready for Warren to contribute to the discourse. His second pointing gesture (CG2) performs a REFER action, contributing to a request for a subsequent state (IS_{t+1}) in which Warren takes the floor and contributes to a particular discourse topic, namely why she switched political party affiliations. This gesture indicates a space closer to Warren than Colbert's previous pointing gesture. This location is unsurprising given that the question relates to Warren's personal choices and history, objects that can be considered metaphorically as her personal possessions. Colbert holds his hand in place until Warren provides a full answer, perhaps as an attempt to maintain joint attention on the topic and prevent Warren from digressing.[13]

Warren then performs the requested actions, taking the turn and directly addressing the question posed. First, she signals a continuation of the proposed discourse topic by pointing to the same region of the Interaction Space that Colbert has just pointed to (WG3). She looks at her gesture as she does so, drawing joint attention to the gesture and thus her uptake of the particular discourse topic. After having taken the turn and confirmed her intent to address the question posed, Warren then looks up at Colbert to continue her turn.

(36) **Management actions by Colbert, ex. 35, Line 11**

$$IS1_{t} = \begin{cases} \text{Participants: } SC_{\text{-S}}, EW_{A} \\ \text{Content : Party affiliation} \\ \text{Management :} \\ \quad \textsc{Disengage}(CP3) \\ \quad \textsc{Refer}(CG2) \\ \quad \textsc{Request}(CP3{+}CG2)\text{: } IS1_{t+1} \end{cases} \qquad IS1_{t+1} = \begin{cases} \text{Participants: } SC_{A}, EW_{+\mathbf{S}} \\ \text{Content : Party affiliation} \\ \text{Management :} \\ \quad \textsc{Engage}(WP) \\ \quad \textsc{Refer}(WG\rightarrow \\ \qquad \text{Party affiliation}) \end{cases}$$

(37) **Management actions by Warren, ex. 35, Line 12**

$$IS1_{2} = \begin{cases} \text{Participants: } SC_{A}, EW_{+S} \\ \text{Content : Party affiliation} \\ \text{Management :} \\ \quad \textsc{Engage}(WP5) \\ \quad \textsc{Refer}(WP4,WG3\rightarrow\text{Party affiliation}) \end{cases}$$

The representations above provide a systematic way to specify the contribution of each gesture to discourse management. Colbert's disengagement from his role as speaker is expressed

[13]Kendon (1995) observes similar behavior during question-answer sequences in Southern Italian speakers.

by the non-manual CP3 gesture. His request for Warren to address a particular topic is expressed compositionally by the disengagement gesture (CP3) and the manual deictic gesture used to locate a particular topic (CG2). The REQUEST action requires both of these actions – Colbert must be willing to cede the floor to Warren, *and* the topic that he wants Warren to address must be specified. The requested Interaction Space state is represented by IS_{t+1}. In the requested state, Warren performs two management actions: an ENGAGE action to begin her turn, and a REFER action specifying the particular ongoing topic that Colbert wants her to address. This is exactly what she does, as represented in (37).

4.4.2 Signalling attitude

The two examples in this section involve nearly identical gestures which seem, on the surface, to perform opposite functions. In both cases, the gesturer holds up a flat open hand oriented away from their body, as if to hold back an object from entering their immediate space.[14] In the first case, the addressee performs the gesture while requesting the speaker to stop an action. In the second, the speaker says something they shouldn't have said and uses the gesture to, in a sense, stop their own action. In both cases, the gesture contributes to a REQUEST action directed at the current speaker.

In the first clip, English actor Eddie Redmayne is telling Colbert about an accident that happened on the set of his recent film. He digresses from his main story by clarifying that the accident had not resulted in injury. During this digression, Redmayne dramatically holds up his right index finger, first toward Colbert (RP1, RG2) then sweeps it across the audience (RP2, RG2) as if to ask everyone to "wait a second" as he corrects a possible misunderstanding. While still holding the "wait a second" gesture, Redmayne briefly looks back to Colbert (RP3) then repeats the rightward sweeping motion (RG3). After this moment of mutual gaze, Colbert speaks over Redmayne, asking him not to give away any "spoilers" as to what actually happens in the film (line 4). During this interruption, Colbert performs a blocking gesture (CG1) and leans back slightly (CP2), increasing the distance between him and Redmayne. Redmayne returns his gaze to Colbert to acknowledge the request and apologizes (line 5, RP5), then performs an emblematic 'facepalm' to signal embarrassment (RG4). Redmayne then returns to the main topic, turning back to Colbert (RP7) and holding up an open hand to offer new, non-controversial, information (RG5).

(38) TRANSCRIPT 2: EDDIE REDMAYNE
[UID:bf03ea3c-1731-11ea-a000-089e01ba0335,2088]

```
1  ER   △*a lot of the film was then shot on green screen but we had*
        △RP1- - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
         *RG1- - - - - - - - - - - - - - - - - - - - - - - - - - - -*
```

[14] This gesture is referred to as the 'holding away' variant in the 'away' gesture family and is associated with refusal and rejection (Bressem & Müller 2014, 2017). A nearly identical form has also been fully conventionalized as a sign meaning roughly "don't do it" or "stop" in both American Sign Language (https://www.handspeak.com/) and British Sign Language (https://www.signbsl.com/).

```
2       *and she was totally fine △ by the way no Felicities* *were
        <- - - - - - - - - - - - - △RP2- - - - - - - - - - - - - - ->
        *RG2- - - - - - - - - - - - - - - - - - - - - - - - - - - * *RG3->
3       △ hurt in the making △ of [this movie* △ *um △ sorry*]
        △RP3- - - - - - - - -△RP4 - - - - - - △RP5- △RP5- - ->
        <- - - - - - - - - - - - - - - - - - - *   *RG4- - - - *
4  SC   △                      △*[no spoilers no spoilers]*△
        △CP1- - - - - - - - - -△CP2 - - - - - - - - - - - -△CP3->
                               *CG1- - - - - - - - - - -*
5  ER   uh △ *but it meant we had the* um we had the memory of that
        <-△RP7- - - - - - - - - - - - - - - - - - - - - - - - - - - ->
              *RG5- - - - - - - - - -*
```

RP1, RP3, RP5: head and gaze turned toward Colbert
RP2, RP4: head and gaze turned toward audience
RP6: head resting in hand, gaze down
RP7: head turned toward Colbert, gaze at gesture then Colbert
RG1: both hands, palm up, loosely open, central gesture space
RG2: Right index finger pointed up, held toward Colbert, lateral sweep right; left hand on desk, index finger pointing up
RG3: Right index finger pointed up, lateral sweep right left then right
RG4: left hand to forehead expressing embarrassment
RG5: loose palm up open hand toward Colbert
CP1, CP3: head and gaze turned toward Redmayne, neutral lean
CP2: head and gaze turned toward Redmayne, slight lean back
CG1: both hands open palm out; right hand held up toward Redmayne; left hand on desk

Figure 4.6: Gesture sequence corresponding to Transcript 2: Eddie Redmayne, ex. 38.

In this example we see both manual and non-manual forms of engagement management, often working in tandem with one another. Colbert performs two concurrent disengagement actions while maintaining the speaker-hearer line. First, he distances himself from the Interaction Space and the contents therein by leaning back (CP2). Colbert also creates a boundary between himself and the Interaction Space through a two-handed blocking gesture (CG1). This blocking gesture can be interpreted in two, non-mutually exclusive ways. First, Wehling (2017) argues for the conception of speaking as forward motion, based on the primary metaphor of any action being conceptualized as forward motion (Lakoff & Johnson 1980). Given this, we can interpret Colbert's blocking gesture as stopping Redmayne from metaphorically moving forward in the discourse. Second, we can interpret the combination of Colbert's blocking gesture and leaning away as Colbert distancing himself from the undesirable topic (spoilers), conceptualized as a metaphoric object within the Interaction Space. The representation in (39) specifies the second interpretation, but does not rule out the first. Because Colbert does not gesturally or verbally signal that he wishes to take the turn, the REQUEST action is interpreted as requesting a content management action of Redmayne. In this case, Redmayne is asked to remove the undesired topic from the Interaction Space, and thus the discourse. This is represented as a REMOVE action (RG → Spoilers) and the "Spoilers" content being crossed out.

(39) **Management actions by Colbert, ex. 38, Line 4**

$$IS1_{t} = \begin{cases} \text{Participants: } ER_S, SC_A \\ \text{Content : Spoilers} \\ \text{Management :} \\ \quad \textsc{Disengage}(CP2) \\ \quad \textsc{Disengage}(CG1 \rightarrow \text{Spoilers}) \\ \quad \textsc{Request}(CP2{+}CG1)\text{: } IS1_{t+1} \end{cases} \quad IS1_{t+1} = \begin{cases} \text{Participants: } ER_S, SC_A \\ \text{Content : } \sout{\text{Spoilers}} \\ \text{Management :} \\ \quad \textsc{Remove}(RG \rightarrow \text{Spoilers}) \end{cases}$$

After Colbert's request, Redmayne performs his own complex disengagement sequence. First, he breaks the speaker-hearer line, turning his head away from Colbert and looking down (RP6). While doing this Redmayne performs an emblematic facepalm gesture, bringing his hand to his forehead (RG4). The facepalm (RG4) and verbal apology (line 3) signal that Redmayne has not only agreed to stop pursuing the current topic, but also acknowledges and accepts Colbert's negative assessment.

Finally, Redmayne also disengages with the Interaction Space between him and Colbert in order to interact directly with the audience (RP2, RP4). This navigation between different Interaction Spaces will be discussed in detail in Section 4.4.3.

In the second example, American actor Jennifer Lawrence is telling Colbert about how spies supposedly tell their children that they're spies. She stops herself abruptly after saying 'he', dramatically signalling that she doesn't mean to give away any information about the spy's identity. At her self-correction (line 2), Lawrence performs a blocking gesture (LG2), holding up her left hand, palm facing outward. As she does so, she looks away furtively (LP3). Colbert begins to take the turn (line 4), but Lawrence continues (line 5). There is then a series of overlapping utterances, as it remains unclear whose turn it is (lines 3-14). Colbert confirms her turn verbally, saying "go head" (line 14) and then "I understand" (line 16) after Lawrence doesn't return immediately to the main discourse. Lawrence finally resumes her story about how spies supposedly tell their children that they're spies (lines 16-20), signalling the return with the discourse marker "anyway" (line 17) and another blocking gesture (LG5), this time with small lateral shakes as if to wipe off a vertical surface.

(40) Transcript 3: Jennifer Lawrence
[UID:22e36ace-3709-11e8-9901-089e01ba0770,2058]

```
1   JL   △and so I was like how do you tell your kids that you're
         △LP1- - - - - - - - - - - - - - - - - - - - - - - - - - - - >
2        spies △*and what he told me* △ *or what somebody told me
                 △LP2- - - - - - - - - -△LP3 - - - - - - - - - - - ->
                  *LG1- - - - - - - - *   *LG2 - - - - - - - - - - >
3        [(laughs)]
         <- - - - ->
4   SC   [don't]
5   JL   [I heard]
         <- - - - ->
6   SC   [don't give] don't give away
7   JL   through the grapevine*
         <- - - - - - - - - - *
8   SC   yes uh-huh
9   JL   △*I'm [gonna give] away our government secrets
         △LP4- - - - - - - - - - - - - - - - - - - - - - - ->
          *LG3- - - - - - - - - - - - - - - - - - - - - - - >
10  SC         [yeah]
11  JL   [I'm not Donald Trump]*     [um]
         <- - - - - - - - - - *
12  SC   [just use a]          use [a name like] Onald-day Ump-tray
13  JL   [(laughs)]
14  SC   [go head]
15  JL   △*no I wouldn't give our country's secrets away
         △LP5- - - - - - - - - - - - - - - - - - - - - - - - - ->
          *LG4- - - - - - - - - - - - - - - - - - - - - - - - - >
16  SC   I understand
17  JL   and I'm me* △ *um anyway* △so the-the he basically said
         <- - - - - △LP6- - - - - △LP7 - - - - - - - - - - - - ->
         <- - - - *   *LG5 - - -*
18       △*that they like when they're at a certain age they take
         △LP7- - - - - - - - - - - - - - - - - - - - - - - - - - - - ≫
          *LG6- - - - - - - - - - - - - - - - - - - - - - - - - - - -≫
19       them to the museum, the spy museum, they show them Spy Kids,
20       which is a great film, and (.)  then they take them through
21       the spy museum and then they're like 'we're spies'
LP1, LP7:  head resting on left hand, gaze down
```

```
LP2:  head toward Colbert, gaze off-center
LP3:  head and gaze turned right, away from Colbert
LP4, LP8:  head and gaze toward Colbert
LP5:  gaze toward Colbert, head leaned forward
LP6:  head toward Colbert, gaze down, small head shakes
LG1:  left hand, loose palm up, toward IS
LG2:  left hand, open palm out toward IS, repeated 3x
LG3:  left hand, loose palm down, flips up and out
LG4:  left hand to chest, palm in, fingers spread, repeated 2x
LG5:  left hand, open palm out toward IS, small lateral shake
LG6:  left hand, open palm up, toward IS, repeated 2x
```

Figure 4.7: Gesture sequence corresponding to Transcript 3: Jennifer Lawrence, ex. 40.

Lawrence uses two strategies of disengagement as she attempts to recover from her self-assessed mistake. First, she breaks the speaker-hearer line by turning her head and gaze away from Colbert (LP3). Second, she distances herself from the interaction by enacting a boundary between herself and the Interaction Space, holding her left hand up, palm facing out, as if to hold something away from her body. I argue that both of these disengagement

strategies are self-directed – she has contributed something to the discourse that she shouldn't have, and she responds to her mistake by disengaging from her contribution.[15] This is represented in (41). As in Redmayne's case, the non-manual disengagement acts upon the discourse as a whole, represented as DISENGAGE(LP3). The manual blocking gesture acts upon the particular negatively assessed topic, represented as DISENGAGE(LP2→Spy_1). The content is listed as "Spy_1" in order to highlight that the contribution Lawrence wishes to cancel concerns the identity of a particular spy. The DISENGAGE(LP3) action signals that the current speaker intends to cede the turn. The DISENGAGE(LP2→Spy_1) action signals a request for the current speaker to remove the negatively assessed topic. Note that these two actions are contradictory in that the first predicts that Lawrence will *not* make a subsequent action, whereas the second requests that she does.

(41) **Management actions by Lawrence, ex. 40, Line 2**

$$IS1_{t} = \begin{cases} \text{Participants: } \text{JL}_{\text{S}}, \text{SC}_{\text{A}} \\ \text{Content : } \text{Spy}_1\text{; Spy parenting} \\ \text{Management :} \\ \quad \text{DISENGAGE(LP3)} \\ \quad \text{DISENGAGE(LG2} \rightarrow \text{Spy}_1\text{)} \\ \quad \text{REQUEST(LP3+LG2):IS1}_{t+1} \end{cases} \quad IS1_{t+1} = \begin{cases} \text{Participants: } \text{JL}_{\textbf{-S}}, \text{SC}_{\text{A}} \\ \text{Content : } \sout{\text{Spy}_{\text{T}}}\text{; parenting} \\ \text{Management :} \\ \quad \text{REMOVE(LG} \rightarrow \text{Spies)} \end{cases}$$

Though Lawrence intends, it seems, to continue her turn, the DISENGAGE action still triggers a kind of turn-transition, resulting in Lawrence and Colbert repeatedly speaking over each other. Lawrence intends her disengagement to be metonymic – she is disengaging from a particular part of the discourse, namely her undesirable contribution, *not* the discourse as a whole. However, the gestures she uses to do so signal complete disengagement. The ambiguity is resolved in context, but still causes a brief derailment in the discourse as the floor is re-negotiated. Recognizing the disconnect between the predicted IS state and Lawrence's actual discourse move helps to understand why a derailment occurs.

Lawrence performs a similar disengagement sequence several lines later as she returns to her story (line 17, LP6 & LG5). This disengagement does *not* result in a derailment despite its formal similarities. I suggest two primary reasons for this. First, the non-manual disengagement is less effortful than in the first disengagement; Lawrence is still turned toward Colbert, only her gaze is averted. Second, the sequence is performed with the discourse marker "anyway" which signals the resumption of a topic (e.g. Ferrara 1997), in this case a topic which Colbert has already urged her to return to.

To summarize the above forms of engagement management, let us directly compare the blocking gestures performed by Lawrence and Colbert. As seen in Figure 4.8 below, there are important formal differences in the disengagement strategies used. Lawrence disengages by breaking the speaker-hearer line (turning hear head and gaze away), whereas Colbert maintains the speaker-hearer line but increases the distance between himself and Redmayne by

[15] We may infer a secondary addressee-directed request to not uptake the aborted topic.

leaning back. There are also contextual differences in (i) whose turn it is, and (ii) who opened the negatively assessed topic. In the case of Lawrence's blocking gesture, it is Lawrence's turn, and it is Lawrence herself who introduced the negatively assessed topic. In the case of Colbert's blocking gesture, it is his interlocutor's turn, and it is his interlocutor who introduced the negatively assessed topic. There is also an asymmetry in who is expressing negative assessment. In Lawrence's case, Lawrence negatively assesses her own contribution; there does not have to be a renegotiation of discourse structure. In Colbert's case, Colbert negatively assesses his interlocutor's contribution. This means that the discourse goals need to be renegotiated; Redmayne can either cooperatively close the goal he has set (which he does), or he can decide to try to negotiate its continuation.

Figure 4.8: Blocking gestures by Jennifer Lawrence (left; ex. 40) and Stephen Colbert (right; ex. 38)

In this way, we can say that both blocking gestures 'mean' the same thing, namely signalling the negative-assessment of an active topic. Coupled with the non-manual disengagement actions, both blocking gestures contribute to the same request – remove the negatively assessed topic from the Interaction Space. Whether this request is self-directed (as with Lawrence) or addressee-directed (as with Colbert to Redmayne) is determined by the discourse and gestural context. That it is Redmayne's turn, that it is Redmayne who contributed the unwanted topic, and that Colbert maintains the speaker-hearer line all contribute to the understanding of Colbert's gesture as addressee-directed. Conversely, that it is Lawrence's turn, that it is Lawrence who contributed the unwanted topic, and that the speaker-hearer line is broken all contribute to the understanding of Lawrence's gesture as self-directed.

4.4.3 Multiparty interactions

In an idealized face-to-face interaction with exactly two participants, the proposed Interaction Space is relatively straightforward. However, actual communication rarely occurs under such neat conditions. We are frequently attending to multiple social, physical, and cognitive tasks at once (Norris 2006). We shift our attention between friends at dinner, juggle attending to our pet, distracted child, and partner every morning, and never quite finish our phone call by the time we get to the counter or board the bus. Indeed, in *all* of the data presented

in this work, speakers must constantly attend to multiple tasks and interlocutors. In talk show interviews, each participant must not only attend immediately to their interlocutor, but also to the live audience and the assumed television audience. In monologues, Colbert shifts his attention between live and television audiences, and occasionally engages with his band on stage. In all of these cases, it is not enough to simply say that there is an Interaction Space between interlocutors which is maintained and managed throughout a discourse. We need systematic ways to account for distractions, digressions, multitasking, and, in short, multiparty interactions.

To address this issue I maintain that an Interaction Space is defined as the space between exactly two participants, but allow for the existence of multiple Interaction Spaces within a single interaction. Importantly, though multiple Interaction Spaces can be open, only one can be managed at a time. In differentiating between these spaces I borrow from Mental Spaces theory (Fauconnier 1994; Fauconnier & Turner 2008). Only one space is 'in focus' at a time. It is the 'in focus' space and its corresponding discourse that is being managed and contributed to.[16] The space in focus is determined by which speaker-hearer line is being maintained. This disambiguates spaces that may partially overlap, such as when three speakers are evenly positioned around a table.[17] We will see in the following examples that this approach has helpful repercussions, providing a principled way to disambiguate addressee reference and exclude digressions and interruptions that may otherwise render a discourse structure incoherent.

In the following two examples there are five open Interaction Spaces: (i) Colbert and his guest, (ii) the guest and the audience, (iii) the guest and the camera, (iv) Colbert and the audience, and (v) Colbert and the camera. Though this may at first seem like an unwieldy proliferation of spaces, the number of *primary* Interaction Spaces in which management consistently takes place is quite limited. For example, in these interviews, the space between Colbert and the guest is the only primary space. This is the space in which the main discourse is being collaboratively built and managed. It is only in this space that both participants have equal access to the space and its contents. The other listed spaces are secondary and are relatively limited in forms of management. In particular, the audience does not have access to Interaction Space management. Through behaviors like cheering and laughing, the audience *can*, in a sense, request a primary participant to attend to the secondary Interaction Space. However, the audience cannot, without extreme disturbance, signal disengagement, nor can they introduce, remove, or organize topics within the Interaction Space. They simply do not have the physical access to do so.

In this section, I will discuss how speakers use bodily reorientation to navigate between different Interaction Spaces. The first example demonstrates a relatively simple naviga-

[16]Drawing this analogy has the added benefit of aligning with a growing body of gesture literature that relates body orientation shifts in sign and gesture in narration to navigation between Mental Spaces (e.g. Dudis 2004; Parrill & Sweetser 2004; Stec 2012; Sweetser 2007).

[17]There are interesting repercussions of this on the "accessibility" of topics as metaphoric objects within the Interaction Space. It is possible, and even *likely*, that gesturer's take advantage of partial overlaps between spaces when engaging in a single discussion with more than one addressee.

tion in which the interviewee (Kathy Griffin) alternates between interacting with Colbert and interacting with the live audience. The second example is more complex in that the interviewee (Lin-Manuel Miranda) interacts not only with Colbert and the live audience, but also opens an additional space between him and his imagined Twitter followers. In both cases the systematic reorientation of the gesturer's body serves to disambiguate the intended addressee.

4.4.3.1 Co-present participants

In the first example, American comedian Kathy Griffin navigates between the primary Interaction Space between her and Colbert and a secondary Interaction Space between her and the audience. This alternation allows her to attend to and manage the primary discourse while also acknowledging audience reactions. In the clip, Griffin is telling Colbert about how she successfully promoted her own shows after a scandal. When she reveals her biggest success, selling out the large New York venue Carnegie Hall, the audience cheers. At this point she turns to thank the audience, clapping along with them (line 5, P2 & G1). Griffin then brings her hands to rest on her lap and turns back to Colbert as she laughs (P3). Colbert, after this moment of mutual gaze, takes the floor to make a comment about never selling out Carnegie Hall (line 6). In response, Griffin brags that she has sold it out five times (line 7). Griffin then reaches toward Colbert (G3) and leans her head dramatically away (P4), looking up and saying "anyway" as she does so. After this, Griffin returns to her story (lines 7-8). This return is signalled in the verbal mode with the discourse marker "so then", and in the gestural mode by returning her gaze to Colbert (P5) and performing a two handed presentational gesture along the speaker-hearer line (G4).

(42) Transcript 4: Kathy Griffin
[UID:c5b60a9c-fabb-11e8-ba14-089e01ba0335,3559]

```
1  KG   so um I had △everyone in Hollywood saying you can't sell any
                    △P1- - - - - - - - - - - - - - - - - - - - - - >
2       tickets, and I sold out Carnegie Hall in less than 24 hours.
        <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
3       And
        <- >
4  --   (applause)
5  KG   △*Thank you* *thank you* △ (laughs)
        △P2- - - - - - - - - - - △P3- - - >
          *G1- - - -* *G2 - - - *
6  SC   I never sold out Carnegie hall
7  KG   I've sold it out △five times.  △*Anyway um* △ *so so
        (off screen)     △P2- - - - - -△P3- - - - -△P4 - - ->
                                       *G3- - - -*   *G4 - ->
```

```
8        then I* decided to promote my own shows
         <- - - - - - - - - - - - - - - - - - - -≫
         <- - -*
P1, P3, P5:  head and gaze toward Colbert
P2:  head and gaze toward audience, head & gaze sweeps right
     across audience then settles in center
P4:  head turned to audience and leaning right toward shoulder,
     gaze up
G1:  clapping
G2:  both hands, flat open palm up, held in gesture space peripheries
G3:  left hand reach toward Colbert, palm flat on desk
G4:  both hands, open palm up, held toward Colbert
```

Figure 4.9: Gesture sequence corresponding to TRANSCRIPT 4: KATHY GRIFFIN, ex. 42.

Griffin's first disengagement from the primary Interaction Space (line 4, P2) is achieved primarily by turning her body, head, and gaze away from Colbert and toward the audience. This postural shift allows Griffin to attend to the audience, who is clapping and cheering loudly in support of her. This simultaneous disengagement and engagement is represented in (43). At time t Griffin disengages from the the primary Interaction Space between her and Colbert (IS1) by turning away (P2). At the same time t, Griffin engages with the Interaction Space between her and the audience (IS2) by turning toward them (P2) and performing manual gestures within IS2.

(43) **Management actions by Griffin, Line 4, ex. 42**

$$IS1_{\text{t}} = \begin{cases} \text{Participants: KG}_{\textbf{-S}}\text{, SC}_{\text{A}} \\ \text{Content : Selling tickets} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{P2}) \end{cases} \qquad IS2_{\text{t}} = \begin{cases} \text{Participants: KG}_{\textbf{+S}}\text{,} \\ \quad \text{Audience}_{\textbf{A}} \\ \text{Content : Selling tickets} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{P2, G1, G2}) \end{cases}$$

In addition to attending to the audience gesturally, Griffin also expresses gratitude verbally by saying "thank you" twice. Importantly, we know that this is addressed to the audience and not Colbert because of Griffin's orientation. Though this may seem obvious, the formalization of the Interaction Space helps to ground this intuition. Because the Interaction Space is explicitly defined as the space between two participants along a speaker-hearer line, there is no ambiguity as to which Interaction Space is in focus and when. The ability of Interaction Space navigation to disambiguate referents, as is the case here, is of clear interest to the discourse analyst.

It is worth contrasting Griffin's navigation between spaces with the disengagement gesture she performs after bragging to Colbert (line 7, P4). In this disengagement, Griffin turns her head away from Colbert, as before. Instead of engaging with the audience, however, Griffin tilts her head and gazes upward at empty space. She also keeps her hand on the desk, central to the Interaction Space between her and Colbert. Both of these gestural features keeps the Interaction Space between her and Colbert in focus while still allowing her to signal topic dismissal through disengagement and the discourse marker "anyway".

4.4.3.2 Imagined participants

In the second example, American actor and writer Lin-Manuel Miranda engages in an imaginary interaction with his Twitter followers after disengaging from the primary Interaction Space and briefly attending to the audience. Orientation toward empty space to engage with imagined interlocutors is a well-documented narrative strategy in both signed and spoken languages (e.g. Janzen 2004, 2012; Parrill & Sweetser 2004; Stec 2012; Sweetser 2007; Sweetser & Sizemore 2008). Thinking about this reorientation in terms of the Interaction Space helps us to systematically understand how the strategy works and why failing to implement it correctly can result in discourse derailment (Laparle 2022).

In the clip, Miranda is telling Colbert about a Twitter habit he has developed (tweeting "goodnight") and the misunderstandings that have resulted from it. He tells Colbert the basic misunderstanding, that people think he is actually going to bed (lines 5-7). As he does so, he performs an outward cyclic gesture (G1), likely metaphorically representing the habitual "thinking" of "everyone".[18] Miranda then denies this (line 7), and performs a full

[18]See Ladewig (2011, 2014a) for a discussion of the cyclic gesture and its metaphoric depiction of mental processes.

two-handed shrug, pursing his lips and tilting his head to signal the rejection.[19] Miranda then begins to address his followers directly and explain what is actually happening (lines 7-9). As he does this, he turns first toward the audience (P3), and then toward empty space above the two established Interaction Spaces (P4 & P5).[20] Finally, Miranda returns to the main discourse between him and Colbert (line 10), signalling the return verbally with the discourse marker "anyway" and gesturally by reorienting toward Colbert (P6) and performing a presentational gesture into the primary Interaction Space (G7).

(44) Transcript 5: Lin-Manuel Miranda
[UID:49c283d6-eef2-11e8-bf71-089e01ba0770,2113]

```
1   LM   I think that uh uh Twitter has given us issues with object
2        permanence.  I think the most damaging phrase we have is
3        "pics or it didn't happen".  It did happen.  Um.  Just
4        because a pic of it didn't happen doesn't mean it didn't
5        happen.  And so sometimes I'll say goodnight, and I'll go
6        and have dinner, and I'll have time with my family.  △*But
                                                            △P1->
                                                             *G1->
7        everyone thinks that I'm literally going to bed.*△*I'm not.*
         <- - - - - - - - - - - - - - - - - - - - - - - - - - △P2- - - >
         <- - - - - - - - - - - - - - - - - - - - - - - - - -* *G2 - - *
8        △*I'm just saying goodnight to you*△*to be polite*△*because
         △P3- - - - - - - - - - - - - - - - - △P4- - - - - -△P5- - >
           *G3 - - - - - - - - - - - - - - - -* *G4- - - - - -* *G5- ->
9        I like you* *and I like the time we've shared on Twitter*
         <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
         <- - - - -* *G6 - - - - - - - - - - - - - - - - - - - - - -*
10       △Um.  △*Anyway.  In the- in the-* y'know over the years,
         △P6- -△P7- - - - - - - - - - - - - - - - - - - - - - - - - ->
                  *G7 - - - - - - - - - - -*
10       people have sort of come to rely on it
P1, P6:  head and gaze toward Colbert
P2:  gaze toward Colbert, head tilted right, pursed lips
P3:  head and gaze toward audience, eyebrows furrowed
P4:  head and gaze toward upper right, eyebrows furrowed
P5:  head and gaze upward, eyebrows furrowed
```

[19] See Debras (2017) for a discussion of compositionality in shrugs and how specific articulatory features correlate with particular use contexts.

[20] P4 seems to be a transitional orientation between fully interacting with the audience (P3) and fully interacting with the imagined twitter followers (P5).

P7: head toward Colbert, gaze toward right hand
G1: hands cupped, small cyclic movements
G2: shrug, both hands open palm up, head tilted, lips pursed
G3-G5: both hands open, upturned, right hand toward interlocutor, moved up and outward, rearticulated three times
G6: both hands open, upturned, repeated lateral movements
G7: right hand open, upturned, held toward Colbert

Figure 4.10: Gesture sequence corresponding to TRANSCRIPT 5: LIN-MANUEL MIRANDA, ex. 44

The hand shape and orientation Miranda performs for his initial shrug toward Colbert (G2) is held through his participation in the secondary spaces (G3-G6). The head tilt and lip pursing characteristic of a shrug is not held beyond G2. I argue that the continuation of this hand shape and orientation takes advantage of the inherent ambiguity of the palm up open hand gesture (Cooperrider et al. 2018; Müller 2004) – the shape and orientation are consistent with both *offering* and *requesting* a metaphoric object. In the shrug, the manual gesture is associated with inaction, expressing that one has nothing to offer by displaying empty hands (Debras 2017). In the case of Miranda, the shrug seems to convey something like "what else can I say?", and aligns with the denial "I'm not" (line 7).[21] As Miranda reorients toward the audience (P3) and then his imagined interlocutors (P5), the palm up open hand gestures are more closely associated with obviousness (Marrese et al. 2021), expressing that the metaphoric objects displayed are within plain view. Within the discourse context, this can be interpreted as expressing a plea to accept what should be obvious explanations. This interpretation is supported by the accompanying speech; each iteration of the gesture

[21]Note that this instance is also the only one in the series of palm up open hand gestures in which Miranda 'reveals' his hands to be empty by turning them upward and moving them apart. This revealing process is specifically associated with the empty-hand interpretation of the palm up open hand gesture (Bavelas et al. 1992; Chu et al. 2014).

corresponds to a statement of what is actually happening ("saying goodnight", line 8, G3) and explanations for why it is happening ("to be polite", line 8, G4; "because I like you", line 8, G5; "I like the time we've shared on Twitter", line 8-9, G6).

This sequence is represented formally in (4.10) below. The top two IS states represent the initial disengagement from Colbert and engagement with the audience. The bottom two IS states represent the subsequent disengagement from the audience and engagement with the Twitter followers.

(45) **Management actions by Miranda, Lines 6-9, ex. 44**

$$IS1_{t} = \begin{cases} \text{Participants: LMM}_{\mathbf{-S}}, \text{SC}_{\text{A}} \\ \text{Content : True actions} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{P3}) \end{cases} \qquad IS2_{t} = \begin{cases} \text{Participants: LMM}_{\mathbf{+S}}, \\ \quad \text{Audience}_{\mathbf{A}} \\ \text{Content : +True actions} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{P3}) \\ \quad \textsc{Present}(\text{G3}\rightarrow\text{True actions}) \end{cases}$$

$$IS2_{t+1} = \begin{cases} \text{Participants: LMM}_{\mathbf{-S}}, \\ \quad \text{Audience}_{\text{A}} \\ \text{Content : True actions} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{P4}) \end{cases} \qquad IS3_{t+1} = \begin{cases} \text{Participants: LMM}_{\mathbf{+S}}, \\ \quad \text{Twitter followers}_{\text{A}} \\ \text{Content : True actions} \\ \quad +\text{Explanations (Expl}_{1\text{-}3}) \\ \text{Management :} \\ \quad \textsc{Engage}(\text{P4, P5}) \\ \quad \textsc{Present}(\text{G4}\rightarrow\text{Expl}_{1}) \\ \quad \textsc{Present}(\text{G5}\rightarrow\text{Expl}_{2}) \\ \quad \textsc{Present}(\text{G6}\rightarrow\text{Expl}_{3}) \end{cases}$$

These representations are particularly helpful in understanding what types of information Miranda wishes to add to each discourse, real or imagined. With his co-present interlocutors (Colbert and the live audience), he focuses on the "true" actions, what is actually happening when he says goodnight. To Colbert, he refers to his Twitter followers as "everyone" (line 7); Colbert is likely not one of his Twitter followers. He then addresses the audience directly with "you" (line 8) and reaches his right hand toward them (G3). It is likely that he has fans in the audience, some of whom may follow him on Twitter. His actual appeal to his Twitter followers, to accept the explanations he offers, is performed in the imagined third space. He uses a combination of "you" and "we" as he explains to them why he says goodnight, repeating a two handed presentational gesture for each of the three reasons.

It is worth noting the compositionality present in Miranda's presentational gestures toward the audience and the imagined space. In all cases, Miranda presents information as metaphoric objects on his open upturned hands. The trajectory of the presentation specifies the addressee. This is especially clear in G3 where his right hand reaches out along the speaker-hearer line between him and the audience. The repeated lateral movements in G6

convey representational meaning related to the repeated exchange of tweets. This is a particularly nice demonstration of the multifunctionality in gesture, and how a 'single' gesture can perform Interaction Space management while also conveying semantic information.

4.4.4 Interim summary

The engagement sequences discussed above highlight some of the functional distinctions between the personal gesture space and mutual Interaction Space. In all examples, the gesturer is gesturing in their personal space. We could use the positioning of the gestures relative to this space to interpret certain aspects of each gesture. For example, two handed gestures performed at the peripheries of the personal gesture space (e.g. Fig. 4.9, G2; Fig. 4.10, G4) may be interpreted as "large" and thereby "important", given the primary metaphor relating IMPORTANCE to SIZE. However, *interactive* meaning conveyed by the gesture must be interpreted relative to the Interaction Spaces involved. This is perhaps most obvious in the Lin-Manuel Miranda case; he is repeating nearly identical gestures in nearly identical regions of his personal gesture space. Despite this, we intuit that the individual gestures are contributing to interactions with different real and imagined participants. This intuition can only be captured by assessing each gesture relative to the Interaction Spaces involved.

4.5 Managing contents

Content management actions are grounded in the 'Conduit Metaphor', a pervasive metaphor in which language users conceive of ideas and topics as metaphoric objects (Reddy 1979). Unlike Reddy's original conception of the Conduit Metaphor which regards communication as metaphoric object *exchange*, the understanding here is that communication is conceptualized as metaphoric object manipulation.[22] This distinction is important because topics that are being actively pursued in the discourse are considered mutually accessible to both participants. A participant does not *give* a topic to their interlocutor. Instead, a participant adds a topic to the Interaction Space such that both they and their interlocutor can continue to interact with it.[23]

A topic can be introduced to the Interaction Space via a PRESENT action. Once established, a topic in the Interaction Space can then be REFERRED to, SEPARATED from other topics, or REMOVED. We have already seen examples of PRESENT and REFER actions in this chapter (ex. 38, 40, 42 & 44 and ex. 35 respectively). For this reason, I will focus on REMOVE and SEPARATE actions here. REMOVE actions are used to signal negative as-

[22] As called for in Sweetser 2007, this is evidence for the use of gesture in informing and updating our cognitive linguistic theories.

[23] Given the *independence of contribution* principle of multimodal discourse management, ongoing discourse topics are considered present in the interaction space regardless of whether they were explicitly introduced in the gestural mode or not.

sessment and the ending of a discourse topic. This is achieved through 'away' gestures (e.g. Bressem & Müller 2014, 2017) in which the gesturer moves the topic as a metaphoric object away from the speaker-hearer line and out of the Interaction Space. Once removed from the Interaction Space the topic is no longer available for interaction by either participant unless it is reintroduced, meaning that it cannot be the target of REFER or SEPARATE actions. SEPARATE actions are composite actions that PRESENT or REFER two topics in different regions of the Interaction Space and are used to express contrast (e.g. Calbris 2008; Hinnell 2019).

4.5.1 Remove

In this section, I provide two examples of REMOVAL actions, one in which the topic is represented metonymically by an actual physical object in the Interaction Space, and one in which the topic is represented as a virtual object in the Interaction Space. These two REMOVAL actions reinforce the understanding of the Interaction Space as fundamentally physical *and* metaphoric. Because it is physical, participants can communicate about and interact with literal objects within the space. Because it is metaphoric, participants can also interact with virtual objects. However, the actions performed on Interaction Space contents, whether they are physical or virtual, express the same forms of discourse management.

In the first example, Israeli-American actor Natalie Portman and Colbert are discussing a tweet Colbert has printed out on a physical card. For simplicity of presentation, I transcribe and discuss only the manual gestures. In the tweet, Mark Hamill, an American actor who stars as Luke Skywalker in the original Star Wars series, claims to have never met Natalie Portman, who plays Luke Skywalker's mother in recent Star Wars episodes. Colbert is amused by the claim that Hamill has never met 'his mother'. Colbert presents the tweet to the audience, placing it on his desk and pushing it toward the camera (CG1). He then brings the tweet back into the Interaction Space between himself and Portman as he summarizes his (faux) confusion (line 4). He performs a palm up open hand gesture, first toward the tweet, then Portman, and then the tweet once more (CG2). Portman plays along with this joke for a while, saying that she'd love to meet her 'son' (lines 8-10). She then interrupts the joke to ask if Hamill had wished her a happy birthday in the tweet (line 12), leaning into the Interaction space and reaching toward the tweet as she does so (PG1). Colbert looks at the tweet, responds "no" (line 13), and continues the joke, repeating his earlier three-part deictic gesture (compare CG2 and CG6). He finishes the joke by claiming that because he knows both Portman and Hamill, he must be the center of the Star Wars universe (lines 18-20). Finally, he closes the topic, verbally through the discourse marker "anyway" (line 22) and gesturally by removing the tweet from the Interaction Space and placing it under his desk (CG8). Colbert then moves to open a new topic about Portman's success as an actor (lines 25-26), performing a presentational gesture toward Portman as he does so (CG9).

(46) Transcript 6: Natalie Portman
[UID:a771d272-b00c-11e8-82c3-089e01ba0770,1635]

```
1   SC   Um uh Mark Hamill uh marked your birthday by sending out
2        this, *he said "fun fact, I've never met this woman"*
                *CG1- - - - - - - - - - - - - - - - - - - - -*
3   NP   (laughs)
4   SC   *But you're his mom*
         *CG2- - - - - - - -*
5   NP   I know, [I know]
6   SC             [You play-] you are his mom, you've never met,
7        you've never met him?  [How is this possible?]
8   NP                            [I know, it's such a  ] shame.
9        I would love to meet him.  Mark, I'd love to meet you.  Where
10       are you?  Come over
11 SC   *I have spent more time-*
         *CG3- - - - - - - - - -*
12 NP   *Did he say happy birthday?
         *PG1- - - - - - - - - - - ->
13 SC   *Uhhh, no.*
         *CG5- - -*
14 NP     * (laughs)
         <-*
15 SC   So maybe [you don't want to meet him]
16 NP            [That would be a good first step] (laughs)
17 SC   *I've spent more time with Mark Hamill than you have, and
         *CG6- - - - - - - - - - - - - - - - - - - - - - - - - - ->
18       I've spent more time with you than Mark Hamill has.* It's
         <- - - - - - - - - - - - - - - - - - - - - - - - - - -*
19       like I'm the center of the Star Wars universe (.)  *I might
                                                            *CG7- ->
20       be, [I might be]*
         <- - - - - - - *
21 NP    [Maybe you could connect] us
22 SC   Mhm, *anyway, congrat- happy birthday,
             *CG8- - - - - - - - - - - - - - - ->
23       [happy birthday.  Good to see you again]
24 NP   [Thank you, thank you]
25 SC   Um,* *now we- y'know, you're an Oscar winning actress*,
         <-* *CG9- - - - - - - - - - - - - - - - - - - - - - - *
```

```
26      everybody loves you as a performer
CG1:  presents card on desk, pushing toward audience
CG2, CG6:  left hand holds card; right hand open palm up toward
     Portman then toward card
CG3:  card flat on desk; right hand open palm up toward card
CG4:  holds card with both hands
CG5:  turns card over with left hand; right hand palm up toward card
CG7:  right hand fans self with card
CG8:  right hand removes card to underneath of desk; left hand loose
     palm up
CG6:  left hand open palm up toward left, then toward Portman
PG1:  left hand on desk reaches toward card, open palm right
```

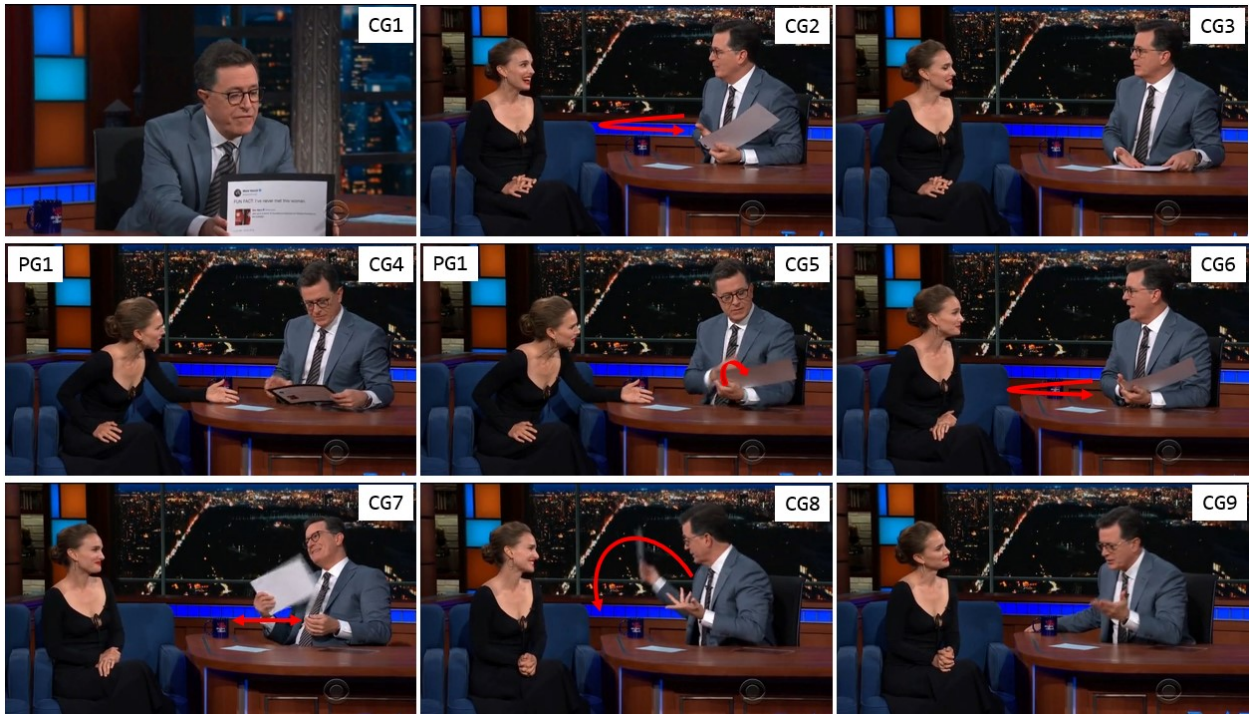


Figure 4.11: Gesture sequence corresponding to Transcript 6: Natalie Portman, ex. 46

Colbert first interacts with the audience to provide the information necessary to understand the jokes he is about to make. This interaction with a secondary space is marked by his orientation and his gesture toward the camera. For the rest of the clip, he contributes to the primary Interaction Space between him and Portman, as they both interact with the printed tweet. Colbert makes two requests for a response from Portman (CG2 & CG6), deictically referring to the topic under discussion, then Portman to elicit a response, and then the topic

under discussion again to specify a focus for her response. We saw a similar request sequence in Section 4.4.1 with Elizabeth Warren.

The removal sequence here is very straightforward. Colbert introduced the topic and has physical control of the card, which metonymically represents the discourse topic as a whole. While over the desk, the card can be located by both participants in order to add or request information about the topic as an object in the Interaction Space. When the card is eventually removed by Colbert (CG8), neither participant is able to locate or interact with the topic. This is formally represented in (47) as a REMOVE action expressed by CG8 and acting upon the tweet (REMOVE(CG8→Tweet)), which is removed from the space's contents (~~Tweet~~). Colbert's next discourse move is to present a new topic for discussion, this is represented as a PRESENT action (PRESENT(CG9→Success)) acting upon the newly introduced topic (+Success). Note that in this state, the tweet is no longer listed in the space's contents.

(47) **Management actions by Colbert, ex. 46, lines 22 & 25**

$$IS1_1 = \begin{cases} \text{Participants: } \text{SC}_{\mathbf{S}}, \text{NP}_{\text{A}} \\ \text{Content : } \sout{\text{Tweet}} \\ \text{Management :} \\ \quad \text{REMOVE(CG8}\rightarrow\text{Tweet)} \end{cases} \qquad IS1_2 = \begin{cases} \text{Participants: } \text{SC}_{\mathbf{S}}, \text{NP}_{\text{A}} \\ \text{Content : +Success} \\ \text{Management :} \\ \quad \text{PRESENT(CG9}\rightarrow\text{Success)} \end{cases}$$

Once the tweet is removed (IS1, CG8), the Interaction Space has no contents. Because there are no longer any topics within the Interaction Space to perform management actions upon, there are only two gestural moves that Colbert can make: he can (i) DISENGAGE and cede his turn as speaker, or (ii) PRESENT a new topic for discussion. As shown above, he chooses the latter. This demonstrates the predictive capacities of the Interaction Space. If Colbert performed some action other than the predicted DISENGAGE and PRESENT actions, the example would be worth closer analysis in order to reassess the interpretation of Colbert's gestures and the assumed discourse structure, and then update our analysis or our model accordingly.

The second example is from an interview with Irish actor Andrew Scott and demonstrates the removal of topics as virtual objects from the Interaction Space. I will consider both manual and non-manual gestures in this example, showing how manual removal actions can be combined with non-manual disengagement to accomplish a complex discourse move.

In this clip, Scott is excitedly talking to Colbert about how to make Shakespeare relevant and engaging for modern audiences. He premises the topic with a hedge, claiming that he doesn't want to "bore" Colbert (line 1-2). This hedge can be interpreted as a reassurance that the topic can be ended if either interlocutor no longer wants to maintain it as a discourse goal. Scott then performs a series of presentational gestures as he argues for the continued relevance of Shakespeare. The last of these presentational gestures is shown as SG1, corresponding to line 9. Scott then retracts his hands to a resting positing and looks down as his sentence trails off (lines 9-10, SP2). Having presumably reached the end of his desired contributions to the topic, Scott then moves to close the topic, using the discourse marker "anyway" (line

10) and a removal gesture, sweeping his hand in a large arc out from the central Interaction space out of the camera frame (SG2). As he closes the topic, Scott also performs a full non-manual disengagement from the primary Interaction Space, turning his head and gaze away, and leaning away from Colbert (SP3). Colbert then, predictably, takes the turn, offering a kind of consolation to Scott, saying that the topic was interesting (line 12), and holding an upturned open hand toward Scott as he does so (CG1). During Colbert's turn, Scott does a full body readjustment, standing up and settling back down in the chair.[24] After resettling, Scott reengages by reorienting his head and gaze toward Colbert and leaning sharply into the Interaction Space (SP6). Scott and Colbert then enter into a closing sequence[25] to end the interview (lines 16-18).

(48) Transcript 7: Andrew Scott
[UID:f00485b2-051e-11ea-8f95-089e01ba0770,3559]

```
1  AS   I'm a bit of a Shakespeare nerd, so- and I don't want to bore
2       you.  But wh-what I find fascinating about it is that,
3       y'know, in order to for Shakespeare to stay relevant, I don't
4       think y- you need to cut it down.  I think you might need to
5       make it as uh- exciting and as thrilling for an audience uh-
6       of today as it would be four-hundred years ago because,
7       y'know, we binge watch TV. Y'know, we watch five hours of
8       television if it's exciting.  △So the idea is don't cut it
                                      △SP1- - - - - - - - - - - ->
9       down, *just make it four hours of really exciting* △um play-
        <- - - - - - - - - - - - - - - - - - - - - - - - - - △SP2- ->
              *SG1 - - - - - - - - - - - - - - - - - - - - *
10      plays.  △*Anyway, I'll stop talking* △Shakespeare△(laughs)
        <- - - -△SP3 - - - - - - - - - - - -△SP4- - - - △SP5- ->
                  *SG2- - - - - - - - - - - *
11 SC  *No not at all*, not at all.  No, that's amazing.
       *CG1- - - - -*
12 AS  △ yeah
       △SP6->
13 SC   That's amazing.  I could talk about Hamlet all night.  They
14      won't- they'll turn the cameras off, but-
15 AS   (laughs)
16 SC   Well, Andrew thank you so
```

[24]This can be related to the primary metaphor Mental states are Physical states such that Scott is physically readjusting in order to mentally prepare for a new topic and discourse goal. Thank you to friend and colleague Kelly Jones for pointing out this interpretation to me.

[25]See Schegloff & Sacks (1973) for a discussion on topic-shifting and turn-negotiation at the end of interactions.

```
17       [much for being here, it was lovely to meet you]
18 AS    [thank you so much, thank you, thank you so much]
SG1:  both hands, open palm up, right hand held higher
SG2:  right hand, open palm out, arched sweep to right
SP1, SP6:  head and gaze toward Colbert, leaning in
SP2:  head and gaze down, leaning in
SP3:  head turned right, eyes closed, straightens posture
SP4:  head and gaze toward audience, straight posture
SP5:  full body readjustment, stands and sits back down
CG1:  both hands open palm up, right hand reaches toward Scott
```

Figure 4.12: Gesture sequence corresponding to TRANSCRIPT 7: ANDREW SCOTT, ex. 48

As we've seen in previous examples, Scott introduces a topic and sub-topics therein through presentational, palm up open hand gestures (SG1). Just like with the physical object discussed in (46), the virtual object introduced into the Interaction Space by Scott can be subsequently referred to, manipulated, and removed by either participant. Because Scott introduces and dismisses the topic within a single conversational turn, he is the only interlocutor actually performing management actions on this particular topic. The final action he performs on the virtual object is, of course, the removal gesture (SG2). This management action is represented in $IS1_1$, ex. 49. Colbert's subsequent presentation of appreciation for Scott's contributions is represented in $IS1_2$.

(49) **Management actions by Scott & Colbert, ex. 48, lines 9-11**

$$IS1_1 = \begin{cases} \text{Participants: AS}_{\textbf{-S}}\text{, SC}_{\text{A}} \\ \text{Content : } \sout{\text{Shakespeare}} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{SP3}) \\ \quad \textsc{Remove}(\text{SG2}\rightarrow \\ \qquad \text{Shakespeare}) \end{cases} \qquad IS1_2 = \begin{cases} \text{Participants: SC}_{\textbf{+S}}\text{, AS}_{\text{A}} \\ \text{Content : +Appreciation} \\ \text{Management :} \\ \quad \textsc{Present}(\text{CG1}\rightarrow \\ \qquad \text{Appreciation}) \end{cases}$$

Scott's Disengage(SP3) action signals his intention to cede his turn, represented in the participants list as $\text{AS}_{\textbf{-S}}$. With the floor open, Colbert is then expected to take the floor ($\text{SC}_{\textbf{+S}}$). Because Scott has just cleared the Interaction Space of is contents, the only content management action Colbert can perform is to Present a new topic, which he does (CG1). Colbert does not perform a distinct Engage action when taking the turn, but this is not particularly surprising. The floor is open, so there is little reason to take actions signalling one's desire to take the floor. Natalie Portman's one gesture sequence in the previous example provides a nice contrast to this (fig. 4.11, PG1). In this sequence, Colbert has the floor and is pursuing a particular topic. Portman wishes to ask an intrusive clarifying question. To do so, she must signal that she wants to take the floor and specify what additional information she wants. She expresses both of these in the gestural mode. Her intention to take the floor is marked by a very noticeable Engage action in which she leans sharply into the Interaction Space. Her request for particular information is marked by the reach directly along the speaker-hearer line toward the location of the main discourse topic (the tweet in Colbert's hand). Neither of these efforts are necessary for Colbert – neither Scott's turn nor an ongoing topic need to be interrupted.

4.5.2 Separate

The final example I will discuss is from an interview with American journalist John Dickerson. Dickerson performs a composite Separate action consisting of three consecutive Present actions in different regions of the Interaction Space. Several gestures later, Dickerson then performs two Refer actions in the form of flat hand points toward the two separated topics. This example shows the ways in which multiple topics can be considered inside the Interaction Space and how they can be meaningfully organized therein. The use of spatial separation, especially along the left-right access, to distinguish concepts from one another is well-documented in the gesture literature (e.g. Calbris 2008; Hinnell 2019; Jannedy & Mendoza-Denton 2005; McNeill 2003). With the exception of McNeill (2003), these observations have not been discussed in the terms of constructing a mutual space.

In this clip, John Dickerson is discussing the ongoing problematic conduct of then-president Donald Trump. Of particular concern is whether or not Trump's conduct warrants impeachment. Dickerson begins this sequence by restating the main topic that has been ongoing for some time (lines 1-2, "the actual conduct of the president"). As he does so,

Dickerson lowers his open, inward facing hands along the speaker-hearer line, as if to place a narrow object on the desk between him and Colbert (DG1). He then contrasts the two possible assessments of Trump's conduct, moving the metaphoric object to the left to signal "wrong" (DG2), and then moving the metaphoric object to the right to signal "impeachable" (DG3). After three very brief gestures, (DG4-DG6), Dickerson repeats the left to right movement. As before, the assessment "wrong" (DG7) is to the left, and "impeachable" to the right (DG8). This second left-right sequence is performed as a one-hand open palm pointing gesture rather than a two-handed containment gesture like in the first iteration. Dickerson then summarizes his point, that people who hold this position don't understand what "they're signing onto" (lines 5). At the start of this summarizing statement, Dickerson performs a loose two-handed containment gesture, as if to gather the objects he has positioned in different places into one "position" (line 4, DG9).

(50) TRANSCRIPT 8: JOHN DICKERSON
[UID:465f22fa-01fa-11ea-8c13-089e01ba0335,2625]

```
1  JD  So you're signing up for a lot if you say *the actual conduct
                                                  *DG1 - - - - - - >
2      of the president* *was wrong* *but not impeachable* *and I
       <- - - - - - - -* *DG2- - -* *DG3- - - - - - - - * *DG4->
3      don't know if* *those* *who hold* *the wrong* *but not
       <- - - - - -* *DG5- * *DG6- - -* *DG7- - - * *DG8- ->
4      impeachable* *position* understand the kind of basket of
       <- - - - -* *DG9- - -*
5      goods they're signing onto, if that's what they decide to do
DG1:  both hands, close together, open palms center, down to desk
     3x beat
DG2:  both hands, close together, open palms center, arch left down
DG3:  both hands, close together, open palms center, arch right down
DG4:  right hand, fingers bunched
DG5:  right hand, index finger point
DG6:  right hand, loose open palm down, sweep right
DG7:  right hand, open palm left, arch left and down
DG8:  right hand, open palm left, arch right and down
DG9:  both hands, loose cupped, palms center, held in front of chest
```

The first left-right gesture sequence (DG1-DG3) expresses a series of three consecutive PRESENT actions. The containment configuration Dickerson maintains throughout suggests the placement and subsequent manipulation of a virtual object in the Interaction Space. The first of these three containment gestures presents the topic itself (Trump's conduct) as a metaphoric object in the central Interaction Space. This is depicted in ex. 51, $IS1_1$ as the PRESENT(DG1→Conduct) management action and the addition of the topic to the space's content (+Trump's conduct). The second and third containment gestures (DG2 &

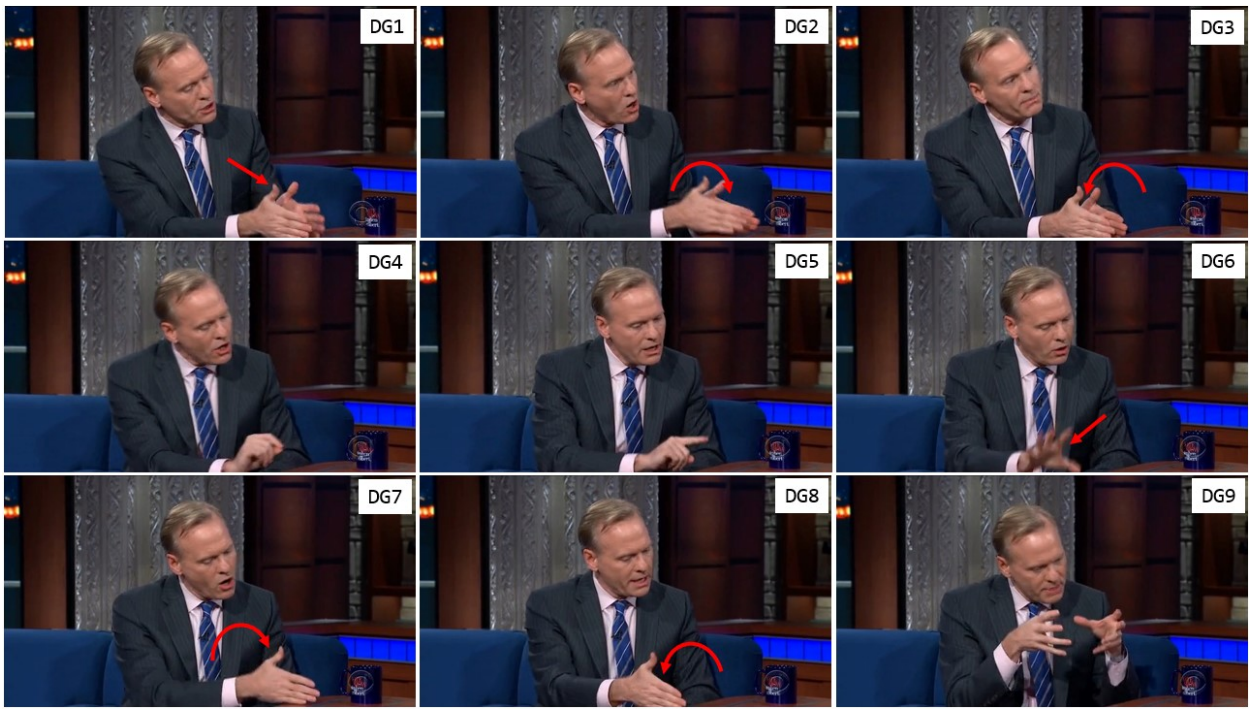


Figure 4.13: Gesture sequence corresponding to Transcript 8: John Dickerson, ex. 50

DG3) "move" the topic between possible assessments ("wrong" and "impeachable"). These two movements are still considered presentational, even though the topic has already been introduced to the Interaction Space – the topic is introduced into each *assessment* space. By moving the topic into a particular region, Dickerson makes that region of the Interaction Space itself meaningful, at least as long as the topic is maintained. The SEPARATE action is a composite action which reinforces the difference in meaning and function of the two spatially distinct regions. This can be related to the primary metaphor SIMILARITY IS PROXIMITY in which the difference between two concepts is understood as the degree of physical separation.

These three movements can also be framed in the terms of Mental Spaces Theory (Fauconnier 1994). In a Mental Spaces Theory approach, Dickerson's first containment gesture (DG1) can be interpreted as introducing a referent (Trump's conduct) into the 'base space', which contains facts about the actual state of the world. He presents this base space as directly in front of him along the speaker-hearer line. Possible assessments of Trump's conduct can then be conceptualized as distinct possibility spaces, in which the conduct is assessed as 'wrong' but not deserving of impeachment, or wrong enough to deserve impeachment.[26]

[26]See Dancygier & Sweetser (2005) for an extensive discussion of hypothetical spaces in Mental Space Theory.

(51) **Management actions by Dickerson, Lines 2 & 3, ex. 50**

$$IS1_1 = \left\{ \begin{array}{l} \text{Participants: JD}_{\mathbf{S}}\text{, SC}_{\text{A}} \\ \text{Content : +Trump's conduct,} \\ \quad \text{+Assessment of conduct} \\ \text{Management :} \\ \quad \textsc{Present}(\text{DG1}\rightarrow\text{Conduct}) \\ \quad \textsc{Present}(\text{DG2}\rightarrow\text{Assess}_1) \\ \quad \textsc{Present}(\text{DG3}\rightarrow\text{Assess}_2) \\ \quad \textsc{Separate}(\text{DG2+DG3}\rightarrow \\ \qquad \text{Assess}_1\text{+Assess}_2) \end{array} \right. \quad IS1_2 = \left\{ \begin{array}{l} \text{Participants: JD}_{\mathbf{S}}\text{, SC}_{\text{A}} \\ \text{Content : Trump's conduct,} \\ \quad \text{Assessment of conduct} \\ \text{Management :} \\ \quad \textsc{Refer}(\text{DG7}\rightarrow\text{Assess}_1) \\ \quad \textsc{Refer}(\text{DG8}\rightarrow\text{Assess}_2) \end{array} \right.$$

As observed in cases of abstract deixis (McNeill 2003; McNeill et al. 1993), the two meaningful regions of space can be subsequently referred to, as is done in DG7 & DG8. This second left-right gesture sequence expresses two consecutive Refer actions, as represented in $IS1_2$. These are considered Refer actions, rather than repeated Present actions, because the hand shape and orientation do not have the affordances necessary for presentation – you cannot present an object on a single vertically oriented hand. Instead, these two gestures are primarily deictic, inciting joint attention toward the two meaningful regions of the Interaction Space.

The two representations above highlight an important distinction between Present and Refer actions – Refer actions are restricted to already-established topics, whereas Present actions are not. This is somewhat complicated by the fact that, given the *independence of contribution* principle of multimodal discourse management, topics do not have to be introduced in the gestural mode. This means that we do not expect Refer actions to *always* be preceded by Present actions. However, where both occur, we expect Present actions to precede Refer actions. This is analogous to our expectations for referential expressions in the verbal mode. We expect the first mention of a new topic to be a full referential expression, and nearby subsequent mentions to occur as definite and reduced forms (e.g. Grosz & Sidner 1986; Prince 1981).[27]

4.5.3 Interim summary

In this section, I have discussed some ways in which gesturers can organize topics in the Interaction Space in order to convey discourse structural meaning.

The first two examples (ex. 46 & 48, Section 4.5.1) demonstrated how speakers can signal the closure of a topic by removing physical and virtual objects from the Interaction Space. In both cases, it was necessary to assess the Remove action relative to the Interaction Space rather than the personal gesture space. In the example with Natalie Portman, both

[27]Given that there are exceptions to the above generalizations in the verbal mode, we may expect analogous exceptions in the gestural mode. For example, Present actions may be used for emphasis of a recently mentioned referent, or be used after a Refer action to mark some shift in discourse unit.

Portman and Colbert interacted with a printed tweet; Colbert by actually physically holding and moving it, and Portman by referring to it via a deictic gesture. When Colbert removed the tweet by placing it under his desk, the physical affordances of the Interaction Space changed for both him and Portman – *neither* participant could continue to interact with an object they had previously interacted with. This mutual change in behavior cannot be fully accounted for without reference to a mutual space. Similarly, the REMOVE action produced by Andrew Scott prevented Colbert from interacting with a topic he didn't have a chance to interact with at all, given that Scott introduced and dismissed the topic within a single turn. The only action immediately available to Colbert after Scott's dismissal was PRESENT, which is exactly the action he performs. As before, this restriction on Colbert's behavior, based on gestures performed not by himself but by Scott, can only be fully accounted for by considering a mutual space.

We also saw examples of PRESENT actions in which gesturer's introduce a topic as a metaphoric object to the Interaction Space, thus making it accessible to both participants for further interaction. We saw several examples where a topic introduced to the Interaction Space by one participant is subsequently referred to by the other participant via a locating gesture (Dickerson fig. 4.4; Warren fig. 4.5; Portman fig. 4.11). That a PRESENT action by one participant enables another participant to perform particular gestural actions cannot be fully accounted for without reference to a mutual space.

Finally, we saw that the Interaction Space can be meaningfully organized into sub-regions based on (i) proximity to each participant, and (ii) spatial separation of topics. In Figure 4.5, we saw that both Colbert and Elizabeth Warren gestured toward a space relatively close to Warren when talking about her actions and attributes. I argued that this was based on the primary metaphor ATTRIBUTES/EXPERIENCES ARE PHYSICAL POSSESSIONS, which, based on our embodied experiences, are kept relatively close to the possessor's body. In the example with John Dickerson (fig. 4.13), we saw that topics can also be organized relative to each other. In the particular example Dickerson takes advantage of this to contrast two potential assessments of Donald Trump's actions by establishing them in separate areas within the Interaction Space.

4.6 Discussion

The process of making predictions, testing those predictions, and then changing our theories based on how the predictions succeed or fail, serves to increase the empirical power and validity of our theories. Because of the apparent polysemy and idiosyncrasy of interactive gestures, attempts at building predictive theories of interactive meaning and multimodal models of discourse structure are more or less non-existent. In this chapter, I have outlined the ways in which a predictive model of interactive gesture can be built based on the observable physical affordances of object-oriented tasks. Especially in cooperative tasks, such as a board game, objects are intentionally organized in mutual space based on how accessible they must be to each participant. We can carry these principles of spatial organization of literal

objects to the widely-accepted metaphorical representation of communication as manipulating objects. The model I propose is based on the Interaction Space, the physical mutual space between interlocutors used to organize discourse through literal and metaphoric object manipulation.

Given the Interaction Space and the manipulation of discourse topics as objects therein, we can make the following set of predictions:

1. Through changes in bodily orientation and posture, gesturers ENGAGE with the Interaction Space when actively contributing to the discourse structure, and DISENGAGE when not contributing to the discourse structure.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, physical engagement with the space is indicative of metaphoric participation in the discourse.

2. PRESENT actions can be used to introduce topics as metaphoric objects into the Interaction Space. Introduced topics are accessible to both participants for further interaction.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, the physical proximity to metaphoric objects determines their metaphoric accessibility as topics to be discussed.

3. REMOVAL actions remove a topic as a metaphoric object from the Interaction Space. Removed topics cannot be further interacted with by either participant.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, the removal of metaphoric objects from the shared physical space limits their metaphoric accessibility as topic to be discussed.

4. Only ongoing discourse topics can be referred to via REFER actions into the Interaction Space.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, the topics that are metaphorically accessible for discussion should be physically accessible within the shared space.

5. The placement of topics in particular regions of the Interaction Space can convey meaning based on their location relative to each participant and other co-present topics.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, locating metaphoric objects in different regions of physical space results in inferences

about the properties of the metaphoric object, especially regarding their association with individual participants and their differentiation from other concepts.

6. Temporary topics not up for further discussion (e.g. asides, parentheticals) should either be (i) established outside of the Interaction Space where they are not accessible to the other participant; or (ii) introduced into the Interaction Space, but removed before further interaction can occur.

   Motivation: Because the Interaction Space is both the physical space in which an interaction takes place and the metaphoric space in which a discourse is constructed, limiting the physical accessibility of metaphoric objects impacts the metaphoric accessibility of topics under discussion.

In the following three chapters I will explore the extent to which these predictions hold for particular kinds of discourse management expressed in the verbal mode. In particular, I consider three lexical discourse markers – *here's the thing, by the way*, and *anyway* – and the interactive gestures accompanying them. Because each lexical discourse marker serves a distinct discourse management function, we can make predictions as to what types of gestures should occur with each.

Here's the thing is used to mark a statement as a particularly important or surprising contribution to an ongoing discourse topic. As such, we expect to see participants increase their engagement with the Interaction Space in order to make a substantial contribution to the discourse structure. We also expect to see presentational and locating gestures as participants (i) PRESENT new points and arguments about existing topics, and (ii) REFER to particular aspects of existing topics.

By the way is used to mark a statement as digressive from the ongoing discourse topic. Assuming that digressive topics are not intended to be taken up by the other participant, we expect participants to PRESENT and REFER to digressive topics on the periphery of the Interaction Space, preventing further manipulation.

Anyway is used to signal a topic-shift, simultaneously dismissing a topic and introducing a topic. As such, we expect to see both REMOVE and PRESENT actions accompanying *anyway*. Furthermore, we expect REMOVE actions to occur first in order to clear the Interaction Space contents and enable PRESENT actions.

As we will see, these predictions are largely supported by the data, but not without exception. By considering a large amount of data, approximately 100 instances of each lexical discourse marker, I identify generalizations as to how lexical discourse markers and gestural discourse markers work together systematically and predictably to express complex features of a discourse's structure. Through these generalizations, I demonstrate that a predictive model of interactive gesture meaning is not only possible, but relatively simple. This, in turn, suggests that an integrated model of multimodal discourse management is an attainable and worthwhile goal. I also argue that careful qualitative analysis of exceptional data directly benefits our theories of gesture meaning and discourse structure by drawing us to new patterns, and inviting us to ask questions that we may not have asked otherwise.

Chapter 5

What is multimodal topic-shifting *anyway*?

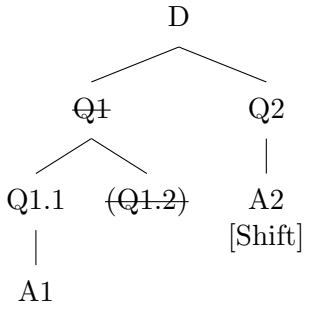
5.1 Introduction

This chapter looks at the gestural expression of topic-shifting as performed with the lexical topic-shift marker "anyway". I consider the distribution of the five classes of action schematic gestures introduced in Chapter 2 (PRESENTATION, REFERRING, REMOVAL, STOPPING and ENGAGEMENT/DISENGAGEMENT) across use-contexts of "anyway". I argue that *removal, stopping* and *disengagement* gestures express the dismissal of the ended topic. This is in contrast with *presentation, referring* and *engagement* gestures which express the introduction of a newly initiated topic or the return to a previous topic. Gestural expressions of topic dismissal and introduction convey meaning relative to the Interaction Space(s) between interlocutors, as defined in Chapter 4. Topic-dismissal involves some movement away from the central Interaction Space and the speaker-hearer line. Topic introduction involves some movement toward the central Interaction Space and re-establishment or continued maintenance of the speaker-hearer line. I also discuss two additional recurrent behaviors, *adjusting* and *shrugs*, that appear to relate to the transition period between dismissed and (re-)introduced topics.

As proposed in Chapter 3, topic-shifting is a type of discourse move that results in the pre-mature closure of an ongoing discourse topic in favor of a higher-level discourse topic. In a question-based approach to discourse structure, this amounts to the cancelling of a potential sub-question and closure of an open question in favor of a new question. This is represented schematically in (52). Here we see a potential sub-question ((Q1.2)) cancelled along with its immediate super-ordinate question (Q1). The topic-shift is completed once Q2 is established. Importantly, the introduced question attaches higher in the discourse structure than the cancelled sub-question.[1]

[1]This movement from a more deeply embedded question to a higher question was referred to as "popping" in early work on hierarchical discourse structure (Polanyi 1988; Scha & Polanyi 1988).

(52)

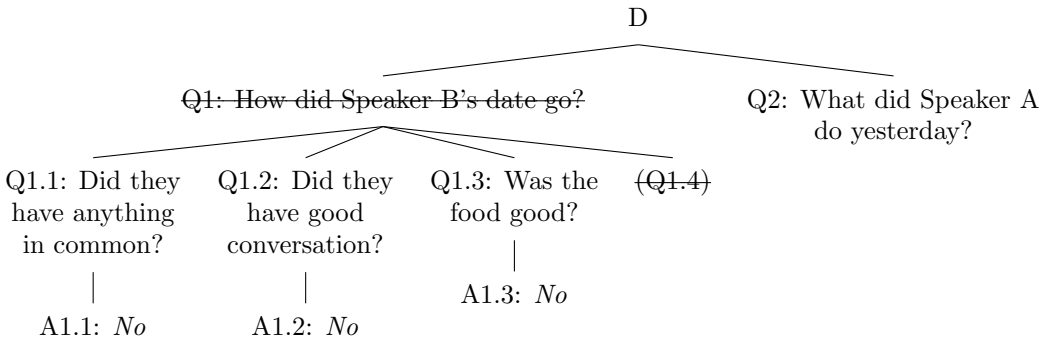


Topic-shifting, under this analysis, is considered a disruptive discourse move that threatens discourse coherence – the expectation is that open questions are addressed until fully answered or mutually deemed unanswerable. Closing a question before either of these have occurred violates the expectations of a cooperative, goal-oriented discourse. As an illustration, consider the fictitious example in (53).

(53) SPEAKER A: So how did the date go? I want to hear all about it.
SPEAKER B: Terribly. We didn't have anything in common, couldn't keep a conversation going, plus the food was bad. (*Anyway*) what did you do yesterday?

Speaker A overtly sets the discourse goal by asking a question about Speaker B's date and further specifies that they want to hear "all about it". After offering three cursory answers, Speaker B shifts the discourse topic to Speaker A by posing a new question. By setting the next discourse topic, Speaker A signals a refusal to answer any more potential questions regarding their date. This is schematically represented in (54).

(54)



The lexical discourse marker "anyway" is grammatically optional – Speaker B could simply ask the new question. However, the inclusion of "anyway" helps to maintain discourse coherence by overtly signalling that a relatively uncooperative discourse move is about to be made. It also serves to reinforce that Speaker B *knows* that Speaker A wants to hear more about the date and that they are refusing to do so. In other words, "anyway" serves to highlight a misalignment in perceived discourse goals (Park 2010).

As alluded to in Chapter 3, not all topic-shifts threaten discourse coherence to the same degree. The degree of disruption caused by a topic-shift depends on the discourse status of

both the dismissed and introduced topics. For example, the dismissed topic may be either (i) a legitimate line of inquiry within the overall discourse structure that helps to achieve discourse goals, or (ii) a digressive line of inquiry that does not serve ongoing discourse goals. In the example above, the dismissed topic is a legitimate line of inquiry, explicitly set by Speaker A. The dismissal of a digressive topic would be less disruptive than the dismissal of a question relating directly to discourse goals. Likewise, the introduced topic can either (i) set an entirely new goal, or (ii) relate to a more general, still-open question under discussion. In the example above, Speaker B introduces an entirely new goal by posing a question unrelated to the one previously posed by Speaker A. The introduction of a topic that is related to previously established discourse goals would be less disruptive than the introduction of an entirely new topic. This structural and functional variation is discussed in detail in Section 5.3.2.

"Anyway", in its use as a discourse marker, can be considered both *multifunctional* and *underspecified*. It is multifunctional in that it simultaneously signals the dismissal of one discourse topic and the (re)introduction of another (Bublitz 1988; Takahara 1998).[2] In this chapter, I show how gestures accompanying "anyway" express only one of these two functions at a time, namely *dismissal* or *introduction*. Furthermore, where *both* functions are expressed in the gestural mode, it is always in the order *dismissal* then *introduction*.[3] "Anyway" is underspecified in that it does not provide information about the relationships of the dismissed and introduced topics to the overall discourse structure (Fraser 2009).[4] In this chapter, I discuss evidence for the ability of accompanying interactive gestures to overtly express the discourse-structural properties that are underspecified by "anyway".

To understand the contributions of both the lexical and gestural discourse markers in maintaining discourse coherence through a topic-shift I carefully consider variation in both the verbal (Section 5.3) and gestural (5.4) modes. As in previous chapters, I use the four principles of multi-modal discourse management proposed in Chapter 1 to guide this discussion. The proposed principles are reiterated for topic-shifting in particular in (55).

(55) **Principles of Multimodal Topic-Shifting**

a. *Multiple strategies*: There are multiple strategies, within both the verbal and gestural modes, for expressing topic-shifting.

b. *Optionality of expression*: The use of each individual expressive strategy is optional and subject to contextual variation.

c. *Independence of contribution*: Strategies may be employed independently and may profile a different aspects of the topic-shift.

[2]Fraser (1988) refers to this compound action as topic 'reorientation', evoking a spatial metaphor (Discourse Space is Physical Space) that is recurrently enacted through bodily reorientation in this data set.

[3]This appears to be an example of gesture as *more* segmented than the accompanying speech, contra claims that gesture is always less segmented (e.g. McNeill 2000, 2005). Also see Chapter 1 for discussion.

[4]Though some analyses have argued that it is specifically a resumptive marker that returns the discourse from a digression to a higher discourse goal. See discussion in Ferrara (1997).

d. *Compositional management*: The strategies employed in both modes are integrated systematically and predictably into a single coherent multimodal message.

Based on both existing literature, and previous discussion in this work, I make the following predictions for the gestural strategies used to express topic-shifts. No predictions are made for *addressing* gestures which are expected whenever action is requested of the addressee.

(56) **Gestural expression of topic-shifting**

a. *Removal gestures*: If the discourse topic immediately preceding "anyway" is dismissed, we expect removal of the topic as a metaphoric object from the Interaction Space. Doing so prevents further interaction with the topic by making it is no longer accessible as a metaphoric object within the Interaction Space. This prediction is supported by previous work on the 'away' gesture family which demonstrates a relationship between movement away from the central gesture space and the dismissal of information (e.g. Bressem & Müller 2014, 2017; Gawne 2021; Harrison 2010; Teßendorf 2014).

b. *Stopping gestures*: If "anyway" marks the discontinuation of a discourse goal, we expect stopping gestures to be used to signal the stopping metaphoric forward motion of discourse participants toward established discourse goals. This prediction is supported by work on "blocking" gestures which demonstrates the relationship between metaphorically holding an object away and a request to stop a line of action (e.g. Wehling 2017).

c. *Presentation gestures*: If a discourse topic is initiated immediately after "anyway", we expect presentation of the new topic as a metaphoric object into the Interaction Space. Doing so enables subsequent manipulation of the topic as a metaphoric object therein. This prediction is supported by previous work on the palm-up open-hand presentational gesture which demonstrates a relationship between the presentation of a metaphoric object on an open hand and the contribution of new information to the discourse (e.g. Bavelas et al. 1992; Cooperrider et al. 2018; Müller 2004).

d. *Locating gestures*: If "anyway" signals topic resumption, i.e. the return to a higher-level discourse topic, we expect locating gestures into the Interaction Space. Doing so draws joint attention to the topic as a mutually accessible metaphoric object that is already in the Interaction Space. This prediction is supported by previous work on abstract deixis which demonstrates the organization of discourse topics as metaphoric objects in shared space (e.g. McNeill & Levy 1982; McNeill et al. 1993).

e. *Alternating engagement with the Interaction Space*: If the utterance marked by "anyway" initiates a topic shift, we expect to see *disengagement* from the Interaction Space at the point of topic dismissal, and *re-engagement* at the point of topic introduction. This is because disengagement signals a lack of participation with the current Interaction Space state, which metaphorically contains the un-

wanted topic; engagement signals participation with the subsequent Interaction Space state and newly presented metaphoric objects therein.

These predictions are largely supported by the gestural behavior observed in the present data set. However, there is also significant variation in gesture performance, including in the articulators involved and in particular formal features, such as hand shape. Potential functional motivations for this formal variation are discussed throughout.

This chapter is structured as follows. Section 5.2 describes the data set and annotation methodologies used. Section 5.3 reviews variation in the form and function of "anyway"-marked topic-shifts as observed in the present data set. In Section 5.4, I describe recurrent gesture patterns and discuss their relationship to shift type and other discourse-structural factors. Section 5.7 presents close analyses of three discourse excerpts that demonstrate the use of multiple gestural strategies during topic-shifts. Section 5.8 concludes.

5.2 Data & Methodology

All data comes from interviews and monologues on the talk show Late Show with Stephen Colbert, collected through the UCLA television news archive in collaboration with the Red Hen Lab. An initial search for the string "anyway" on the Late Show with Stephen Colbert between 2016 and 2019 yielded 645 results (including repeated clips and commercials). The most recent 150 unique instances were annotated, resulting in a data set ranging from October 2018 to December 2019.[5] The unique corpus ID (UID) is provided for each example. Where only speech is of concern (Section 5.3), the speaker's name and relevant discourse fragment is presented as a simple transcript. Where gesture is also of concern (Section 5.4), a multi-tier transcript of the discourse fragment is provided, using conventions from Embodied Conversation Analysis (Mondada 2018) to highlight gesture-speech alignment.

All examples were coded for the following variables. Detailed definitions of variables are discussed where relevant throughout the chapter.[6]

(57) a. Discourse type: *interview, monologue*[7]
b. Speaker: *Colbert, Guest*
c. Lexical unit: *anyway* and any accompanying discourse markers (e.g. *so anyway*)
d. Word sense: *discourse marker, attitude marker, semantic*
e. Topic shift variety: *full, partial, return*
f. Shift trigger: *speaker, addressee, audience*
g. Turn structure: *cede, keep, pass, take*

[5]Note that this is 50 more instances than the other two case studies in this work. An additional 50 instances were annotated in this case study to account for the polysemy of "anyway".

[6]All annotations, including for those examples excluded from the present study, are available upon request.

[7]For interview data, the name and demographic information of the guest were also recorded, including *occupation, gender, race*, and *nationality*.

All 150 examples were also coded for disengagement and re-engagement actions of the hands, gaze, head, and body using the scheme in (58). Finally, each instance was notated with a 1-2 sentence qualitative description of gestural activity and communicative context.

(58) a. Disengagement & Re-engagement: *yes, no, unsure*
b. Alignment with "anyway": *before, during, after*
c. Disengagement Class: *adjust, break, distance, remove, shrug, stop, other, n/a*
d. Engagement class: *address, focus, locate, present, other, n/a*

Remove, stop, address, locate, and present correspond directly to the action schemas of the same name presented in Chapter 2. *Adjust* refers to self-adaptors (Ekman & Friesen 1969), movements that are self-directed and typically considered non-communicative (e.g. fixing tie or itchin nose). *Break* refers to non-manual gestures that result in the breaking of the speaker-hearer line through gaze aversion or head/body reorientation. *Distance* refers to a non-manual gesture in which the gesturer increases the physical difference between them and their addressee. *Focus* refers to non-manual gestures that reorient the gesturer toward their addressee.

For discussion of how these classes are derived from a gesture's formal features, see Chapter 2. I discuss each gesture class in detail and provide examples in Section 5.4.

5.3 Variation in the verbal mode

The lexical discourse marker "anyway" conveys different meaning and serves different functions in different use contexts, as is the case for discourse markers generally (e.g. Fischer 2000; Fraser 1999). "Anyway" is also subject to phrasal and syntactic variation, co-occurring with other discourse markers, and appearing in different sentential positions. The topic-shifts marked with "anyway" are also subject to variation as to why they occur and the effects they have on the overall discourse structure. In this section, I describe and provide examples of this variation as it is represented in the present data set. I begin by detailing the meanings of "anyway" itself (Section 5.3.1), and then focus on the structural and functional variation in its use contexts. I focus on variation in *shift type* (5.3.2), *shift trigger* (5.3.3), *turn-structure* (5.3.4) and *co-occurring lexical discourse markers* (5.3.5), each of which is likely to affect gestural expression.

5.3.1 Polysemy

Anyway has been noted for its polysemy in English (e.g. Ferrara 1997; Takahara 1998; Urgelles-Coll 2010). As with most work on polysemous lexical items, there is disagreement as to the number and nature of distinct senses.[8] In this work, I do not intend to argue for a particular set of senses. Instead, I settle for making only the distinctions relevant to the goal

[8]See, for example, Tuggy 1993 for a review of the complexities of word senses.

of describing the multimodal expression of topic-shifts. To that end, I identify three broad senses of *anyway*, as exemplified in (59).

(59) a. **Discourse marker (DM)**
TRANSCRIPT 1: DREW BARRYMORE
[UID:658edac8-1ee1-11e9-b705-089e01ba0335,2071]

```
1  DB   So that's part of who I am on the inside and then I got
2       this one the outside, and that brings a whole other thing
3       But sometimes (silly noises).  Anyway_DM, what I'm saying
4       is, I'm stuck here
```

b. **Attitude marker (AttM)**
TRANSCRIPT 2: JOHN LEGUIZAMO
[UID:a9bd5ca4-b038-11e9-86ac-089e01ba0770,2002]

```
1  SC   So would you- would you recommend, perhaps, that he drop
2       out and come back to the city and take care of us here?
3  JL   Yeah, just- just-, y'know, there's too many people
4       running for office anyway_AttM.  Why don't you just step off
5       and fix this and we- we'll talk, then we'll talk
```

c. **Semantic (Sem)**
TRANSCRIPT 3: STEPHEN COLBERT,
[UID:355cfbd2-2c3d-11e9-92ef-089e01ba0770,1576]

```
1  SC   My first guest this evening is so talented she can boil
2       bunnies and try to make coats out of dogs, and we love
3       her anyway_Sem
```

Anyway$_{\text{DM}}$ is, of course, the primary concern of this chapter and consists of those instances in which the proceeding utterance pertains to a higher topic in the discourse structure than the immediately preceding utterance. In (59a), American actor Drew Barrymore is describing her perception of her appearance, her internal persona, and how both shape her experience of the world. After stating that there is a distinction between her inner and outer self (lines 1-2), Barrymore implies that the distinction sometimes makes her feel crazy, making silly noises and moving her index fingers in circles near her temple in an emblematic gesture of insanity. After this brief theatrical performance, she uses *anyway*$_{\text{DM}}$, in tandem with the phrase *what I'm saying is*, to signal that she is returning to address the more general topic of how she experiences the world.

Anyway$_{\text{AttM}}$ signals a dismissive stance held by the speaker toward some referent or proposition. In (59b), Colombian-American actor John Leguizamo uses *anyway*$_{\text{AttM}}$ to emphasize his negative assessment of their being "too many people running for office". When Colbert asks him if one of the primary candidates at the time should drop out of the presidential race, Leguizmo provides both a direct answer ("yeah", line 3), and a justification

for this answer which begins with the negative assessment marked by $anyway_{\text{AttM}}$ and ends with a suggestion for what the candidate should do instead.[9]

$Anyway_{\text{Sem}}$ denotes a kind of exhaustivity roughly equivalent to "regardless" or "despite that". In (59c), Colbert uses $anyway_{\text{Sem}}$ to highlight that *despite* certain undesirable traits, his upcoming guest is well-loved. Note that this is the only sense to *change* the interpretation of the sentence. If, for example, Colbert did not include *anyway* (59c), you would be free to interpret the final proposition *and we love her* as an item in a list of "things that are true about our guest", or even as the *result* of her boiling of bunnies and skinning dogs. In the other two senses, the negative assessment would be maintained and the topic-shift would be achieved without the presence of *anyway*.

In addition to being functionally distinct, the senses are also distinct in their distribution and prosody. Both $anyway_{\text{Sem}}$ and $anyway_{\text{AttM}}$ occur at the end of a clause and are a part of the same prosodic unit. $Anyway_{\text{DM}}$, on the other hand, is generally utterance initial and constitutes an independent prosodic phrase.[10] See Ferrara (1997) for a detailed discussion of prosodic differences.

Though all three senses were present in the data set, $anyway_{\text{DM}}$ was by far the most common, occurring over three times as frequently as $anyway_{\text{Sem}}$ and over ten times as frequently as $anyway_{\text{AttM}}$.[11] The exact breakdown of the 150 annotated examples by sense is given in Table 5.1.

Discourse marker	Attitude marker	Semantic	Total
111	11	28	**150**
(74%)	(7%)	(19%)	(100%)

Table 5.1: Distribution of *anyway* senses

For the remainder of this chapter, I will focus solely on $anyway_{\text{DM}}$. Going forward, all mentions of "*anyway*" refer specifically to its use as a discourse marker. Note that 5 of the 111 instances of $anyway_{\text{DM}}$ occurred immediately adjacent to another *anyway*, as in "anyway, anyway". These are not counted as independent examples when discussing discourse context in the remainder of Section 5.3. Because of this, counts will be presented out of 106 rather than 111.

[9] The "anyway" in the title of this chapter is an example of $anyway_{\text{AttM}}$ sense.

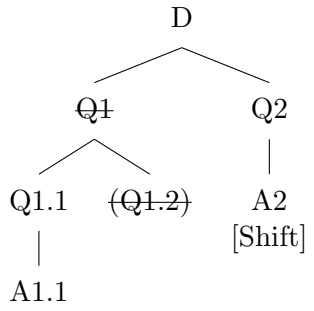
[10] Discourse markers in general tend to occur at clausal peripheries and frequently constitute an independent prosodic unit (e.g. Zwicky 1985).

[11] This is not surprising given the frequency of $anyway_{\text{DM}}$ in spoken, informal discourse. For example, in a study of client-therapist discourse, Ferrara (1994) found that $anyway_{\text{DM}}$ occurred as frequently as once every 44 words (as reported in Ferrara 1997:345). This distribution would likely look very different for written texts, in which discourse topics are not being actively negotiated by participants.

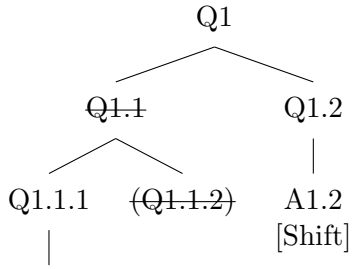
5.3.2 Varieties of shift

In my data, I identify three distinct types of *anyway*-marked topic-shifts which I call *full shifts*, *partial shifts*, and *returns*. All three shift types share two significant discourse-structural features. First, the shift serves to close a question, thus cancelling any potential subquestions, regardless of whether or not the question is considered fully answered by both participants. Second, the utterance that achieves the shift, i.e. the *new* topic, attaches higher in the discourse structure.The three shifts *differ* in (i) the point of attachment of the initiated topic, and (ii) the status of the dismissed topic. Schematic representations of each of the three shifts is given in (60)-(62).

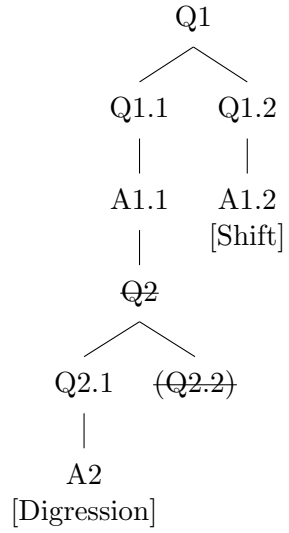
(60) **Full shift**



(61) **Partial shift**



(62) **Return shift**



In both *full shift* (60) and *partial shift* (61) cases, an established line of inquiry is ended, where 'established' means contributing to the main question under discussion. In the diagram above, this is represented as a cancelling of a potential subquestion (e.g. ~~(Q1.2)~~) and a closure of its immediately dominating question (e.g. ~~Q1~~). The difference between the two lies in whether the closed question constitutes a main discourse topic (full shift), or a subtopic within a main topic that is continued by the proceeding utterance (partial shift). In other words, the immediately proceeding utterance in a partial shift pertains to a question that has already been posed, whereas the immediately proceeding utterance in a full shift pertains to an entirely new question. Finally, *returns* (62) are those cases that end a digressive line of inquiry, that is an utterance, or set of consecutive utterances, that do not contribute to answering the main question under discussion.[12] I will discuss an example of each in turn.

[12]These three types of shift are themselves subject to variation and may be differentiated into subclasses. As with identifying distinct senses of *anyway*, I have decided here to limit myself to only making distinctions that are relevant to the data analysis at hand.

In (63), Colbert begins his show by complimenting the energy of the audience. He then performs a *full shift*, marked by *anyway*, to begin his monologue. As it is the beginning of the show, there is no higher discourse topic to return to, and so the shift serves to open an entirely new line of inquiry. This can be mapped to the structure in (60) such that questions regarding "quality of the audience" are considered under Q1, and questions regarding the first act of the show, Colbert's monologue, are considered under Q2.

(63) **Full shift**

TRANSCRIPT 4: STEPHEN COLBERT

[UID:e9570b7e-17fa-11ea-840e-089e01ba0335,116]

```
1  SC   That, that is such a good Thursday crowd.  I wish they could
2       stick around and be the Friday crowd.  Y'know, amazing.
3       Never gonna happen though.  Anyway.  Welcome to the Late Show
4       everybody, I'm your host Stephen Colbert.
```

In (64), American actor Matt Bomer is talking about how he came to know former president Bill Clinton, and begins the topic by recounting a story of seeing the former president at a dinner. Bomer then uses a partial shift marked with *anyway*, shifting up from detailing the reasons for a misunderstanding (lines 3-6) to the punchline of the story – that Bill Clinton really wanted to meet him (lines 7-10). Both the utterance before and after *anyway* address the same general discourse topic, i.e. the story that is being told. However, the utterance before *anyway* is adding specific *explanatory* information – *how did Bomer interpret a specific event?*. The utterance after *anyway*, on the other hand, is adding directly to the narrative structure, answering the recurring narrative question *and what happened next?*.

(64) **Partial shift**

TRANSCRIPT 5: MATT BOMER

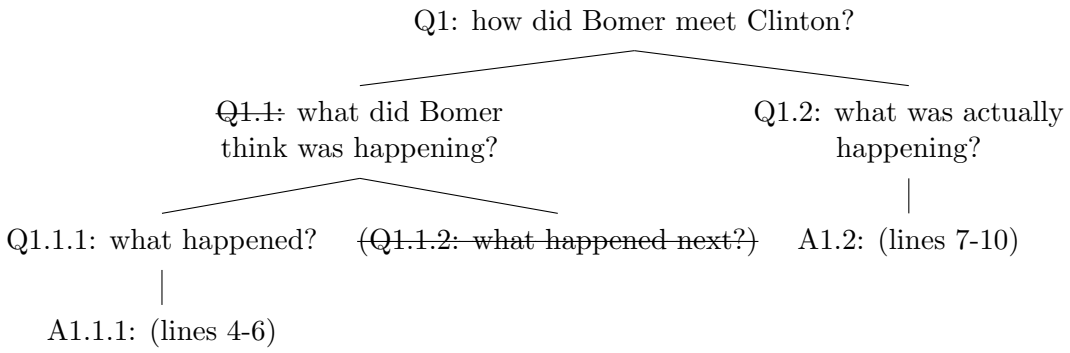
[UID:8b2cf030-89b7-11e9-9a2d-089e01ba0335,2527]

```
1  MB   I looked over across the room a-and Bill Clinton was sitting
2       at a table across the room and he kept looking over at my
3       table.  And so obviously I didn't think he was looking at me.
4       I was like looking behind me, and there was a secret service
5       agent behind me.  Uh, so then I thought there'd been some
6       kind of security breech and we're all in trouble.  Um,
7       anyway, I ended up leaving early.  and the next day a secret
8       service agent approached me and said "hey why'd you leave
9       early last night?  President Clinton really wanted to meet
10      you.  He never misses an episode of White Collar"
```

All utterances in the discourse fragment pertain to a main discourse question – "how did Bomer meet Clinton?". He begins by setting the scene – where he was when he *almost* met Clinton. The utterances immediately preceding "anyway" relate to a sub-question answering Q1.1, roughly "what did Bomer *think* was happening when Clinton was looking at him?".

The utterances immediately proceeding "anyway" relate to a subtopic answering the newly opened Q1.2, roughly "what was *actually* happening when Clinton was looking at Bomer?". This is represented in (65) and can be mapped directly to the schematic representation for partial shifts given in (61) above.

(65)



The most important thing about this structure is that the utterances before and after "anyway" are partial answers to an single ongoing discourse topic regarding how Bomer met Clinton. Rather than marking a change to the discourse goal, "anyway" here marks a change in how the discourse goal is being pursued. Bomer has no more to say about his misunderstanding and signals this with "anyway" before moving on.

In (66), Irish actor Saoirse Ronan is beginning to answer a question posed by Colbert concerning one of her co-stars hiding a pregnancy. Ronan is quickly derailed by Colbert who pokes fun at her, stating that there are generally straightforward ways to tell a woman is pregnant (line 9), and asking her if she'd had the euphemistic "talk" (line 11). Ronan awkwardly engages in this joke (line 12), laughing as Colbert makes a further euphemistic remark (line 13). Ronan once again awkwardly acknowledges this joke (line 14-15), and then tries to resume her story with "anyway" and a framing expression "7 months later".[13] She is then abruptly derailed again as Colbert continues his joke. Again she plays along, and tries again to resume her story, marking the shift with the complex discourse marker string "but so anyway" (line 17). After getting derailed for a third time (line 18), Ronan interrupts Colbert, finishes his thought, and finally successfully continues her story (lines 19-20).

(66) **Return**

TRANSCRIPT 6: SAOIRSE RONAN

[UID:92fedfb8-1b1f-11ea-9518-089e01ba0335,2060]

```
1   SC   I understand, uh, I understand that, uh, Greta was pregnant
2        for the entire shoot and you didn't know.
3   SR   None of us knew.  None of us knew-
4   SC   How did you not know?
5   SR   We didn't, so, we didn't know because she was just like
```

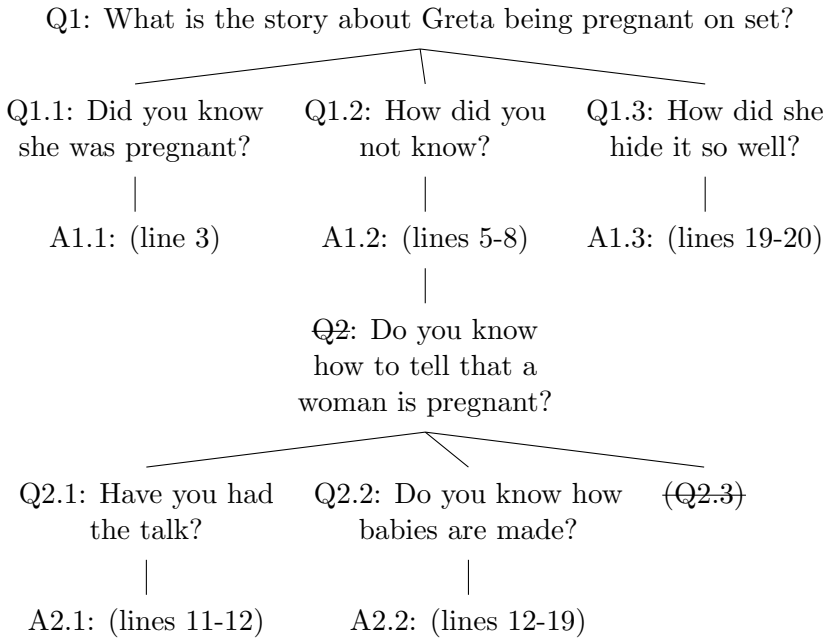
[13] See Krifka (2008: Section 6) for a summary of "frame setters" in discourse and their use in managing common ground.

6 amazing and like had this incredible like command over the
7 set. And we just didn't, it didn't even come into our head
8 But also-
9 SC There are ways to tell that a woman is pregnant, by the way
10 SR Yeah
11 SC I don't know if you've had the talk but...
12 SR I know- *(mumbles)* there- there were babies made. (laughs)
13 SC uh-huh, a firm handshake
14 SR Right *(laughs)*. She, um, so she, y'know, must've done that
15 And she um she- I don't know *(laughs)* Anyway, 7 months later-
16 SC We don't know, we don't know
17 SR We don't know what happened. But s- anyway-
18 SC It could have been immaculate, but-
19 SR There's a baby cooking in there. And it was like seven
20 months in none of us knew

This excerpt can be mapped to the *return* structure in (62). The ongoing topic under discussion relates to the story being told – "what is the story about Greta being pregnant on set?". Ronan gets through two sub-questions in this story (i) "did you know she was pregnant?", and "how did you not know?". After offering her answer to the second sub-question, Ronan begins an elaboration to continue the story, but is derailed. The digression introduced by Colbert, basically a joke about sex, maps to Q2 and is embedded under Ronan's answer to "how did you not know Greta was pregnant?".[14] Ronan's answers to this digressive question does not contribute information about the story itself. Eventually Ronan successfully returns to her elaboration, answering a third sub-question in the narrative – "how did Greta hide her pregnancy so well?". This is represented in (67).

[14]Colbert's contribution can also be considered digressive in that it questions not-at-issue information, i.e. information that should already be in the common ground, and is not immediately relevant for answering the open question (e.g. Simons et al. 2010). Instead of accepting Ronan's answer to the question posed, Colbert questions a presupposition of the question – whether Ronan *could* know that Greta was pregnant. Ronan's answer to this question does not contribute to the story of Greta's pregnancy on set. At best, it may help us understand something about Ronan's reporting of the story.

(67)



The utterance after "anyway" in both the partial shift (64) and return (66) offer partial answers to an already open question, returning the discourse to a more general, but already established goal. The two shift types differ in the status of the utterance immediately preceding "anyway". In the partial shift in (64), the utterance preceding "anyway" contributes the question under discussion by contributing information about what Bomer *thought* was happening when he was at a dinner with Bill Clinton. The same cannot be said for the utterance preceding "anyway" in a return context (66).

As stated, *anyway* signals the closure of a potential question in all shift types. In full and partial shifts, this involves cancelling the expectation that the current line of inquiry will be continued. In return shifts, this involves cancelling the potential for continuing a digression. However, the three shifts differ in how they relate to principles of cooperation. Returns are relatively *cooperative* in that the speaker returns to the cooperative pursuit of a still-open question. Partial shifts are also relatively cooperative in that the speaker maintains their cooperative pursuit of a still-open question, but signals a potential misalignment in the perception of strategies as helpful or not. Full shifts are the least cooperative in that the speaker unilaterally ends the pursuit of a still-open question and poses another.

The distribution of shift types as observed in the present data set is given in Table 5.2, divided by discourse type (*interview* vs. *monologue*). In both contexts, return shifts are most common. This is likely due to the overall prevalence of digressions in free and partially-bound discourses, such as the informal interviews and comedy bits that make up the present data. It is also in line with previous work arguing for "anyway" as marker of resumption (e.g. Ferrara 1997). Partial shifts are noticeably more common in interviews than in monologues. This could reflect the particular interactive function of partial shifts – the speaker acknowledges a potential misalignment in what strategies are perceived as most

helpful in answering the open question by overtly signalling that they are shifting strategies.

	Full shift	Partial shift	Return	Total
Interview	10/55 (18%)	12/55 (22%)	33/55 (60%)	55/55 (100%)
Monologue	16/51 (31%)	4/51 (8%)	31/51 (61%)	51/51 (100%)
Total	26/106 (25%)	16/106 (15%)	64/106 (60%)	106/106 (100%)

Table 5.2: Distribution of shift types in data set

There are also several idiosyncrasies of the monologue format that are worth noting. First, 9 of the 16 full shifts observed during monologues were performed during impersonations of former president Donald Trump. Given Colbert's penchant for making fun of Trump's disfluent, often nonsensical speech, it seems likely that full shifts occur here not as a legitimate discourse move, per say, but as a way to efficiently convey Trump's non-sequiturs. Another 3 of these 16 full shifts occur at transitions in show segments, such as from a commercial break to a monologue, or from pre-show banter to begin the actual show (as in ex. 63). Control over these transition points are unique to Colbert's role as show host. In contrast, the full shifts observed during interviews tend to occur at points of embarrassment or social faux pas – a topic is abruptly discontinued because a participant is made uncomfortable, often to performative and comedic effect.[15]

Second, the majority of return shifts during monologues (22/31) are a kind of recurrent joke construction in which Colbert makes a tangential joke about some referent or phrase in the preceding utterance, and then returns to reporting the news story. These stand in stark contrast to the returns that occur during interviews that are often triggered, not by the speaker themselves, but by some feedback or interruption by an interlocutor. Because of these differences observed between discourse formats, I will focus on variation as it occurs in topic-shifts during interviews for the remainder of Section 5.3.

Based on the predictions for multimodal topic shifting in the introduction to this chapter, we can further specify our predictions based on the form and function of each shift type. These predictions are outlined in (68). For each of the expected gestural discourse markers – *disengagement, manual removal, re-engagement* and *presentation* – shifts types are ordered by how often the particular gestural discourse marker is expected to occur, beginning with the type for which it is predicted to be most prevalent. A "∼" symbol means that no difference is predicted. An explanation of the predictions of relative frequencies is given for each.

[15]These uses align well with Park's (2010) analysis of "anyway" as an impasse marker.

(68) **Gesture and shift type**

a. *Removal gestures*: Full shifts ∼ Partial shifts ∼ Returns
Because (i) removal gestures are used to remove topics as metaphoric objects from the Interaction Space, and (ii) a topic is ended in all shift types, we expect removal gestures to occur with all shift types. However, we expect to only see *clearing* gestures, which clear the Interaction Space of all metaphoric objects, with full shifts – the only shift type to close all open-questions. *Flicking* and *throwing* gestures are likely to be more common with other shift types because they are compatible with the removal of singular objects rather than clearing the space entirely.

b. *Stopping gestures*: Full shifts ∼ Partial shifts ∼ Returns
Because (i) stopping gestures are used to metaphorically stop the forward motion of discourse participants, and (ii) a potential discourse question is being cancelled in all shift types, we expect stopping gestures to occur with all shift types. Continuing the a discourse as expected can be metaphorically understood as forward motion (e.g. Wehling 2017). By extension, *dis*continuing a discourse as expected can be metaphorically understood as stopping forward motion.

c. *Presentation gestures*: Full shifts > Partial shifts ∼ Returns
Because (i) presentation gestures are used to introduce a topic as a metaphoric object into the Interaction Space, and (ii) topics are introduced in all shift types, we expect to see presentation gestures with all shift types. However, because full shifts are the only shifts to introduce an entirely new main discourse topic, we expect presentational gestures to be most frequent in full shifts. This is the only shift type in which the introduced topic is not already in the Interaction Space. Because partial shifts and returns continue a topic already present as a metaphoric object in the Interaction Space, we expect presentational gestures to be the less frequent.

d. *Locating gestures*: Returns ∼ Partial shifts > Full shifts
Because locating gestures are used to draw attention to a metaphoric object that is already in the Interaction Space, we expect to see them only in topic-shifts where the introduced topic is one that has been previously established in the discourse. This means that we expect an absence of locating gestures with full shifts, in which the introduced topic is completely new. We also expect the presence of locating gestures with partial shifts and returns, because the 'introduced' topic is subordinate to a topic that is already present as a metaphoric object within the Interaction Space.

e. *Engagement with the Interaction Space*: Returns > Full shifts ∼ Partial shifts
Because engagement signals participation with the Interaction Space and the metaphoric objects therein, we expect to see *disengagement* at points of topic dismissal, and *engagement* at points of topic introduction, which occurs in all shift types. However, because digressions are (prototypically) not up for further

> discussion, and thus not present in the Interaction Space, we expect less disengagement for returns – if the digressive topic is not in the Interaction Space to begin with, there is no unwanted topic therein to disengage from.

In the proceeding subsections I will discuss how three discourse-structural factors, *shift-trigger, turn structure* and *lexical discourse marker compositionality*, interact with each of the three shift-types. I will also discuss how these interactions further affect our predictions for expressing topic-shifts in the gestural mode.

5.3.3 Shift trigger

As Ferrara (1997) observes, a topic-shift can be motivated in different ways. Ferrara identifies two triggers - the speaker and the addressee. Speaker-triggered topic-shifts serve as a kind of self-regulation – when a speaker feels they have gone off topic, or are perhaps confusing or boring their interlocutor, they may shift topics, saying "anyway" to signal their intention of doing so. Addressee-triggered topic-shifts, on the other hand, occur when the addressee requests some action of the speaker, be it through an overt clarifying question or a less conspicuous backchannel like "really?". Examples of these two trigger types are given in (69).

In (69a), Irish actor Andrew Scott is discussing how to get modern audiences excited about Shakespeare, a topic he is passionate about. In lines 1-2, Scott expresses insecurity, dismissing his own enthusiasm as he says "I don't want to bore you". He then goes on to describe how he thinks audiences could and should enjoy four-hour long Shakespeare plays. At line 10, he ends the topic by saying "anyway I'll stop talking Shakespeare", again dismissive of his own enthusiasm. After Colbert affirms Scott and his interests, Colbert moves to end the interview, making this a full shift. This shift is *speaker*-triggered because it is Scott who moves, unprompted, to end the ongoing topic, likely a result of his expressed insecurity of 'boring' his audience. In Ferrara's terms, this is thus a case of speaker self-regulation – Scott fears he's gone astray of the discourse goals and his interlocutor's interests, and thus regulates his perceived misstep by topic-shifting.

In (69b), Canadian actor Keanu Reeves is describing how he was able to do his own stunts on the set of a recent action film. He is describing a particular scene involving fighting while riding a horse when Colbert interrupts to ask him to clarify his explanation (lines 4-5, "what are you talking about?"). After denying his request for clarification by saying "it's very complicated" (line 6), Reeves shifts back to finish his description of the scene. This is an atypical *partial* shift in which a subquestion of an ongoing topic is posed by Colbert, but rejected by Reeves. Because an answer to Colbert's question would contribute to the established discourse goal of describing how Reeves does his stunts, the discourse segment introduced by Colbert is *not* digressive, despite it being disruptive.[16] This shift is *addressee-*

[16]There is another possible analysis where Colbert's question is interpreted as rhetorical, acting as a kind of comment on Reeve's being unclear, rather than as a genuine request for information. Under such an analysis, Colbert's interruption would be considered digressive, and the shift would be considered a *return*

triggered because Reeve's comment, "it's very complicated", and subsequent shift is a direct response to an action performed by Colbert.

(69) a. **Speaker-triggered**

TRANSCRIPT 7: ANDREW SCOTT
[UID:f00485b2-051e-11ea-8f95-089e01ba0770,3559]

```
1  AS   I'm a bit of a Shakespeare nerd, so- and I don't want to
2       bore you.  But wh-what I find fascinating about it is
3       that, y'know, in order to for Shakespeare to stay
4       relevant, I don't think y- you need to cut it down.  I
5       think you might need to make it as uh- exciting and as
6       thrilling for an audience uh- of today as it would be
7       four-hundred years ago because, y'know, we binge watch
8       TV. Y'know, we watch five hours of television if it's
9       exciting.  So the idea is don't cut it down, just make it
10      four hours of really exciting um play- plays.  Anyway,
11      I'll stop talking Shakespeare.  (laughs)
12 SC   No not at all, not at all.  That's amazing.
```

b. **Addressee-triggered**

TRANSCRIPT 8: KEANU REEVES
[UID:b902d0c2-8112-11e9-92e2-089e01ba0335,1841]

```
1  KR   But they had safety systems.  So they- created this rig
2       that the horse behind the truck with the thing with the
3       wires, so-
4  SC   They strapped a horse behind a truck?  What are you
5       talking about?
6  KR   Well kind- it's- it's very complicated.  But anyway, you
7       won't see that in the movie.  I'll just be riding and it
8       will look- hopefully people will enjoy it
```

Because of the communicative setting, an interview observed by a live audience, I identified the *audience* as a third type of trigger. The audience is distinct from the addressee because the audience is not actively participating in the construction of the discourse; the speaker doesn't *have* to respond. When the speaker does decide to respond, it most usually is in the form of a derailment and subsequent return shift, as is the case in (70). In this example, American comedian Conan O'Brien is talking about his family's history in Massachusetts. When he mentions the city of Worcester, someone in the audience "woo"s. This derails O'Brien's story as he turns to the audience to jokingly reprimand them. O'Brien then marks his return to his story with *anyway*.

rather than a *partial* shift.

(70) **Audience-triggered**

TRANSCRIPT 9: CONAN O'BRIEN

[UID:31292f26-3a62-11e9-928f-089e01ba0335,2230]

```
1  COB  We came to Boston around the time of the civil war.  We just
2       all moved into a very small area, a corner of Worcester
3       Massachusetts, [and we just-]
4  --   [woooo!]
5  COB  Don't "woo" Worcester, no one "woo"s Worcester.  (laughs)
6       It's unwooable.  Anyway.  They moved into a small corner of
7       Worcester Massachusetts, all lived in one house, and married
8       each other for a hundred and eighty years
```

In exceptional cases, such as lengthy digressions, multiple triggers can be involved. For example, in (71), English actor Emma Thompson is describing a recent movie in which she plays a late night talk show host. She makes a joke, saying that it is "basically science fiction" to comment on the lack of female representation in late night. At this point the audience erupts in cheers and laughing. In response to this, Thompson comments on her joke (lines 4-5) and then attempts to return to describing her film, using *anyway* to signal the shift. This shift is audience-triggered because her digression is a direct response to the audience's reaction. Colbert than derails this attempted shift by also commenting on the joke (line 6). After acknowledging his comment, Thompson uses *anyway* again to signal, and this time achieve, the return shift. This second shift is addressee-triggered because the shift is necessitated by Thompson's acknowledgement of Colbert's contribution.

(71) **Mixed trigger**

TRANSCRIPT 10: EMMA THOMPSON

[UID:227f3a58-c570-11e9-a890-089e01ba0335,2100]

```
1  ET   She's a woman late night talk show host, so it's basically
2       science fiction.  And um (.)
3  --   (cheers)
4  ET   Did you see the way I slipped that in?  (.)  Sneaky sneaky
5       little political remark there from Dame Thompo.  Anyway, um
6  SC   We don't need politics in late night
7  ET   Oh nooo, we really don't, good grief.  Anyway, so she's come
8       into the writer's room to talk them and tell them, give them
9       a row, and say "you have got to make this better"
```

As in Ferrara's (1997) findings, speaker-triggered shifts are the most common. This leads Ferrara to conclude that *anyway* was primarily a self-regulating lexical marker. However, the picture is more complicated when we look at the rate of triggers across shift types, as is presented in Table 5.3. In my data, return shifts are are actually triggered slightly more often by someone other than the speaker. External triggers with full and partial shifts, on the other hand, appear to be exceptional.

	Full shift	Partial shift	Return	Total
Speaker-triggered	8/34 (24%)	11/34 (32%)	15/34 (44%)	34/34 (100%)
Addressee-triggered	1/15 (7%)	1/15 (7%)	13/15 (87%)	15/15 (100%)
Audience-triggered	1/6 (17%)	-	5/6 (83%)	6/6 (100%)
Total	10/55 (18%)	12/55 (22%)	33/55 (60%)	55/55 (100%)

Table 5.3: Shift triggers by shift type

This suggests that shift type correlates with particular interaction patterns. Digressions can be triggered by any participant. Closing a non-digressive topic, as in full shifts and partial shifts, are more consistently under the speaker's control. With partial shifts, this may be because of asymmetric knowledge states, as in (64) where only the speaker knew the whole story, and thus only the speaker knew when a certain strategy for answering a question was no longer productive. With full shifts, this may be a matter of politeness – it would be rude for the addressee to trigger the end of a topic that the speaker wanted to continue.[17] In an interview setting, this is especially true for the show's host who is expected to show deference to their guest.[18]

Because each shift trigger corresponds to a distinct interaction between speaker and interlocutor we can make predictions as to gestural expressions of topic-shifts for each shift trigger. These predictions are summarized in (72).

(72) **Gesture and shift trigger**

a. *Removal gestures*:
Speaker-triggered > Addressee-triggered > Audience-triggered
Because the dismissed topic is wholly in the speaker's control in speaker-triggered shifts, we may expect them to also exert the most control over metaphoric objects in these cases. Because of the association of "away" gestures with negative assessment (Bressem & Müller 2014, 2017), we expect fewer removal gestures in addressee- and audience- triggered shift, *except* when the speaker intends to be dismissive of their interlocutor's actions. Doing otherwise would violate the

[17]This can be thought of as a violation of Lakoff's (1973) *give options* rule of polite discourse. Unilaterally ending a discourse topic prevents the other interlocutor from *choosing* to continue. See Chapter 3, Section 2 for more discussion of 'politeness' as a communicative imperative.

[18]See, for example, Goffman (1956) for early discussion of *deference* displays in every day communication and particular social rituals.

politeness principle *be friendly* by overtly devaluing the addressee's contribution (Lakoff 1973). We also expect audience-triggered shifts to exhibit the fewest manual removal gestures because the audience does not have access to the Interaction Space to present metaphoric objects for subsequent removal.

b. *Stopping gestures*:
Addressee-triggered ∼ Audience-triggered > Speaker-triggered
Because stopping gestures prevent the movement of an incoming metaphoric object, we expect to see stopping gestures with addressee-triggered and audience-triggered shifts, but not with speaker-triggered shifts. This is because stopping actions are compatible with stopping the forward movement of another participant (e.g. Wehling 2017), but not self-regulating one's own movement.

c. *Presentation gestures*:
Speaker-triggered ∼ Addressee-triggered ∼ Audience-triggered
Because new topics are introduced regardless of shift-trigger, we do not expect to see differences in presentation gestures across trigger types during topic introduction. Presentational gestures into the primary Interaction Space are expected for all shift-triggers.

d. *Locating gestures*:
Speaker-triggered ∼ Addressee-triggered ∼ Audience-triggered
Because the status of a topic as new or old is independent of who introduced the topic, we do not expect to see differences in locating gestures across trigger types during topic introduction. So long as the introduced topic is related to a metaphoric object already in the primary Interaction Space, locating gestures into the primary Interaction Space are expected.

e. *Engagement with the Interaction Space*:
Audience-triggered > Speaker-triggered ∼ Addressee-triggered
Because interacting with the audience requires the gesturer to orient away from the primary Interaction Space (between Colbert and his guest), we expect to see more disengagement-reengagement sequences in audience-triggered shifts than with speaker- or addressee-triggered shifts. Because navigation between multiple Interaction Spaces is not required speaker- or addressee-triggered shifts, we do not expect to see a difference in engagement patterns.

These predictions are distinct from those made for shift types in several ways. First, unlike with shift types, the rates of *removal* and *stopping* gestures are predicted to differ across trigger types. This is a result of the social pressures that result from different shift-triggers. For example, if a topic is introduced by one interlocutor, the other runs a risk of offending their addressee by ending the topic. If, on the other hand, it is the gesturer that introduced a topic, there is less at stake in the interaction if the speaker also decides to end the topic, removing a metaphoric object as they do so. To alleviate the risk of offense when dismissing a topic introduced by the addressee, speakers may opt out of expressing the dismissal through a removal gesture. This prediction is supported by the observation that 'away' gestures in

general are associated with negative assessment (Bressem & Müller 2014, 2017). Second, the consideration of the audience also makes unique predictions about patterns of engagement with the primary Interaction Space. Because addressing the audience requires the speaker to reorient themselves away from their interlocutor, disengagement and re-engagement with the Interaction Space between the two primary participants should be consistently marked.

5.3.4 Turn regulation

"Anyway" can occur at any point in the turn-taking cycle. Different turn positions indicate distinct interactive functions of "anyway". When the speaker *keeps* the turn after "anyway" they control both the dismissal and initiation of a topic. We have seen an example of this in Matt Bomer's retelling of how he almost met Bill Clinton (ex. 64). When a speaker *cedes* the floor, ending their turn with an "anyway"-marked topic-shift, their interlocutor is left to initiate a new topic. We have seen an example of this in Andrew Scott's self-deprecating topic-shift when he simultaneously ends a topic and his turn out of fear of boring his interlocutor (ex. 69a). When the speaker *takes* the floor, beginning their turn with an "anyway"-marked topic-shift, they also control both the topic dismissal and initiation, but risk offense in ending a topic that their interlocutor may still want to pursue. Likely because of this social risk, the majority of topic-shifts in a *take* position in this data set occur in three very particular contexts: (a) where there has been a lengthy pause and it seems unclear whose turn it is supposed to be, (b) where the addressee has just 'passed' (Schegloff & Sacks 1973), opting to not contribute to the discourse structure, and (c) where the speaker prefaces their topic-shift with a kind of *buffer* acknowledging their addressee's contribution. An example of each of these is given in (73).

In (73a), American actor Nicholas Braun is describing the time he met Bill Clinton at a party. His story is derailed by Colbert holding up a picture of Braun standing next to picture and the audience erupting in loud cheers. Braun stops speaking in embarrassment, attempts to start again using "and" (line 4), but stops as the cheering continues. Once the cheering has finally stopped, Braun *reclaims* the floor with "anyway". This is considered an audience-triggered return shift, as it is the audience's cheers that cause the derailment and Braun's story is continued immediately after "anyway".

In (73b), American actor Andrea Savage is discussing all of the inappropriate things her television network True TV allows on air. After a string of swears and inappropriate descriptions by both Savage and Colbert, the discourse is completely derailed, the audience is laughing, the interlocutors are laughing, and it is not entirely clear how they can return to the discourse. Savage comments on the state of the discourse by yelling "what is happening?" (line 5). They then joke about this being the last of Colbert's shows because of their misbehavior (lines 6-8). Finally, Colbert initiates the return with a passing turn saying "okay, so" to prompt Savage to continue her story (line 9). Savage takes this prompt, using "anyway" to return to the main discourse, and begins telling the story she intended to tell about a time when her and her co-worker were trying to decide if they had taken things too far. This is considered an addressee-triggered return because Colbert both (i) derails the

conversation by escalating the inappropriate situation and (ii) encourages her to return to her story after the lengthy derailment by 'passing' his turn.

In (73c), American actor Sarah Jessica Parker is telling Colbert about how her mother homemade elaborate Halloween costumes for her as a kid, including a lion costume she really liked. Colbert interrupts her story in a jokingly competitive way to say that his mom would have done the same but it would have been too much with eleven children. Parker tries to diffuse the situation by saying she can "see where this is headed" (line 3). Colbert then dramatizes his competition in line 4. Parker takes the floor, acknowledges Colbert's drama by saying "oh my god", and then uses *anyway* to transition back into her story.

(73) a. Transcript 11: Nicholas Braun
[UID:a7ed73a8-b4ef-11e9-a249-089e01ba0770,3431]

```
1  NB   But I see Bill- Bill's also there.  And he's (.)  he's
2       regaling a group of people
3  --   [picture of Braun and Clinton held by Colbert] (cheers)
4  NB   and- (.)
5  --   (cheers)
6  NB   (.)  and so anyway I kind of lurk over him long enough for
7       somebody to say "Oh, hey, Bill, this is cousin Greg, like,
8       he's on Succession, you should-, y'know, you should watch
9       this show, you're gonna love the show"
```

b. Transcript 12: Andrea Savage
[UID:67f2c08a-f628-11e9-920c-089e01ba0335,3472]

```
1  AS   We got away with more than you think
2  SC   But the sway back (beeped swearing) in the middle, you
3       gotta
2  --   (audience laughs)
3  SC   Can we say that?  Can we say that on CBS? None of this,
4       none of this.  [This is our last show]
5  AS                    [What is happening?]
6  SC   This is our last television show.  Thank you for being
7       here
8  AS   Thrilled to be a part of anything [that's the last]
9  SC                                     [Okay] so
10  AS   So anyways, so- so we went back and forth.
```

c. Transcript 13: Sarah Jessica Parker
[UID:67f2c08a-f628-11e9-920c-089e01ba0335,1704]

```
1  SC   Well if my mom only had eight, I would have been a lion
2       but eleven, she can't-
3  SP   Oh I see where this is headed, okay
```

4 SC yes, contest of the wombs
5 SP uh- it's uh, yeah- oh my god. Uh anyway, I liked being a
6 lion

'Buffered' takes are relatively common in the present data set, occurring in 8/18 of "anyway" examples in a take position (73c). In fact, three other such examples have already been discussed. In (66), Saoirse Ronan partially repeats what Colbert has just said before returning to her story, marking the return shift with "but, so, anyway". In (71), Emma Thompson dramatically agrees with Colbert before returning to a summary of her latest film, marking the return shift with *anyway*. And finally in (69b), Reeves rejects Colbert's attempted digression before continuing to discuss an action sequence in his recent film, again marking the return shift with *anyway*. Even though the utterance containing *anyway* is technically not the first in the speaker's turn, these cases are still considered takes because the first utterance, the 'buffer', is playing a direct role in the topic-shift, closing the dismissed topic by acknowledging but not contributing.

A speaker can also 'pass', signaling that they would like to close the immediately open question, and not offering a new topic. Park (2010) discusses these uses in particular as signalling an 'impasse' in which an interlocutor doesn't know what to say, but also wants to signal that they do not wish to continue the current discourse topic. Such examples are rare in the current data set, occurring only twice, but the two that do occur align with Park's observations. Consider the example in (74). In this example, O'Brien is discussing the new format of his show *Conan*, and how it is significantly shorter than the previous format. His explanation is self-deprecating, as he claims that his audience stops caring after half an hour (line 2). When Colbert pushes him to explain more (line 3), O'Brien remains silent for a moment, looking sad, and then says *anyway*. Colbert continues to push the topic (line 5). After O'Brien continues his self-deprecation (line 7), Colbert completes the full shift O'Brien initiated and compliments O'Brien's outfit (line 9). O'Brien's pass in this case is likely for comedic effect – he doesn't want to continue the topic, but also does not initiate a new topic, and instead continues a dramatic portrayal of self-pity.

(74) **Anyway in a *pass***
TRANSCRIPT 14: CONAN O'BRIEN
[UID:31292f26-3a62-11e9-928f-089e01ba0335,1809]

1 COB I said, let's do half an hour because we noticed that our
2 crowd, after a half an hour, they were like "we're good"
3 SC Really?
4 COB (.) anyway
5 SC I think [in show business]
6 COB [I- I got sad-]
5 SC we call that "leave them wanting some"
7 COB yeah (.) no, even at half an hour, they're like you know
8 you could go to fifteen so (.)

```
9  SC   Now there's a- tonight, obviously, you're dressed beautifully
```

The relationships between the turn position of "anyway" and the other two discourse-structural factors discussed are summarized in Tables 5.4 & 5.5. The observed trends are unsurprising and align with previously discussed observations.

	Full shift	Partial shift	Return	Total
Keep	-	10/27 (37%)	17/27 (63%)	27/27 (100%)
Cede	7/8 (87.5%)	1/8 (12.5%)	-	8/8 (100%)
Take	2/18 (12%)	-	16/18 (%)	18/18 (100%)
Pass	1/2 (50%)	1/2 (50%)	-	2/2 (100%)
Total	10/55 (18%)	12/55 (22%)	33/55 (60%)	55/55 (100%)

Table 5.4: Turn structure by shift type

First, only partial and return shifts occur when "anyway" is in the middle of a turn. This reflects the use of "anyway" in self-regulation – you are likely to go on digressions or pursue a question in an unproductive way, and then self-correct. In contrast, you are unlikely to completely change the topic you are talking about in the middle of a turn and without any feedback from your interlocutor. Second, nearly all examples of "anyway" at the end of a turn (7/8 cede contexts) occur with a full shift. This is unsurprising given the potential risk full shifts pose to politeness – if you dismiss a topic that the your interlocutor wanted to pursue, you may be perceived as taking away their choice in how the discourse should progress. By ceding your turn after a topic shift, your interlocutor is invited to determine the next topic, thus alleviating the risk of offense. Finally, examples of "anyway" in a take position occur overwhelmingly with returns (16/18). This is unsurprising given that the majority of return shifts (18/33) are triggered by someone other than the speaker themselves, and thus occur as a response to another participant's contribution.

The correlations between turn-position and shift-trigger are similarly unexceptional. The majority of examples of "anyway" in a keep position (21/27) occur in speaker-triggered shifts. This again reflects the use of *anyway* as a self-regulator – the speaker realizes the digression or unhelpful contribution they've made and shifts accordingly within their turn. The majority of examples with "anyway" in a cede position also occur with speaker-triggered

shifts. This also reflects the self-regulating function of "anyway" – the speaker has deemed their contribution digressive or no longer worth pursuing, but instead of self-correcting and continuing their turn, they simply stop and allow the interlocutor to make the next move. Finally, the majority of examples with "anyway" in a take position (12/18) occur when the shift is externally triggered (by the addressee or audience). All 12 are return shifts, and all are performed by guests, not Colbert. In these contexts, the shift constitutes the speaker's attempt to regain control over the conversation after a derailment.

	Speaker-triggered	Addressee-triggered	Audience-triggered	Total
Keep	21/27 (%)	4/27 (%)	2/27 (%)	27/27 (100%)
Cede	6/8 (75%)	1/8 (12.5%)	1/8 (12.5%)	8/8 (100%)
Take	6/18 (33%)	9/18 (50%)	3/18 (17%)	18/18 (100%)
Pass	1/2 (50%)	1/2 (50%)	-	2/2 (100%)
Total	34/55 (62%)	15/55 (27%)	6/55 (11%)	55/55 (100%)

Table 5.5: Turn structure by shift trigger

Because each turn position of *anyway* corresponds to a distinct interaction between speaker and interlocutor we can again make predictions as to the gestural expression of topic-shifts given particular turn positions. These predictions are summarized in (75).

(75) **Gesture and turn-structure**

a. *Removal gestures*: Keep ∼ Cede > Take ∼ Pass
In keep and cede positions, it is the speaker who made the last contribution to the dismissed topic. This means that they have immediate access to the topic as a metaphoric object for removal, and we expect to see removal gestures. In take and pass positions, it is the addressee who made the last contribution to the dismissed topic. This means that removal gestures in this context would be removing the addressee's contribution rather. Given the relationship between the removal action schema and negative assessment (e.g. Bressem & Müller 2014, 2017), we do not expect removal gestures in this position, so long as the speaker is abiding by principles of politeness (*be friendly, give choices*).

b. *Stopping gestures*: Take $\sim$ Pass $>$ Keep $\sim$ Cede
Because stopping gestures are used specifically to stop an external, incoming force, we expect stopping gestures to occur in cases in which someone other than the speaker has made the last contribution to the dismissed topic. This means that stopping gestures are expected in take and pass positions, but not keep and cede positions. In cede positions, the addressee must make a subsequent action, making a request to *stop* an action infelicitous. In keep positions, the speaker's actions are unimpeded, also making a request to *stop* an action infelicitous (unless the addressee attempts an interruption).

c. *Presentation gestures*: Keep $\sim$ Take $>$ Cede $\sim$ Pass
Because presentation gestures introduce a new topic into the Interaction Space as a metaphoric object, we expect to see presentation gestures whenever the speaker makes the first contribution to the introduced topic. This is the case for keep and take positions. In cede and pass positions, the speaker ends their turn at "anyway", leaving their addressee to make the first contribution to the new topic. Presentation gestures are not expected in these contexts.

d. *Locating gestures*: Keep $\sim$ Take $>$ Cede $\sim$ Pass
Because locating gestures deictically refer to a topic that is already present in the Interaction Space as a metaphoric object, we expect to see locating gestures whenever the speaker is contributing to a reintroduced topic. This is the case for keep and take positions. In cede and pass positions, the speaker ends their turn at "anyway", leaving their addressee to make the first contribution to the reintroduced topic. Locating gestures are not expected in these contexts.

e. *Engagement with the Interaction Space*: Take $>$ Keep $>$ Cede $\sim$ Pass
When an interlocutor is engaged in a turn as speaker, they are expected to be engaged with the Interaction Space in order to signal participation in the discourse and maximize the accessibility of metaphoric object in the Interaction Space. When an interlocutor cedes a turn or passes their turn, they are expected to disengage with the Interaction Space in order to provide room for their interlocutor to engage.

These predictions are distinct from those made for the previous discourse-structural factors because they depend on the interaction structure of the discourse, i.e. *who* is contributing and when, rather than the information structure of the discourse, i.e. *what* is being contributed and when. Because of this, these predictions are relatively independent from those made above and are likely to result in contradictory predictions in some circumstances. For example, stopping gestures are predicted for all shift types and for addressee- and audience-triggered shifts, but are expected *not* to occur in keep and cede turn positions. This means that there are contradictory predictions for gestures occurring with, for example, an addressee-triggered shift in a keep or cede position. I do not make predictions as to whether certain discourse-structural factors have a greater impact on gestural expression

than others. However, the potential for tension between factors is worth keeping in mind when 'unexpected' examples are encountered.

5.3.5 Lexical variation

As discussed throughout this section, the lexical discourse marker "anyway" serves to signal the topic-shift itself, but is underspecified for other discourse-structural factors, including shift type, shift trigger, and turn position. "Anyway" also does not specify the relationships that hold between the dismissed and introduced topics and existing the discourse structure. For example, in a return shift, the reintroduced topic may provide an explanation for a previous utterance or may continue a narrative. The relationships between adjacent discourse segments can be expressed by other lexical markers adjacent to "anyway". In interviews in the present data set, lexical discourse markers adjacent to "anyway" include "and" (3 instances), "but" (15 instances), "so" (14 instances), "uh"/"um"[19] (17 instances), and "well" (1 instance). "Anyway" appears unaccompanied by other lexical discourse markers in only 18 of the 55 interview examples. This stands in stark contrast with monologues in which 46 of 51 examples are unaccompanied by adjacent lexical discourse markers.

A summary of where adjacent lexical discourse markers occur relative the other discourse-structural factors is provided in Table 5.6.[20] Given the preponderance of returns, speaker-triggered shifts, and keep turn-positions makes much of this distribution unsurprising. There are, however, three things to note.

First, "so" occurs most frequently in addressee-triggered shifts and in take positions, despite the relative infrequency of these use contexts. This aligns well with analyses of "so" as a "turn-entry" device (Schegloff 1987) that signals "incipient actions" (Bolden 2009). Under both analyses, "so" signals that the speaker intends to take some action relevant to the discourse structure. In tandem with "anyway" this can serve to signal both the topic-shift and the speaker's intention to immediately contribute to the introduced topic.

Second, partial shifts most frequently occur with "but". Bell (1998) convincingly argues that *but*, in its use as a discourse marker, plays a 'cancellative' function, in that it serves to preemptively cancel some expectation, be it semantic, inferential, or discourse functional. In the current terms of QUD this would mean that *but* serves to preemptively signal a closure of a possible question. This aligns with both the use of topic-shifts generally, and with partial shifts in particular. I argued above that partial shifts are particularly associated with self-regulation – the speaker realizes that their line of inquiry is no longer helpful and shifts to another strategy for answering the still-open discourse question. The cancellative function of "but" may serve to reinforce this self-regulation.

[19]The status of "uh" and "um" as discourse markers is controversial. Often, they are seen as 'hesitations' and 'fillers' rather than substantive markers of discourse structure. See Gunnel (2014) and Clark & Tree (2002) for discussion.

[20]Note that multiple lexical discourse markers can occur in the same example. This means that the counts for each use-context may appear inflated. For example, the string "and so anyway" would appear in the table twice, once as an occurrence of "and" and once as an occurrence of "so".

Finally, full shifts and shifts in turn-cede positions are the only use contexts which appear with bare "anyway" the majority of the time. This makes sense given the particular discourse-structural and interactional contexts in which these shifts appear. In full shifts, the previous discourse topic is completely dismissed and replaced with an entirely new topic. This means that there is no meaningful relationship between the new topic and existing discourse structure that can be marked lexically. In cede contexts, the speaker is simply not making a contribution to the new topic, and thus does not have the opportunity to marks its relation to existing discourse structure.

	and	*but*	*so*	*uh/um*	*well*	∅
Shift type						
full	-	1	-	1	-	8
partial	1	5	2	2	-	4
return	2	9	12	14	1	6
Trigger type						
speaker	2	12	5	10	-	12
addressee	-	3	7	4	1	4
audience	1	-	2	3	-	2
Turn position						
keep	2	10	5	12	-	6
cede	-	1	1	-	-	6
take	1	4	8	5	1	4
pass	-	-	-	-	-	2

Table 5.6: Lexical discourse markers adjacent to "anyway"

To concretize the function of lexical discourse markers adjacent to "anyway", I will briefly consider four examples.

In the present data set, *so* can appear both before and after *anyway*, as demonstrated in (76). The excerpt in (76a) demonstrates the use of the discourse marker string "and so anyway" in a recovery from an audience-triggered derailment (the audience's loud applause). The "and" functions as a continuation marker (Schiffrin 1986), reiterating the intended continuation that was marked with "and" immediately before the derailment. The "so" signals Braun's incipient action – returning to his story, despite the continuing cheers from the audience. "Anyway", of course, marks a topic-shift. The excerpt in (76b) demonstrates the use of "anyway so" in a buffered take position after an addressee-triggered digression. Here, "so" preemptively marks a frame-setting phrase, "she's comes into the writer's room" that centers us in the narrative again after a digression. Both examples have been in previous sections and so will not be re-described in detail here.

(76) a. TRANSCRIPT 15: NICHOLAS BRAUN
[UID:a7ed73a8-b4ef-11e9-a249-089e01ba0770,3431]

```
1  NB   But I see Bill- Bill's also there.  And he's (.)  he's
2       regaling a group of people
3  --   [picture by Colbert toward audience] (cheers)
4  NB   and- (.)
5  --   (cheers)
6  NB   (.)  and so anyway I kind of lurk over him long enough
7       for somebody to say "Oh, hey, Bill, this is cousin Greg,
8       like, he's on Succession, you should-, y'know, you
9       should watch this show, you're gonna love the show"
```

b. TRANSCRIPT 16: EMMA THOMPSON
[UID:227f3a58-c570-11e9-a890-089e01ba0335,2100]

```
1  SC   We don't need politics in late night
2  ET   Oh nooo, we really don't, good grief.  Anyway, so she's
3       come into the writer's room to talk them and tell them,
4       give them a row, and say "you have got to make this
5       better"
```

Of the 15 occurrences of "but" during interviews, all but one occurs before "anyway". An example of "but" in a partial shift is given in (77). In this example, British actor Keira Knightley is talking about the difficulties she's been having with her three-year-old daughter. She first exemplifies the problems by describing how her daughter has been yelling at her and how irrational it is (line 1-3). She then performs a partial shift from this example to the punchline of story in lines 3-6, signaling the shift with *anyway* and reinforcing the cancellation of her previous strategy with "but".

(77) TRANSCRIPT 17: KEIRA KNIGHTLEY
[UID:73d49c28-cb8d-11e8-b552-089e01ba0335,1592]

```
1  KK   They're vocal.  And it's a lot of like "go away!" and
2       slamming these doors, and you think "I can't go away, you're
3       three, you can't look after yourself" but anyway, I was sort
4       of trying this- she was having a tantrum- and I was like
5       "You can't have a tantrum at me, I'm an officer.  You should
6       salute me" or something.  But she doesn't seem to want to do
7       that either
```

Finally, the excerpt in (78) demonstrates the use of a particularly complex string of lexical discourse markers "um but um anyway so". In this clip, American comedian Tig Notaro is describing her comedic interview show "Under a Rock with Tig Notaro". After introducing the show, Colbert asks a clarifying question (lines 3-4), which Notaro pursues for several utterances, providing an explanation and example of how she doesn't recognize famous people.

After providing an example, Notaro initiates a partial shift, deeming the background information she has been providing as no longer helpful in pursuing the main discourse question regarding the nature of her show. Colbert provides a backchannel, "sure" (line 8), signalling that he accepts the clarification she has offered. Notoro then marks the pursuit of a higher level question with "anyway". "But" reinforces the cancellation of the lower level question (i.e. the clarification regarding why she doesn't know famous people). "So" reinforces the intended continuation of an existing topic. The two "um"'s likely signal Notoro's intention her turn in response to Colbert's backchannel (e.g. Clark & Tree 2002).

(78) Transcript 18: Tig Notaro
[UID:88917b30-9325-11e9-b54c-089e01ba0770,3323]

```
1  TN  "Under a rock" is uh it was a show created because I have a
2      hard time recognizing famous people and so-
3  SC  Cause you're not like- you're not plugged into popular
4      culture?
5  TN  I don't follow pop culture.  I follow music and I follow
6      documentaries, but I really kind of miss everything else.
7      And why I didn't maybe didn't know about Shailene
8  SC  Sure
9  TN  um but um anyway so I have people on my talk show, uh and
10     it's me interviewing them trying to figure out who they are
```

The use of multiple lexical discourse marker with a topic-shift move highlights the importance of the principles of multimodal topic-shifting outlined in Section 1, especially the principles of *multiple strategies* and *independence of contribution*. These two principles highlight that for every discourse move there are multiple strategies speakers can use for expression, and when multiple strategies are used at once, each can profile a different aspect of the topic-shift. For example, whereas "anyway" signals the topic-shift as a whole, "but" reinforces the cancellation of a potential question, and "so" reinforces the introduction of a new, but related, question.

When considering the interactive gestures that align with "anyway" it is important to foreground these two principles of multimodal topic-shifting. Just as lexical discourse markers can profile particular parts of the discourse move, so can gesture. The function of an accompanying gesture may align more closely with one lexical marker than another, or may profile an entirely different aspect of the discourse move.

5.3.6 Summarizing verbal variation and multimodal predictions

For each discourse-structural factor discussed above, I outlined specific predictions regarding gestural expression of the topic-shift. Across these factors, there are generalizations to be made. For example, we expect patterns of engagement and disengagement to be correlated with who is speaking when and to whom. We also expect gestural performance of removal

and presentation to correlate with (i) who last contributed to the dismissed topic, and (ii) who is responsible for initiating the next topic. The predictions made in Sections 5.3.2-5.3.4 are summarized in Table 5.7. Checks ("✓") signify that a gestural expression is expected in the given discourse-structural context. An "*X*" signifies that a gestural expression is expected *not* to happen in the given discourse-structural context. Cells are left blank when the discourse-structural factor is not predicted to impact the form of gestural expression.

	Removal	Stopping	Presentation	Locating	Disengagement
Shift type					
full	✓	✓	✓	*X*	✓
partial	✓	✓	✓	✓	✓
return	✓	✓	✓	✓	*X*
Trigger type					
speaker	✓	*X*			*X*
addressee	*X*	✓			*X*
audience	*X*	✓			✓
Turn position					
keep	✓	*X*	✓	✓	*X*
cede	✓	*X*	*X*	*X*	✓
take	*X*	✓	✓	✓	*X*
pass	*X*	✓	*X*	*X*	✓

Table 5.7: Summary of predictions for gestural expression of topic-shifting

The purpose of this section has been to show how different discourse-structural factors can independently contribute to the function and profile of a given topic-shift. Based on the principles of multimodal topic-shifting stated in the introduction, any accompanying gestural behavior may independently profile any of these properties, regardless of their overt expression in the verbal mode. When interpreting the function of a gesture and its contribution to a coherent multimodal, multi-strategic message, we must consider the presence of all structural factors.

5.4 Variation in the gestural mode

In this section, I discuss the observed variation in gestures accompanying "anyway"-marked topic shifts. Unlike in the sections on verbal variation, I discuss gestural expressions in both interview and monologue settings, but maintain that there are likely systematic differences between the two settings. Because of this, all counts are presented out of 106, and the distributions of gestures in interviews and monologues are presented separately.

Based on the predictions made in the introduction, we expect to see independent gestural expressions of topic dismissal and introduction. Two strategies for dismissal and introduction were predicted: (i) managing engagement with the Interaction Space through *stopping* gestures and non-manual reorientation, and (ii) managing the contents of the Interaction Space through *removal*, *presentation*, and *referring* gestures. As shown in Table 5.8, both topic dismissal and introduction are expressed in over half of the instances of *anyway* (61/106). Even more striking, there are only two cases in which neither topic dismissal nor introduction are marked in the gestural mode.

	Dismissal only	Introduction only	Both	Neither	Total
Interview	13/55 (24%)	9/55 (16%)	32/55 (58%)	1/55 (2%)	55/55 (100%)
Monologue	6/51 (12%)	15/51 (29%)	29/51 (57%)	1/51 (2%)	51/51 (100%)
Total	**19/106** (18%)	**24/106** (23%)	**61/106** (58%)	**2/106** (2%)	**106/106** (100%)

Table 5.8: Gestural expression of topic dismissal and introduction with *anyway*

The ordering of topic-dismissal and topic-introduction gestures was strict. There was no example in which a topic-introduction gesture occurred immediately before a topic-dismissal gesture, and vice-versa. The temporal alignment of topic-dismissal and topic-introduction gestures with "anyway" is summarized in Table 5.9. When only one or the other occurs, the interactive gesture overwhelmingly occurs during "anyway". However, when both occur, the distributions get pushed to either side – it becomes more likely that a topic-dismissal gesture will occur *before* "anyway" and that a topic-introduction gestures will occur *after* "anyway".

Gesture type	Before	During	After	Total
Topic Dismissal				
Dismissal only	0/19 (0%)	18/19 (95%)	1/19 (5%)	19/19 (100%)
with Introduction	17/61 (28%)	42/61 (69%)	2/61 (3%)	61/61 (100%)
Total	**17/80** (21%)	**60/80** (75%)	**3/80** (4%)	**80/80** (100%)
Topic Introduction				
Introduction only	0/24 (0%)	18/24 (75%)	6/24 (25%)	24/24 (100%)
with Dismissal	0/61 (0%)	18/61 (30%)	43/61 (70%)	61/61 (100%)
Total	**0/85** (0%)	**36/85** (42%)	**49/85** (58%)	**85/85** (100%)

Table 5.9: Temporal alignment of dismissal and introduction gestures with *anyway*

In addition to the expected expressions of topic dismissal and introduction, I identified a third functional class which I call expressions of *transition*. These gestures are often ambiguous as to their relationship to the topic-shift. Instead of relating to the dismissal and introduction functions, they seem to relate to some mental state of the speaker, such as an unwillingness to act or a readiness to begin anew. Two types of transition gestures were identified – *adjusting* and *shrugs*. Adjusting gestures are self-directed gestures that serve to indicate that the speaker is preparing for a new task or activity without actually initiating that activity. This class of gestural marker may be considered controversial as they resemble a class of bodily behaviors called *adaptors* which are frequently considered *informative*, but not *communicative* (e.g. Efron 1941; Ekman & Friesen 1969; McNeill 1992). Section 5.7.2 will address this issue. Shrugs are complex gestures that maximally involve movement of the hands, shoulders, head, mouth, and eyebrows. They have been associated with a expressions of epistemic stance in the literature (e.g. Debras 2017; Jehoul et al. 2017; Marrese et al. 2021).

The recurrent recurrent classes of topic-shifting gestures identified in the present data set are summarized in Table 5.10.[21] I will by discussing each class individually (Sections 5.5-

[21]For detailed discussion of the action schema associated with each class and related previous findings, see Chapter 2. Action schematic meaning and related literature will be reviewed where relevant.

5.6). After this, I provide close analyses of three discourse excerpts, analysing the extended management of the Interaction Space during topic-shifts (Section 5.7).

Type	Description	Section
Dismissal	marks discontinuation of a discourse topic	
Removal	hand gestures that literally or metaphorically remove an object from Interaction Space, including *clearing*, *flicking* and *throwing* gestures away from the speaker-hearer line	*5.5.1*
Stopping	hand gestures requesting a stop of the discourse through convention (*pausing* gestures) or iconicity (*blocking* gestures)	*5.5.2*
Disengagement	head and body gestures that break the speaker-hearer line including head and body turns away from the Interaction Space and gaze aversion	*5.5.3*
Transition	marks transition between topics	
Adjusting	actions of self-directed re-adjustment including flicking hair and fixing clothing	*5.7.2*
Shrug	full body performance marked by upturned open hands, shoulder raises, head tilts, and facial scrunches,	*5.7.1*
Initiation	marks beginning of a new discourse topic	
Presentation	hand gestures that metaphorically introduce an object into the Interaction Space, including *palm-up open hand* gestures, as well as *containment* and *locating* gestures	*5.6.1*
Addressing	deictic gestures toward addressee without the affordances of presentation	*5.6.3*
Re-engagement	head and body gestures that re-orient the speaker to their interlocutor, including leaning and turning toward the Interaction Space	*5.6.4*

Table 5.10: Types of topic-shifting gestures accompanying anyway

Annotated transcripts and co-indexed screenshots are provided for each example. For particularly complex examples, a representation of the relevant Interaction Space state and question-answer structure are also provided.

5.5 Topic Dismissal

The first primary function of the lexical discourse marker "anyway" is to dismiss the immediately preceding discourse topic, closing any potential questions regarding it. In the data collected, the gestural expression of this function falls into three recurrent gestural classes which I am calling here *removal*, *stopping*, and *disengagement*. These names reflect the action schema associated with the gestures' formal features (Chapter 2), as well as the Interaction Space actions associated with them (Chapter 4). Removal gestures *remove* metaphoric or literal objects from the Interaction Space, preventing further interaction with them (e.g. Bressem & Müller 2014, 2017). Stopping gestures *stop* an action by stopping metaphoric forward movement (e.g. Wehling 2017). Disengagement gestures signal *disengagement* from the primary Interaction Space, and thus the discourse, by turning or leaning away from the center of the Interaction Space (e.g. Gill et al. 2000; Hall 1995; Kendon 2010). Examples of these three dismissal classes are given in Figure 5.1.

Figure 5.1: Three classes of dismissal gesture

The removal gesture performed by Colbert (Fig. 5.1, left) is a diagonal, palm down, *clearing* gesture, beginning in the central Interaction Space between him and the audience and moving toward its lower left periphery. This signals the clearing of all metaphoric objects from the space. This particular example is discussed in more detail in Section 5.5.1, ex. 81. The stopping gesture performed by American actor Drew Barrymore (Fig. 5.1, center) is a *pausing* gesture, in which an index-finger point is held up as if to ask the addressee to "wait a moment". This example is discussed in more detail in Section 5.5.2, ex. 89. The disengagement gesture performed by British adventurer Bear Grylls is a complex non-manual gesture involving both gaze aversion and a head turn away from Colbert.

Topic dismissal is expressed using at least one of these gestural strategies in three-quarters of the data set (79/106). The most frequently used strategy is non-manual disengagement, which occurs in 65 of the 106 examples analyzed. This gesture class involves reorientation away from the central Interaction Space by at least one non-manual articulator (including gaze, head, and torso). Manual removal gestures occurred in a little under one quarter of the data set (25/106). The removal gestures observed exhibit variation in hand shape and orientation, but share an 'away' trajectory. The least frequent gestural dismissal strategy was *stopping*, occurring in only 10% of the data set (11/106).[22] The majority of manual dismissal gestures (23/36) co-occurred with non-manual disengagement gestures. These findings are summarized in Table 5.11.

Gesture class	Interview	Monologue	Total
Removal			
Removal only	1	7	8
Removal+disengagement	9	8	17
Total	**10/55**	**15/51**	**25/106**
	(18%)	(29%)	(24%)
Stopping			
Stopping only	2	4	6
Stopping+disengagement	0	5	5
Total	**2/55**	**9/51**	**11/106**
	(4%)	(18%)	(10%)
Disengagement			
Disengagement only	**32/55**	**11/51**	**43/106**
	(58%)	(22%)	(41%)
Total	**44/55**	**35/51**	**79/106**
	(80%)	(69%)	(75%)

Table 5.11: Gestural expression of topic dismissal *anyway*

In following three subsections, I discuss each of these dismissal classes and their relation to other discourse-structural factors. Though I will be discussing each class individually, it is important to note the compositionality often present in gestural performances of dismissal. For examples, speakers often perform a manual gesture that exhibits properties of both *removal* and *stopping*. Even more frequently, speakers will simultaneously perform manual and non-manual dismissal gestures. This compositionality will be mentioned where relevant and

[22] An additional 4 gestures categorized as *removal* gestures also exhibited formal features of *stopping*, namely open palms oriented away from the gesturer's body. This will be discussed in Section 5.5.2.

addressed directly in Section 5.7. Throughout, the principle of *independence of contribution* is relevant – where multiple gestural strategies are used, each may profile a different aspect of the topic-shift.

5.5.1 Formal and functional variation in removal gestures

Removal gestures are hand gestures that exhibit the physical affordances of removing an object from the gesturer's immediate space. In their use as interactive gestures, *removal* gestures involve movement away from the center of the Interaction Space and the speaker-hearer line. Different hand shapes, orientations, and types of movement may be involved, but in all cases share the trajectory is 'away'. This makes *removal* gestures fundamentally deictic – it is not REMOVAL unless the movement happens in a particular direction relative to the shared space. Formally and functionally similar gestures have been observed across a variety of languages, including English (Streeck 2009)[23], German (Bressem & Müller 2014, 2017), Italian (Kendon 2004), Spanish (Teßendorf 2014), and Syuba (Gawne 2021).[24]

5.5.1.1 Three types of removal

The examples given in Figure 5.2 demonstrate three subtypes of removal gestures which I call *clearing* (G1 & G2), *flicking* (G3), and *throwing* (G4). These three variants are well-documented in the above-mentioned literature on 'away' gestures.

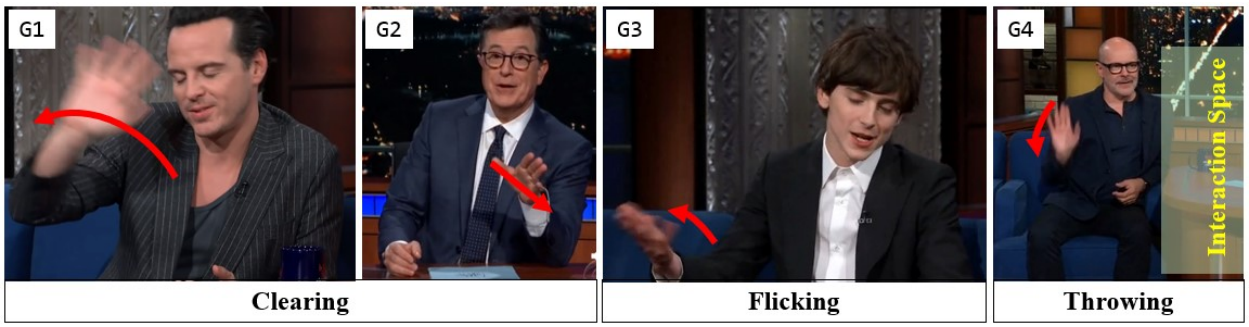


Figure 5.2: Removal gestures corresponding to TRANSCRIPTS 17-20, ex. 79-86

The discourse segment in (79) is an example of a speaker-triggered full shift, in which the speaker (Andrew Scott) cedes their turn, as was discussed previously (Section 5.3.3). In it, Irish actor Andrew Scott ends a lengthy turn in which he appears to be excitedly describing how to make Shakespeare relevant for modern audiences. He is insecure about

[23]In fact, one of Streeck's (2009:192) examples of 'away' gestures as used in discourse management is aligned with "anyway".

[24]As stated, I opt for the term *removal* here, as opposed to Bressem & Müller's (2014; 2017) "away", to foreground the function over the form. That being said, it is the form, in particular the shared "away" trajectory, that determines the presence of a shared *removal* action schema.

boring his audience, and ends his turn in a self-deprecating manner (line 1). As Scott says "anyway, I'll stop talking", he looks away from Colbert, and moves his right hand in a large arcing sweep out from the center of the Interaction Space to his right periphery, as if to clear a vertical surface of objects (Fig. 5.2, G1). While performing this removal gesture, Scott averts his gaze and turns his heard away from Colbert, *disengaging* from the primary Interaction Space.

(79) TRANSCRIPT 19: ANDREW SCOTT
[UID:f00485b2-051e-11ea-8f95-089e01ba0770,3559]

```
1  AS   *Anyway, I'll stop talking* Shakespeare.  (laughs)
        *G1- - - - - - - - - - - -*
2  SC   No not at all, not at all.  That's amazing.  No, that's
3       amazing.  I can talk about Hamlet all night.  They won't-
4       they'll turn the cameras off.
5  AS   (laughs)
6  SC   Well, Andrew, thank you so much for being here
G1:  right hand, open, palm out, arcing sweep out of the Interaction
     Space
```

This sequence can be formally represented as the Interaction Space management move provided in (80). The removal gesture performs a REMOVE action, clearing the contents of the Interaction Space by removing the main topic – *Shakespeare*. Scott's reorientation away from Colbert serves as a DISENGAGE action which signals an intention to cede his turn.

(80) **Removal action by Scott, ex. 79, line 1**

$$IS1_1 = \begin{cases} \text{Participants: AS}_{\textbf{-S}}, \text{SC}_\text{A} \\ \text{Content : } \sout{\text{Shakespeare}} \\ \text{Management :} \\ \quad \textsc{Remove}(\text{G1}\rightarrow\text{Shakespeare}) \\ \quad \textsc{Disengage} \end{cases}$$

This type of gesture was predicted for full shifts in which all open questions are closed, resulting in the metaphoric clearing of all objects from the Interaction Space. Though Colbert's response to Scott's removal action is polite (line 2), he does not actually move to re-engage with the dismissed topic. Instead, he begins a closing sequence to end the interview (line 6).

In (81), Colbert makes a somewhat cruel joke about the relationship between Trump and his son Eric. The audience erupts in cheers and laughter at the end of the joke. Colbert then uses "anyway" to signal a return to his news commentary, swiping an open, down-turned hand down and to his left as he does so (Fig. 5.2, G2), as if to clear a flat surface of objects. Because the audience continues to cheer, Colbert pauses, mouths "what" as if to convey faux-innocence, and then uses "anyway" again to actually achieve the shift. This is an example of an idiosyncratic return shift that is recurrent in Colbert's monologues in

which he momentarily derails the discourse to make a joke, usually a pun. Colbert remains fully oriented toward the audience throughout the sequence.

(81) TRANSCRIPT 20: STEPHEN COLBERT
[UID:eb31a3d4-1317-11e9-81bc-089e01ba0770,1242]

```
1  SC   Trump said "you know what, it's yours, I'm leaving".  Fun
2       fact, that's also what he said at the custody hearing for
3       Eric [(.)  *anyway* (.)  (mouths "what") (.)]
                   *G2- - *
4  --        [(cheers from audience)]
5  SC   anyway Trump declared that the US would be pulling troops out
6       of Syria within thirty days with no conditions
G2:  left hand, palm down, open hand, sweep down and left
```

This sequence can be represented by the Interaction Space management move in (82). As before, the *clearing* gesture serves to remove all contents from the Interaction Space. Colbert remains fully engaged with his addressee, the audience, throughout, and so no DISENGAGE action occurs. This is expected when the speaker keeps their turn, which is indeed the case in this example.

(82) **Removal action by Colbert, ex. 81, line 3**

$$IS1_1 = \begin{cases} \text{Participants: } SC_S, \text{ Audience}_A \\ \text{Content : } \sout{\text{The news}}; \sout{\text{Joke}} \\ \text{Management :} \\ \quad \textsc{Remove}(\text{G2}\rightarrow\text{The news; Joke}) \end{cases}$$

This is a relatively surprising action sequence because *clearing* gestures are predicted *not* to occur in return shifts – there is an established topic that is continued after the shift. The continued topic (*the news*), as a metaphoric object, should not be removed from the Interaction Space. I believe there are two contextual factors that make this seemingly exceptional gesture not so exceptional. First, the joke Colbert has just made is particularly effective, as evidenced by the derailment caused by the audience's laughing and cheering. To recover, it may be easier, in a sense, to start over, clearing the space of all disruptive material, even if that means having to re-introduce a previously established topic. This interpretation is supported in the verbal mode by the use of "Trump" and "Syria". Both referents have been previously mentioned in the discourse, "Trump" just a couple of lines before, and so should be able to be referred to with reduced forms (e.g. "he" and "there"). Instead, the full forms are used, serving to reestablish the referents as contents in the Interaction Space. The following example provides a nice contrast to this.

In (83), American actor Timothée Chalamet is telling a story about being rejected from a job. He cuts himself off to shyly ask if he is allowed to say the name of the producer. Colbert assures him that he is allowed to do so, but Chalamet falters again. Instead of offering the name, Chalamet performs a speaker-triggered partial shift, marked by "anyway", dismissing

the subquestion of the producer's identity to continue the main topic of how he got rejected. As he says "anyway", he flicks his right hand away, as if to toss a small object to the side. Chalamet averts his gaze during this removal, but maintains his head orientation toward Colbert.

(83) TRANSCRIPT 21: TIMOTHÉE CHALAMET
[UID:4b3223f4-ea3b-11e8-985b-089e01ba0770,1862]

```
1   TC   there was an old Letterman producer named- can I say- am
2        I allowed to say that while I'm on here?
3   SC   uh sure
4   TC   Yeah, okay, alright, uh, that uh, and his name-
5        *uh anyway* he was casting a movie here
         *G4- - - -*
G4:  right hand, palm up, loose, flicks out at wrist twice
```

This sequence can be represented by the Interaction Space management move in (84). The *flicking* gesture performs a REMOVE action. Unlike with the *clearing* gestures discussed above, this REMOVE action does not result in the full clearing of Interaction Space contents. The flicking gesture Chalamet performs exhibits the physical affordances of removing a single object. As such, the REMOVE action targets only the most recent topic – the producer's name. The producer's name is an embedded topic under Chalamet's rejection; this is represented by the embedded parentheses.

Chalamet's gaze aversion performs a DISENGAGE action which should signal an intention to cede his turn, but this does not seem to be the case. As noted, the gaze aversion here is qualitatively very different from the disengagement performed by Andrew Scott in (79). Scott's disengagement involves his entire body and does seem to signal his intention to cede the turn. Considering this, Chalamet's relatively mild disengagement may serve to signal disengagement from the particular topic being dismissed, rather than the discourse as a whole.

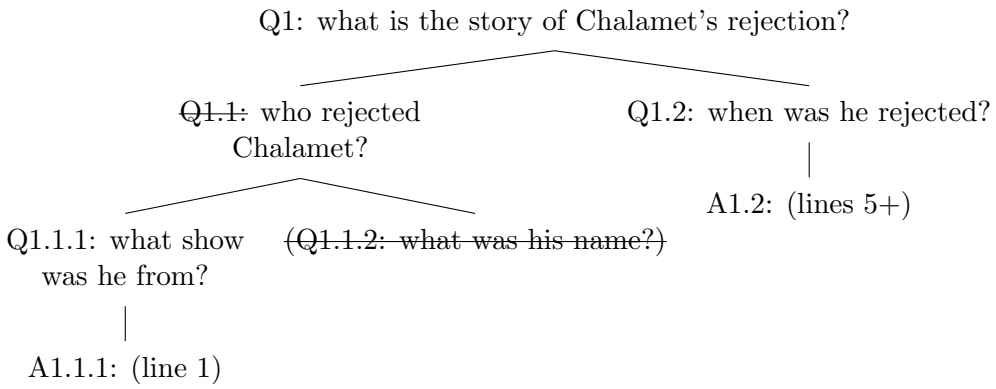
(84) **Removal action by Chalamet, ex. 83, line 5**

$$IS1_1 = \begin{cases} \text{Participants: } TC_{\text{-S}}, SC_{\text{A}} \\ \text{Content : Chalamet's rejection(Producer(}\sout{\text{name}}\text{))} \\ \text{Management :} \\ \quad \text{REMOVE(G3}\rightarrow\text{name)} \\ \quad \text{DISENGAGE} \end{cases}$$

To clarify the embedded structure of this example, consider the diagram in (85). Chalamet sets the scene of his story by providing information about who rejected him. This serves as the first sub-question in his narrative (Q1.1). He begins by providing information about what show the producer was from, the Letterman Show (line 1). This serves as the answer to a further embedded sub-question (Q1.1.1). He then begins to provide additional information, the producer's name, but stops himself. This serves as another embedded sub-question

(Q1.1.2) which is subsequently cancelled, unanswered, by Chalamet. This cancellation is marked by his use of "anyway" and the removal gesture. This cancellation closes Q1.1 – Chalamet will provide no further information about who rejected him. Instead he moves on with his main narrative, addressing the next sub-question regarding when he was rejected and what from (Q1.2).

(85)



If Chalamet were to use a *clearing* gesture in this context, we would expect all open questions to be closed. The *flicking* iconically enacts the removal of a small object. This enactment, in context, is interpreted metaphorically as the cancellation of a sub-structure within the ongoing discourse.

The crucial difference between this and the previous example (ex. 81) is the effect the gesture seems to have on the verbal referent forms used immediately after "anyway". In the previous example with Colbert, we saw that he used a full form, "Trump" and "Syria", as he would if they were new referents to the discourse, even though they weren't. Chalamet, on the other hand, maintains a reduced form, referring to the producer as "he" (line 5). This supports an interpretation in which the relevant referent is still accessible (both verbally and physically) in Chalamet's case, but not in Colbert's case. In other words, the *clearing* gesture performed by Colbert removes all contents, and so referents must be fully reintroduced. The *flicking* gesture performed by Chalamet removes only the question of the producer's name, leaving a representation of the producer himself present in the space and available for subsequent management.

The final sub-type of removal gesture is a *throwing* gesture, depicted in Figure 5.2, G4. In the relevant discourse excerpt, given in (86), American actor Rob Corddry has just told Colbert he has been married for 17 years. As the audience begins cheering loudly, Corddry first claps along with them, then dismisses their enthusiasm, claiming "they weren't going to clap at that" (line 1). After Colbert points out that the audience gets instructions when to cheer (line 2), Corddry dramatically sighs and dismissively flicks his hand down toward the audience, as if to toss a small unwanted object to the ground (line 3, G3). He then returns his attention to Colbert and returns to his story about meeting his wife. This is an audience-triggered return shift, in which Corddry is openly dismissive of the audience's interruption. The dismissal of the digressive topic (the audience's cheers) is signalled by "anyway" and the *throwing* gesture. The *throwing* gesture may also contribute

to the expression of Coddry's negative assessment of the audience, which his co-expressed in the verbal mode by his dramatic sigh. The continuation of his story is signalled by the subsequent lexical discourse marker "so".

(86) Transcript 22: Rob Corddry
[UID:e5b2d83e-fd3a-11e9-aee3-089e01ba0335,2766]

```
1   RC   they weren't going to clap at that
2   SC   We'll hit the button, we'll hit the button
3   RC   *(sigh)* anyway um so uh I was uh I had a date scheduled
         *G3- - *
4        with this beautiful woman for Saturday
G3:  right hand, palm out, open hand, flicks down at wrist
```

Similar to the *flicking* gesture produced by Chalamet, Coddry's *throwing* gesture exhibits the physical affordances of removing a single object. In this case, the removed object is the digressive topic regarding the audience's reaction. The one interesting difference between this and the *flicking* gesture is the trajectory of movement. *Flicking* gestures are characterized by quick, loose lateral movements, as if throwing something aside. *Throwing* gestures, on the other hand, are characterized by forward movement. One may argue that this is not movement away from a *shared* space, but rather only the speaker's immediate space. However, the non-manual gestural context suggests otherwise. Before the digression, Coddry is engaged in the primary Interaction Space between him and Colbert as he begins to tell his story. He then *disengages* from this space to interact with the audience. Upon ending the digression, Coddry turns back. In this case then the 'forward' movement is still movement away from the shared space between him and Colbert (see again the location of the Interaction Space relative to the performed gesture in Figure 5.2).

Despite the variation in size, shape, and orientation, all of these gestures serve to remove metaphoric objects from the main Interaction Space. In G1 and G2 this is accomplished through *clearing* movements, performed with a flat hand and lateral movement. In G3 and G4 this is accomplished through loose tossing movements, one to the side (*flicking*) and one forward toward a secondary addressee (*throwing*). All three sub-types of removal gestures have been observed elsewhere in the literature on 'away' gestures (e.g. Bressem & Müller 2014, 2017; Teßendorf 2014).

5.5.1.2 Idiosyncratic removal

In this section a final type of removal gesture that takes advantage of the unique spatial affordances of Colbert's opening monologue. In these idiosyncratic removal gestures, Colbert finishes a joke by holding a very specific metaphoric object, usually involved in a pantomime, with one hand, and then moves the metaphoric object off screen. This is idiosyncratic both in that it is not observed in the data outside of monologues, and that it depends on two affordances unique to the setting and format of the opening monologue. First, it requires that Colbert can move his hand outside of the camera frame, which is not possible in the

format of interviews where performers are seated in the center of the shot, limiting their mobility. Nor is it possible in everyday conversation, where there is generally no camera frame to move out of. Second, the performance is distributed across Colbert, who performs the gesture, and his pianist, Jon Baptiste, who plays a discordant chord as the metaphoric object is placed onto a metaphoric surface off screen. Of course, Colbert's interviewees do not have access to these rehearsed multiparty performances, nor do people in every day life. The point of discussing this idiosyncratic gesture, despite it not being generalizeable across speakers or contexts, is that it nonetheless abides by the generalization made about removal gestures – some metaphoric object is removed from the primary Interaction Space.

In this excerpt, Colbert is reporting on a then recent drone attack on a Saudi oil field. The perpetrator had not been identified, but the Iranian government was suspected. Colbert comically critiques the evidence that Iran did it by joking about the availability of drones to everyday consumers – governments have drones, but so does Best Buy (lines 1-2). He then acts as if he is "your dad" with a drone, trying to take a family photo (lines 3-6). During this enactment, he pantomimes controlling a drone's joy stick (G2 & G4). He ends the joke with the dad accidentally blowing up an oil refinery (line 6), after which he removes the "joy stick" from the frame (G5).

(87) Transcript 23: Stephen Colbert
[UID:47198e46-d915-11e9-9fe6-089e01ba0335,200]

```
1  SC   yes, drones are highly advanced tech, so they've narrowed
2       down the suspect to *Iran / or your dad who just bought a
                           *G1a-/G1b- - - - - - - - - - - - - - ->
3       quad-copter at Best Buy.  "Totally sweet.* *We're taking the
        <- - - - - - - - - - - - - - - - - - - - -* *G2 - - - - - ->
4       family picture, we're taking the family picture this way this
        <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - >
5       year.* *We're doing the Christmas card.* *Everybody in the
        <- - * *G3- - - - - - - - - - - - - - -* *G4- - - - - ->
6       yard, everybody.  Oh, I blew up an oil refinery"* *[     ]
        <- - - - - - - - - - - - - - - - - - - - - - * *G5 - - >
7  --                                                   [(piano)]
8  SC   anyway* *our top intelligence officials* think Iran did it
        <- - -* *G6- - - - - - - - - - - - - - - *
```

G1: right index finger pointing up held to right, then left index finger pointing up held to left

G2; G4: right hand clenched on open left hand, pantomime of controlling a joint stick

G3: open right hand held to upper right in repeated beckoning motion

G5: right hand in cup shape, moves from surface of left hand rightward off screen

G6: right index finger pointing up held toward audience

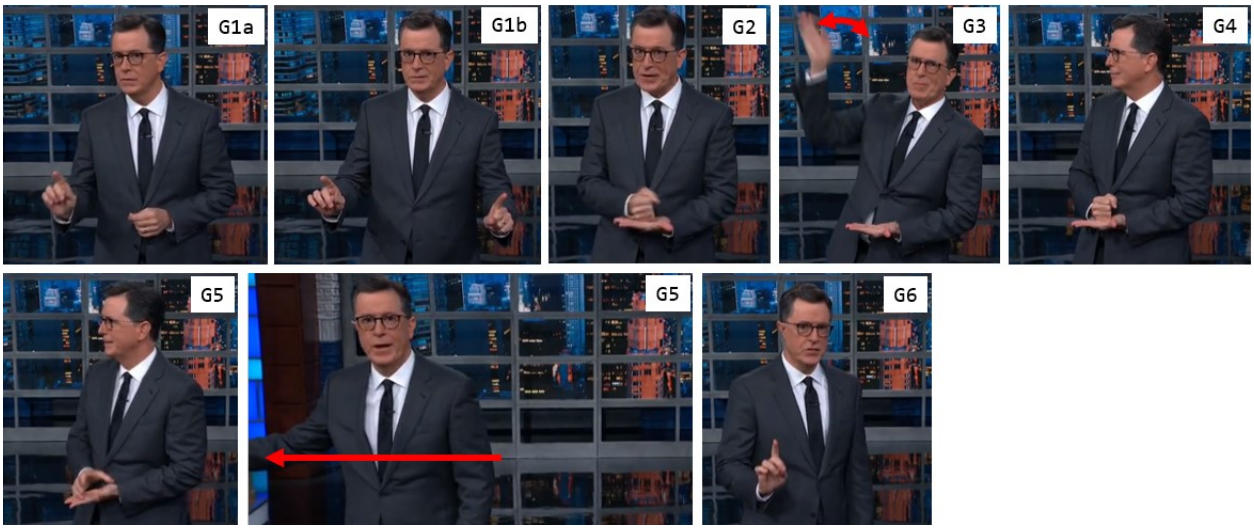


Figure 5.3: Gesture sequence corresponding to TRANSCRIPT 23: STEPHEN COLBERT, ex. 87.

Note that Colbert uses a reduced verbal form immediately after the removal gesture and "anyway", using "it" to refer to the drone attack (line 8). As in the Chalamet example, Colbert's removal gesture is consistent with the removal of a singular object rather than the clearing of the entire space. This serves to metaphorically remove the joke from the Interaction Space while maintaining the presence of the ongoing topic.

This removal gesture is exceptional. It is dramatically and intentionally performed, rehearsed by both Colbert and his band lead, and is dependent on the particular physical space in which it occurs. Despite all this, the generic REMOVAL action schema is evoked, and its use as a means for Interaction Space management is the same. This can be analogized to a creative use of a very conventional metaphor. Saying "rents have shot into the stratosphere" evokes the very primary metaphor MORE IS UP (Lakoff & Johnson 1980) while also being in some way *special* – it is said by a particular type of person in a particular circumstance. Despite its relative idiosyncrasy, we can still learn from it. From particularly creative uses of both verbal metaphors and gestures, we can learn something about how more general patterns can be exploited in effective communication.

5.5.1.3 Distribution of removal gestures

Now that we have seen the forms and use contexts of removal gestures in the data set, let us compare their occurrences to the predictions made in Section 5.3. First, it was predicted that removal gestures would be equally distributed across shift types because topics are dismissed in all cases. As shown in Table 5.12, this prediction is not met. Nearly all of the observed

removal gestures (23/25) occur with full or return shifts. They also occur at the highest rate in full shifts.

	Full shift	Partial shift	Return	Total
Interview	4/10 (40%)	1/12 (8%)	5/33 (15%)	10/55 (18%)
Monologue	7/16 (44%)	1/4 (25%)	7/31 (23%)	15/51 (29%)
Total	11/26 (42%)	2/16 (13%)	12/64 (19%)	25/106 (24%)

Table 5.12: Distribution of removal gesture by shift type

This unexpected finding invites us more carefully consider the functional differences of shift types and removal gesture variants. First, partial shifts, which occur the least frequently with removal gestures, are the only shifts in which the two utterances adjacent to "anyway" contribute to answering a single discourse question. This suggests that the information conveyed before "anyway" may still be relevant to the ongoing discourse goal. If this is the case, removal of the topic may be infelicitous – the information is still relevant and so should remain in the Interaction Space. Second, removal gesture types are not evenly distributed across shift types. For example, four of the five removal gestures that occur in interviews with partial shifts are either *flicking* or *throwing* gestures, i.e. gestures that serve to remove particular metaphoric objects rather than clear the space of all objects. In contrast, none of the four removal gestures that occur in full shifts during interviews are *flicking* gestures; two are *clearing* gestures, one removes a physical prop from the stage, and one pantomimes the placement of a book onto a shelf.[25]

Finally, this 'unexpected' finding is somewhat complicated by the fact that full shifts most frequently occurred (i) when triggered by the speaker and (ii) when the speaker ceded the turn. Removal gestures were predicted to be more common in both of these contexts. This suggests that removal gestures may be more closely related to the expression of shift trigger and turn position, rather than shift type.

When we considered different shift triggers (*speaker*, *addressee* and *audience*), I suggested that removal gestures should be most common in speaker-triggered shifts because it was this context in which they were in most direct control of the Interaction Space contents – it is the speaker's *right*, in a way, to remove an object when they are the cause of the shift. In contrast, I suggested that removal gestures would be less common in addressee-triggered shifts because of the association of the "away" trajectory with negative assessment (e.g. Bressem & Müller 2014, 2017) – a removal gesture in an addressee–triggered shift gesture

[25]This last example is discussed in Section 5.7.1.

my be construed as the speaker being dismissive of their interlocutor's contribution, thus violating principles of politeness. The data supports these predictions, as summarized in Table 5.13.[26]

	Speaker-triggered	Addressee-triggered	Audience-triggered	Total
Interview	6/34 (18%)	2/15 (13%)	2/6 (33%)	10/55 (18%)
Monologue	15/47 (32%)	- -	0/4 (0%)	15/51 (29%)
Total	21/81 (26%)	2/15 (13%)	2/10 (10%)	25/106 (24%)

Table 5.13: Distribution of removal gestures by shift trigger

The four exceptional examples occurring with addressee- and audience-triggered shifts are significantly less exceptional upon further inspection. One audience-triggered shift with a removal gesture was presented in (86). As discussed, Rob Coddry is being actively dismissive of the audience's reaction in this example. This makes a removal gesture charged with negative assessment fully appropriate. The other is a complex sequence performed by Renée Zellweger in which she first removes a metaphoric book she has been discussing and then performs what appears to be a self-effacing shrug as the audience cheers. This example will be discussed in more detail in Section 5.7.1.

One of the two removal gestures occurring with an addressee-triggered shift is performed by Irish actor Saoirse Ronan after Colbert has interrupted her story twice. This removal gesture is compositional, performed with a blocking hand orientation associated with stopping an action. This example is discussed in detail Section 5.7.3, but, needless to say, also seems to be intentionally dismissive.[27] This leaves only one unexplained outlier, a small flicking gesture toward the audience performed by American actor Matthew McConaughey.[28] The relative under-articulation of his gesture performance likely minimizes any perception of McConaughey as intentionally aggressive toward Colbert.[29]

Finally, removal gestures were expected to be more frequent with shifts in which the speaker keeps or cedes the floor than when they take the floor or pass. This prediction is motivated by the fact that in the first two contexts, the speaker has been engaged with the

[26]It is worth noting that speaker-triggered shifts are disproportionately represented in the data, making up 79% of the total shifts when one includes the data from monologues. Monologues are never addressee-triggered (since there is no interactive addressee) and only rarely audience-triggered.

[27]See Sweetser & Sizemore (2008) for discussion of emphatic gestures used when speakers feel they have a particular 'right' to the floor, such as after an interruption.

[28]Clip retrievable at: UID:8642fe04-3bf4-11e9-9fda-089e01ba0770,2861

[29]Thank you to Eve Sweetser for pointing out this interpretation.

dismissed topic up until the topic shift – it is their immediately preceding utterance that serves as the final contribution to the closed topic. Thus, when a speaker performs a removal gesture while keeping or ceding the floor they are, in a sense, removing their own contribution. A removal gesture in a take or cede context would be removing the addressee's contribution, again risking an interpretation of negative assessment and a violation of politeness. These predictions are again supported by the data, as summarized in Table 5.14.

	Keep	Cede	Take	Pass	Total
Interview	2/27 (7%)	4/8 (50%)	4/18 (22%)	0/2 (0%)	10/55 (18%)
Monologue	15/47 (32%)	- -	0/4 (0%)	- -	15/51 (29%)
Total	17/74 (20%)	4/8 (50%)	4/22 (18%)	0/2 (0%)	24/106 (24%)

Table 5.14: Distribution of removal gestures by turn position

Three of the four exceptional cases occur with buffered takes. This means that the relevant turns start with some brief reference to the topic that is subsequently dismissed. These buffers may serve as sufficient contributions for the purposes of removal. The final 'exception' is the 'throwing away' gesture performed by Rob Coddry. Again, this particular gesture is directed toward the audience, not toward Colbert. Because Coddry is dismissing the actions of the audience, he is not at risk of impolitely dismissing Colbert's immediately preceding contribution.

5.5.2 Stopping

As described in Chapter 2, two stopping gesture variants were identified in the present data set: *blocking* and *pausing*. *Blocking* gestures exhibit the physical affordances of stopping an incoming object from entering the gesturer's immediate space, and are used to metaphorically stop the addressee's actions or the discourse from 'moving forward' (e.g. Wehling 2017; Sweetser 2022). *Pausing* gestures are performed with an index-finger pointing upward, as if to ask the addressee to "wait a second". These gestures exhibit the same palm orientation as *blocking* gestures, partially evoking the same BLOCKING action schema, and compose with the conventional numerical '1' hand shape.

The removal gestures discussed above perform discourse management by managing the *contents* of the Interaction Space. Stopping gestures, on the other hand, perform discourse management by managing *engagement* with the Interaction Space, primarily by requesting

that the addressee cease some action.[30] Stopping gestures are inherently underspecified for exactly what actions the interlocutor is instructed to discontinue. A given stopping gesture may serve as a request to stop either an inference or a line of inquiry.

Stopping gestures are relatively infrequent in the present data set, occurring only 15 times total, and only 4 times in interviews. 4 of the 15 are complex gestures composing a *blocking* hand shape and orientation with a *removal* trajectory. To complicate matters, 8 of the 15 occur as Colbert is impersonating Trump. Colbert appears to use a two-handed blocking gesture as a cue for his Trump persona, and so it is unclear whether or not these examples should count as interactive gestures with the rest. Because of all of this, there is not sufficient data against which to test the predictions made in Section 5.3. Nonetheless, I discuss the apparent function of stopping gestures with "anyway"-marked topic-shifts in two examples below.

In (88), Colbert is in the middle of a joke in which he is impersonating Trump. As Trump, Colbert makes an imaginary phone call to a pizza delivery service (lines 1-2), holding his hand to his ear in an emblematic phone gesture. Instead of a pizza delivery service, his attorney, Rudy Giuliani, supposedly answers the call. He apologizes for mistaking "shaky Italian guy" as a pizza service (lines 2-3). Colbert, as Trump, then signals a shift with two consecutive "anyway"'s. At the first "anyway", he performs a one-hand blocking gesture with his palm facing into the imagined Interaction Space (G4). At the second *anyway*, he maintains this hand shape and orientation while sweeping outward, turning the stopping gesture into a compositional removal gesture (G5). The outward movement is repeated four times, indicated by the "/" marks in the gesture tier below. Half-way through this repeated removal gesture, Colbert gets to the punchline of the joke, reiterating that he called to order a pizza. He then acts as if to hang up the phone (G6), humorously exclaims that he "forgot the garlic knots", and then finally ends the impersonation and returns to his monologue.

(88) Transcript 24: Stephen Colbert
[UID:bf03ea3c-1731-11ea-a000-089e01ba0335,441]

```
1  SC   Hello (.)  Shaky's Pizza?  I'd like to order a large
2       pepperoni (.)  *Oh hi Ruddy, I'm sorry you're just-* *you're
                       *G1- - - - - - - - - - - - - - - - - -* *G2- ->
3       in my contacts* *as "shaky Italian guy" (.)* *Anyway*
        <- - - - - - -* *G3- - - - - - - - - - - -* *G4- - *
4       *anyway/ sorry about that/ sorry about that (.)  Look/ the
        *G5- - /- - - - - - - - /- - - - - - - - - - - - - -/- ->
5       point is I would like a large pepperoni, thank you* [*(.)*]
        <- - - - - - - - - - - - - - - - - - - - - - - - - -* *G6-*
6  --                                                        [(piano)]
7       *oh!  Forgot the garlic knots (.)* *another guy whose phone
        *G7- - - - - - - - - - - - - - - - * *G8- - - - - - - - - -≫
```

[30] See discussion of Request actions in Chapter 4.

```
8        records are sparking interest is GOP congressman and man
9        indicating how much dignity he has left, Devin Nunes
G1:  left hand, loose, palm up
G2:  left hand, flat, palm up, indicating imagined book of contacts
G3:  left hand, flat, palm away, fingers toward audience
G4:  left hand, flat, palm away, blocking
G5:  left hand, flat, palm away, sweep up and out
G6:  left hand held in block, right hand places imaginary phone off
     screen
G7:  both hands, fists, palm center, held at chest level
G8:  both hands to rest, step forward toward audience
```

Figure 5.4: Gesture sequence corresponding to Transcript 24: Stephen Colbert, ex. 88

There are four gestural performances that of interest. First, the two gestures coinciding with the two instances of "anyway" (G4 & G5) demonstrate two formal variants of the *blocking* gesture. The first is simple, held at chest level, and directed at an imagined interlocutor in front of him. The second is complex, composing with an *away* trajectory to create a

clearing gesture, much like the one we saw performed by Andrew Scott in Section 5.5.1.1. Both gestures may additionally serve as a cue for the Trump persona, along with G3 and the left hand in G6. This is a clear demonstration of the multifunctionality of interactive gesture.

Second, Colbert performs a variation on the idiosyncratic removal gesture discussed in Section 5.5.1.2. As he enacts hanging up his imaginary phone, he reaches his hand outside of the camera frame. This is accompanied, as before, by a discordant piano chord as the imagined phone is place on an imagined surface off screen. This is an important demonstration of the principle of *optionality of expression*. The idiosyncratic removal gesture performed both here and in the previous section express then ending of a discourse topic (a joke, in both of these cases). In the previous example we saw, the gesture was accompanied in the verbal mode by "anyway" such that the topic-dismissal was, in a way, double-marked. In this example, the ending of the joke is only marked in the gestural mode. This demonstrates the ability of gestural topic-dismissal markers to occur *without* an accompanying lexical marker.

Finally, Colbert is turned away from the camera for the entirety of the joke, turning to his right to have the imaginary phone call. In this orientation, Colbert is *disengaged* from the primary Interaction Space between him and his audience, and *engaged* in a secondary imaginary space to his right.[31] He only returns to the primary Interaction Space between him and the audience after he has 'hung up' the imaginary phone and ended his impersonation. As he returns to his monologue, he directly faces the camera and steps forward, further engaging in the primary Interaction Space. Once again, this reorientation signals topic introduction *without* an accompanying lexical marker.

In (89), American actor Drew Barrymore performs a *pausing* gesture, aligned with "anyway", as she returns from a speaker-triggered digression. In the excerpt, Barrymore is describing the duality of her internal life and her public persona, relating it tangentially to a movie in which she plays two different characters. The gestures she performs during this sequence are large and theatrical. The first two convey primarily representational information. The first is a self-directed gesture, deictically referring to herself as she says "who I am on the inside" (line 1, G1). She then points to her face, moving in large circles, as she contrasts who she is on the outside to who she is "on the outside" (lines 1-2, G2). G3 is the first interactive gesture in the sequence – a two-handed shrug signalling that she won't offer a further explanation of the difference (e.g. Debras 2017). She then performs four consecutive gesture with the same index-finger point hand shape. The first is an emblematic gesture signaling insanity (G4) which she performs while making a silly face and making random noises. The second is the *pausing* gesture of interest (G5), which is performed as she says "anyway", signaling that she is stopping the silliness and returning to the main discourse. The third maintains the orientation of the pausing gesture, but adds repeated lateral movement across her body space (G6). The final one again maintains the hand shape, but changes orientation

[31] As discussed in Chapter 4, this reorientation to engage with imagined characters is a well-documented narrative strategy in both signed and spoken languages (e.g. Janzen 2004, 2012; Parrill & Sweetser 2004; Stec 2012; Sweetser 2007; Sweetser & Sizemore 2008).

as she points to herself (G7).

(89) TRANSCRIPT 25: DREW BARRYMORE
[UID:658edac8-1ee1-11e9-b705-089e01ba0335,2071]

```
1  DB   *so that's part of who I am on the inside* *and then I got
        *G1- - - - - - - - - - - - - - - - - - - - * *G2- - - - - ->
2       this on the outside* *and that brings a whole other thing.
        <- - - - - - - - - * *G3- - - - - - - - - - - - - - - - - ->
3       But sometimes* *(silly noises)* *Anyway* *what I'm saying is*
        <- - - - - - * *G4- - - - - -* *G5 - -* *G6- - - - - - - -*
3       *I'm stuck here*
        *G7 - - - - - -*
G1:  both hands, loose, palm self, touching chest
G2:  right hand, index-finger point toward face, circular motion
     around face
G3:  full two-hand shrug, performed twice
G4:  both hands, index finger point toward head, repeated cyclic
     motion, ("crazy" emblem)
G5:  right hand, index finger pointed up, held toward audience
G6:  right hand, index finger pointed up, 3x beats while leaning
     right
G7:  right hand, index finger point toward self
```

Figure 5.5: Gesture sequence corresponding to TRANSCRIPT 25: DREW BARRYMORE, ex. 89.

The main reason for including this sequence is to highlight how gestural context can complicate the classification and interpretation of single gestures. G5 has all of the physical characteristics of a pausing gesture and is performed in a discourse context (i.e. a topic-shift) in which we expect to see pausing gestures. However, the gesture occurs amid a string of other gestures sharing the same hand shape. It is thus unclear how intentionally the pausing gesture is produced, or whether it is the result of co-articular with temporally adjacent gestures.

5.5.3 Non-manual disengagement

Non-manual disengagement is performed by reorienting ones gaze, head, or body away from the primary Interaction Space and the speaker-hearer. This makes non-manual disengagement an incredibly heterogeneous class of gestural behavior. Everything from leaning back, to looking down, to turning one's head to the side counts as non-manual disengagement. What all these forms have in common is a very basic action schema of distancing oneself from unwanted objects and activities. Because the Interaction Space is the metaphoric space in which participants manage discourse structure, distancing oneself from the Interaction Space signals that something in it is undesirable. In the case of topic-shifting, the undesirable 'thing' is the dismissed topic, present within the Interaction Space as a metaphoric object.

Non-manual disengagement occurs in the majority of the present data set (65/106). 22 of these occur in tandem with a manual dismissal strategy (i.e. with a *removal* or *stopping* gesture). We have already seen instances of this above, such as when Timothée Chalamet looks down while performing a *flicking* gesture and saying "anyway" (ex. 83). In the remaining 43 examples, the non-manual disengagement occurs as the only gestural marker of dismissal. The following example demonstrates the use of non-manual disengagement, independent of manual dismissal strategies, during an audience-triggered return shift.[32]

In (90), American comedian Kathy Griffin is discussing how she dealt with the fall out of a scandal by promoting her own shows. As Griffin reveals her biggest success, selling out the large New York venue Carnegie Hall, the audience cheers, and she turns to clap along with them. She then returns her hands to rest on her lap and looks back at Colbert as she laughs (line 4, P1). Colbert, after this moment of mutual gaze, takes the floor to make a comment about never selling out Carnegie Hall (line 5). In response, Griffin brags that she has sold it out five times (line 6), and then reaches toward Colbert as if to comfort him (G1). Simultaneously, she dramatically leans her head away (P3), looking up and saying "anyway" as she does so. Griffin then returns to her story (lines 6-7). The return to the main discourse is signalled in the gestural mode by returning her gaze to Colbert (P5) and performing a two-handed presentational gesture into the primary Interaction Space (G4). The lexical discourse marker phrase "so then" signals the continuation of a narrative (line 6).

[32]This example also appears in the discussion of *engagement management* in Chapter 4.

(90) TRANSCRIPT 26: KATHY GRIFFIN
[UID:c5b60a9c-fabb-11e8-ba14-089e01ba0335,1868]

```
1   KG    so um I had everyone in Hollywood saying you can't sell any
2         tickets, and I sold out Carnegie Hall in less than 24 hours
3   --    (applause)
4   KG    Thank you thank you △ (laughs)
                               △P1- - - ->
5   SC    I never sold out Carnegie hall
6   KG    I've sold it out △five times.  △*Anyway um* △ *so so then
          (off screen)     △P2- - - - - -△P3- - - - - △P4 - - - ->
                                          *G1- - - -*   *G2 - - - ->
7         I* decided to promote my own shows
          <- - - - - - - - - - - - - - - -≫
          <*
P1, P4:  head and gaze toward Colbert
P2:  head facing audience, eyes closed, lean left toward Colbert
P3:  head facing audience, gaze up, lean right away from Colbert
G1:  left hand reach toward Colbert, palm flat on desk
G2:  both hands, open palm up toward Colbert
```

Figure 5.6: Gesture sequence corresponding to TRANSCRIPT 26: KATHY GRIFFIN, ex. 90.

This example demonstrates the importance of the principle of *independence of contribution*. The gestural performance notated as P3;G1 conveys social and discourse structural meaning simultaneously and independently. The lexical discourse marker and reorientation of Griffin's head and gaze co-express topic-dismissal. Her hand gesture, on the other hand, conveys social information, performing a kind of apology. To maintain the politeness principle of *friendliness*, Griffin must make up for her bragging. She could do this through a verbal apology or compliment. Instead, she reaches her hand toward Colbert in an expression of intimacy.[33] Griffin's articulators then coalesce upon her re-engagement with the main

[33]This interpretation is grounded in everyday physical experiences of comfort. To comfort a grieving friend, we may reach out across the table to take their hand, or place our hand on their shoulder. This primary scene of physical comfort is also reflected in the primary metaphors INTIMACY IS PHYSICAL PROXIMITY and MENTAL STATES ARE PHYSICAL STATES (Dancygier & Sweetser 2014; Lakoff & Johnson 1980).

discourse, as her return is simultaneously signalled by the orientation of her body, gaze, and head, the trajectory of her hand gesture, and the lexical marker "so then".

In Section 5.3, I made predictions regarding the influence of various discourse-structural factors on non-manual disengagement. I predicted that disengagement would be least frequent with return shifts because digressive topic are less likely to occupy a position within the Interaction Space. I also predicted that disengagement would be most frequent in audience-triggered shifts because interacting with the audience requires the interview guest to reorient their body away from Colbert. Finally, I predicted that disengagement would be frequent when ceding or passing a turn, since disengagement can independently serve as a marker of turn-transition.[34] Conversely, disengagement was predicted to be *infrequent* when taking and keeping a turn because these contexts independently require engagement with the Interaction Space.

The three tables below provide the distributions of non-manual disengagement gestures in relation to shift type, shift-trigger, and turn-position. None of the predictions are clearly supported by the data. It is possible that this is a result of insufficient data. For example, there are only 10 audience-triggered shifts in the entire data set. More likely, however, is that non-manual disengagement is too heterogeneous of a class to make coherent predictions. Sufficiently exploring the diversity of non-manual disengagement strategies is outside the scope of the present work. The reader is directed to work on *proxemics* for relevant discussion (e.g. Gill et al. 2000; Hagemann 2014; Hall 1995).

	Full shift	Partial shift	Return	Total
Interview	6/10 (60%)	7/12 (58%)	19/33 (58%)	32/55 (58%)
Monologue	2/16 (13%)	3/4 (75%)	6/31 (19%)	11/51 (22%)
Total	8/26 (31%)	10/16 (63%)	25/64 (39%)	43/106 (41%)

Table 5.15: Distribution of non-manual disengagement by shift type

[34]See Chapter 4, Section 4 for discussion.

	Speaker-triggered	Addressee-triggered	Audience-triggered	Total
Interview	22/34 (65%)	7/15 (47%)	3/6 (50%)	32/55 (58%)
Monologue	11/47 (23%)	- -	0/4 (0%)	11/51 (22%)
Total	23/81 (28%)	7/15 (47%)	3/10 (30%)	43/106 (41%)

Table 5.16: Distribution of non-manual disengagement by shift trigger

	Keep	Cede	Take	Pass	Total
Interview	17/27 (63%)	4/8 (50%)	10/18 (56%)	1/2 (50%)	32/55 (58%)
Monologue	11/47 (23%)	- -	0/4 (0%)	- -	11/51 (22%)
Total	28/74 (38%)	4/8 (50%)	10/22 (45%)	1/2 (50%)	43/106 (41%)

Table 5.17: Distribution of non-manual disengagement by turn position

5.5.4 Interim summary of topic-dismissal

The findings reported in this section support the proposal that *removal* gestures serve as gestural topic-dismissal markers. This class of gesture occurs in approximately one quarter of the data set, and largely aligns with predictions made in Sections 5.1 and 5.3. Furthermore, there are contextual explanations for examples that appear to violate predictions. For example, removal gestures unexpectedly occurred four times in turn-take positions. Three of these occurred with *bufferred takes* in which the speaker briefly acknowledges the topic immediately before dismissal. This suggests that *buffered takes* perform a particular interactive function – allowing the speaker to take control of the topic in order to remove it from the discourse. We also saw one example (Section , ex.) in which a removal gesture is performed *without* an accompanying lexical discourse marker. As with the removal gestures accompanying "anyway", this example occurred at during a topic-shift. This supports the principles of *optionality of expression* and *independence of contribution* which state that interactive gestures should be able to perform discourse management with and without accompanying lexical markers.

The reported findings do not support the use of *stopping* gestures as gestural-topic dismissal. These gestures in only 10% of the data set, and are complicated by factors unrelated to the topic-shift itself. In particular, over half of *blocking* gestures in the data set occur while Colbert is impersonating Trump. Colbert recurrently uses the *blocking* shape and orientation as a cue for these impersonations, suggesting that their use with "anyway" may be unrelated to the expression of topic-dismissal.

Finally, non-manual disengagement was the most common gestural strategy associated with topic-dismissal in the data set, occurring in 65 of the 106 examples. However, the heterogeneity of this class of gesture prevents us from drawing clear conclusions without further study and sub-classification.

5.6 Topic Introduction

The second primary function of the lexical discourse marker "anyway" is to signal the introduction of a new topic through the opening of a new discourse question. In the data collected, the gestural expression of this function falls into four recurrent gestural classes which I am calling here *presentation*, *locating*, *addressing*, and *re-engagement*. Note that *locating* and *addressing* gestures are sub-types of the REFER action schema. These names reflect the action schema associated with the gestures' formal features (Chapter 2), as well as the Interaction Space actions associated with them (Chapter 4). Presentation gestures *present* metaphoric or literal objects into the Interaction Space for inspection, and enable further interaction with them (e.g. Bavelas et al. 1992; Müller 2004). Locating gestures *locate* a literal or metaphoric object in order to direct joint attention toward it (e.g. McNeill et al. 1993; McNeill 2003). Re-engagement gestures signal renewed participation with the primary Interaction Space, and thus the discourse, by turning or leaning toward the center of the Interaction Space (e.g. Gill et al. 2000; Hall 1995; Kendon 2010; Streeck & Hartge 1992). Examples of these four gestural classes are given in Figure 5.7.

Figure 5.7: Classes of topic-introduction gestures aligned with "anyway"

The leftmost gesture in Figure 5.7 shows American actor and playwright Lin-Manuel Miranda performing a palm-up open hand gesture (e.g. Cooperrider et al. 2018; Müller 2004) into the central Interaction Space, as if to present an object for inspection. This

is a sub-type of *presentational* gesture that is discussed in Section 5.6.1. The center-left gesture shows British actor Emma Thompson performing a *locating* gesture in the central Interaction Space, pointing with her index finger to a location on the desk between her and Colbert. This example is discussed in Section 5.6.2. The center-right gesture shows Colbert performing an *addressing* gesture, pointing his index finger directly toward the audience as if to remind them to take some action. The rightmost gesture shows American comedian Conan O'Brien performing a non-manual re-engagement gesture, re-directing his head and gaze toward Colbert after looking down. This strategy serves to re-assert O'Brien's participation in the discourse by overtly signaling where his attention is directed. This example is discussed in Section 5.6.4.

The introduction of a new topic is signalled using at least one of these gestural strategies in 81% of the data (86/106). Non-manual re-engagement co-occurs with the vast majority of manual topic-introduction gestures (38/45). This is summarized in Table 5.18.

Gesture class	Interview	Monologue	Total
Presentation			
Presentation only	2	0	2
Presentation+re-engagement	17	10	27
Total	**19/55**	**10/51**	**29/106**
	(35%)	(20%)	(27%)
Locating			
Locating only	2	1	3
Location+re-engagement	4	2	6
Total	**6/55**	**3/51**	**9/106**
	(11%)	(6%)	(8%)
Addressing			
Addressing only	0	2	2
Addressing+re-engagement	2	3	5
Total	**2/55**	**5/51**	**7/106**
	(4%)	(10%)	(7%)
Re-engagement			
Re-engagement only	**15/55**	**26/51**	**41/106**
	(27%)	(51%)	(39%)
Total	**41/55**	**44/51**	**86/106**
	(75%)	(86%)	(81%)

Table 5.18: Gestural expression of topic introduction with *anyway*

There are three things worth noting about the general distribution of *presentation* gestures in the data set. First, presentation gestures are the most common manual strategy for topic-introduction. This makes sense given that a topic is introduced in all topic-shift contexts, regardless of shift-type or trigger-type. *Locating* gestures and *addressing* gestures, as sub-types of the REFER action schema, are comparatively infrequent, occurring in only 15% of the data combined. Locating gestures are only expected in partial and return shifts in which an ongoing discourse topic is maintained. This is because there is an ongoing topic in the Interaction Space as a metaphoric object at the time of the gesture. This prediction holds with only one exception in which Colbert uses a pointing gesture during a full-shift in a Trump impersonation. Addressing gestures are not directly related to topic-introduction itself, but rather to eliciting a contribution from the addressee. As such, they are expected to occur only in cede and pass turn positions where the addressee – the two contexts in which the addressee will make the first contribution to the new topic. This prediction does not hold for the data set, largely because the majority of *addressing* gestures occur in monologues in which Colbert never cedes or passes his turn.

Second, though there are very few examples of locating and addressing gestures, it is interesting that their distribution in interviews and monologues are opposite one another – *locating* gesture occur more in interviews, and *addressing* gestures occur more in monologues in the same proportion. The relatively high frequency of locating gestures in interviews as opposed to monologues is likely a result of spatial affordances – there is a literal surface (the desk) on which to organize metaphoric objects. This is not the case in monologues, especially the opening monologue in which Colbert is standing and constantly moving around the stage.

The relatively high frequency of addressing gestures in monologues over interviews is likely related to the social affordances of the two settings. In all but one of the 55 examples of "anyway" in interviews, the guest is speaking. We expect addressing gestures when the gesturer is eliciting a response from their addressee. In an interview setting, Colbert, as interviewer, is responsible for eliciting responses from his interviewees. Interviewees, on the other hand, are not expected to request answers from Colbert. As a result, we expect addressing gestures to not occur in interviews in this particular data set.

Lastly, non-manual re-engagement gestures appear much more frequently *without* manual gestures in monologues than in interviews. I suggest that this is related to the physical affordances of the two settings. In monologues, especially the opening monologue, Colbert has access to more non-manual strategies. He is constantly moving around, turning from side-to-side, stepping toward and away from the camera. Re-engagement actions in interviews are limited to relatively subtle reorientation strategies.

In the remainder of this section I discuss examples of each of the four classes of topic-introduction gestures and how they relate to the predictions made in Section 5.3.

5.6.1 Presentation

Presentation gestures exhibit the physical affordances of presenting an object for inspection. In the current data set, I observed three sub-types of presentation gestures co-occuring with

"anyway" – *containment, palm-up open-hand* (henceforth PUOH), and a complex gesture composed of the PUOH hand shape and orientation and cyclic movement. An example of each is depicted in Figure 5.8.

As discussed in Chapter 2, these different forms of presentation gestures have different affordances as to the properties of the metaphoric object they introduce. Containment gestures evoke the presentation of a flat-sided rectangular object, such as a box which can contain other objects. This is conducive to the introduction of high-level topics that are composed of sub-questions that can be individually pursued. The PUOH gesture is relatively underspecified and homophonous. The gesture does not convey information about the properties of the metaphoric object other than that it can be supported by an open upturned hand. The physical properties of the PUOH gesture are also known to be consistent with *opposite* actions – the gesture can evoke both a PRESENTATION and a REQUEST action schema (e.g. Cooperrider et al. 2018). The final sub-type, *cyclic PUOH*, only occurs with "anyway" twice in the present data set. Both examples occur immediately after an interruption. Ladewig (2011) notes that this complex gesture combines the PRESENTATION action schema with the CYCLE image schema to represent an iterative presentation of information.

Figure 5.8: Sub-types of presentation gestures

In (91), American actor Andrea Savage performs a containment gesture (Fig. 5.8, G1) as she returns to her story after a lengthy derailment. This discourse excerpt was discussed in detail in Section 5.3.4, ex. 73b, and will not be fully reviewed here. It is enough to say that (i) the topic-shift is signalled by "anyways" (line 3), (ii) the re-introduction of the discourse topic is signalled by the containment gesture into the central Interaction Space (G1), and (iii) the continuation of a previous narrative is repeatedly signalled by "so" (lines 2-3).

(91) TRANSCRIPT 27: ANDREA SAVAGE
[UID:67f2c08a-f628-11e9-920c-089e01ba0335,3492]

```
1  AS   Thrilled to be a part of anything that's the last
2  SC   Okay, so
3  AS   *So anyways, so* I- so we went back and forth
        *G1- - - - - - *
```

G1: two hand containment gesture held toward Colbert

There are two important points to consider in regards to the use of a containment gesture in (91). First, this topic-shift is an example of an addressee-triggered return in which the speaker takes the turn. Presentation gestures are explicitly predicted for such a context – new information is being presented, and it is the first contribution of the speaker for the given turn. Second, this particular gesture occurs after a long and very disruptive derailment. This renders the still-open question of the main discourse relatively in accessible – a lot has happened in the Interaction Space since the last time it was addressed. This context favors a presentation gesture over a locating gesture, despite the prediction that locating gestures should occur more frequently in return shifts.

In (92), American actor and playwright Lin-Manuel Miranda re-engages with the primary Interaction Space between him and Colbert after engaging in an imaginary conversation with his Twitter followers. In the discourse excerpt, Miranda is describing a particular habit he has to say "goodnight" on Twitter and is complaining that his followers interpret this too literally (lines 1-2). Beginning in line 3, Miranda engages with his imagined Twitter followers rather than Colbert, as signalled by the use of the second person pronoun "you". He then signals his return to the main discourse with "anyway" as he turns back to Colbert, and performs a PUOH presentation gesture into the central Interaction Space (Fig. 5.8, G2).

(92) Transcript 28: Lin-Manuel Miranda
[UID:49c283d6-eef2-11e8-bf71-089e01ba0770,2113]

1 LM And so sometimes I'll say goodnight, and I'll go and have
2 dinner, and I'll have time with my family. But everyone
3 thinks that I'm literally going to bed. I'm not. I'm just
4 saying goodnight to you to be polite because I like you and
5 I like the time we've shared on Twitter. Um. *Anyway, in
*G1- - - >
6 the- in the-* y'know over the years, people have sort of
<- - - - - *
7 come to rely on it
G2: right hand, flat, palm up, held toward Colbert

The shift-type of this example is difficult to determine. On the one hand, it seems like a *return* in that the imagined interaction between Miranda and his followers can be interpreted as a completely distinct interaction between a different set of participants. This interpretation is supported by Miranda's pattern of engagement – he is oriented toward Colbert, reorients his head, gaze, and body during the entirety of the imagined interaction, and then physically *returns* to Colbert. On the other hand, the information conveyed during this imagined interaction is directly related to the question under discussion, namely his Twitter habits. Given this, the move more closely resembles a *partial shift* – Miranda was pursuing one line of inquiry in which he specified and corrected his Twitter followers' misunderstanding; he then moves onto a different, more general sub-question regarding how long he has engaged in

the habit and what the repercussions are. I do not think that there is a conclusive argument for one analysis of the other. However, it is interesting to note that Miranda's gestural behavior, especially his pattern of reorientation, rather clearly supports the first analysis.

In (93), Irish actor Saoirse Ronan performs a cyclic PUOH gesture as she attempts to recover from a derailment caused by Colbert. This example was discussed in Section 5.3.3 and will be returned to once more in Section 5.7.3. For this reason, I will not review the discourse context fully here. What is important for the purposes of the presentation gesture is that Ronan was derailed by Colbert immediately before and appears repeatedly flustered, laughing and mumbling as she tries to regain the turn and continue her story (lines 4 & 6-7). Ronan signals the attempted topic-shift with "anyway" (line 7), and reasserts the current open question with the frame-setting phrase "7 months later" which places us in the center of the story being told. The cyclic PUOH gesture spans both of these phrases.

(93) Transcript 29: Saoirse Ronan
[UID:92fedfb8-1b1f-11ea-9518-089e01ba0335,2060]

```
1  SC   There are ways to tell that a woman is pregnant, by the way
2  SR   Yeah
3  SC   I don't know if you've had the talk but...
4  SR   I know- (mumbles) there- there were babies made.  (laughs)
5  SC   uh-huh, a firm handshake
6  SR   Right (laughs). She, um, so she, y'know, must've done that
7       And she um she- I don't know (laughs) *Anyway, 7 months
                                             *G1- - - - - - ->
8       later-*
        <- - -*
9  SC   We don't know, we don't know
10 SR   We don't know what happened
G3:  both hands, loose palm up, fast cyclic motion
```

Given the verbal, social, and gestural context of interpretation, this cyclic PUOH gesture appears to be semiotically rich, contributing meaning at multiple levels. First, the cyclic movement expresses semantic meaning about the time frame of the pregnancy by evoking the primary metaphor TIME IS MOTION THROUGH SPACE (Ladewig 2011, 2014a). This parallels the verbal phrase "several months later". The cyclic movement also expresses interactive meaning related to Ronan's attempts to recover from the derailment and return to her story, as if to propel herself forward after being stopped by Colbert. This evokes the metaphor MIND IS MACHINE (*ibid*). Together with Ronan's averted upward gaze, associated with a "thinking face"(Goodwin & Goodwin 1986), the cyclic movement serves as a cue that she intends to keep the floor through initial disfluencies. Finally, the PUOH shape and orientation contributes to the interactive meaning by signalling continued contributions of metaphoric objects into the Interaction Space.

The distribution of presentation gestures in the data set partially support the predictions made in Section 5.3. As predicted, all presentation gestures in interviews occurred in a keep (11/19) or take (8/19) turn position.[35] This was predicted because it is in these contexts that the speaker makes the first contribution to the introduced topic. When "anyway" occurs in cede and pass turn position, it is the addressee who makes the first contribution to the new topic.

It was also predicted that presentation gestures would occur at the same rate across shift types. This is not the case, as demonstrated in Table 5.19. Presentation gestures occur most frequently with return shifts, more noticeably so in interviews that in monologues. Only one presentation gesture occurred in a full shift. This may be related to the fact that full shifts (in interviews) occurred almost exclusively in cede or pass turn positions (8/10).

	Full shift	Partial shift	Return	Total
Interview				
Containment	-	2	4	6
PUOH	-	1	10	11
Cyclic PUOH	-	-	2	2
	0/10	**3/12**	**16/33**	**19/55**
	(0%)	(25%)	(48%)	(35%)
Monologue				
Containment			3	3
PUOH	1	1	5	7
Cyclic PUOH	-	-	-	-
	1/16	**1/4**	**9/31**	**10/51**
	(6%)	(25%)	(29%)	(20%)
Total	**1/26**	**4/16**	**24/64**	**29/106**
	(4%)	(25%)	(38%)	(25%)

Table 5.19: Distribution presentation gestures by shift type

It may also be worth noting that 8 of the 11 examples of "so anyway"/"anyway so" in interviews co-occur with a presentation gesture. Given that this is a significantly higher rate than the occurrence of presentation gestures with topic-shifts more generally, this suggests that the function of "so" and presentation gestures may be more closely related than the function of "anyway" and presentation gestures. The co-occurrence of presentation gestures with "so" in other use contexts is worth further study.

[35]This also true for monologues. However, because all but four monologue examples occur in a keep position, the finding is not particularly meaningful.

5.6.2 Locating

Locating gestures are deictic gestures that lack the physical affordances of presentation. In their use as interactive gestures, they serve to direct joint attention toward a particular discourse topic by indicating the topic's location as a metaphoric object in shared space (e.g. McNeill et al. 1993; McNeill 2003).

In (94), British actor Emma Thompson performs two locating gestures into the central Interaction Space as she returns to a narrative after a digression. We saw this discourse excerpt previously as an example of a topic-shift with a 'mixed-trigger' (Section 5.3.3, ex. 71). In this excerpt, Colbert and Thompson are discussing a clip from Thompson's recent film *Late Night* in which she plays a late night talk show. Thompson digresses from the description of the clip to make a joke about the lack of women in late night (lines 1-2). After commenting on her own joke (lines 4-5), Thompson moves to return to the main discourse. This return is signalled in the verbal mode by "anyway" and in the gestural mode by *locating* gesture (G1). Colbert disrupts Thompson's attempted return by also commenting on her joke (line 6). Thompson then takes the turn, acknowledging Colbert's contribution before attempting a return again. Having kept her finger on the desk through Colbert's interruption, she taps the desk at "anyway" and successfully returns to the main discourse. Immediately after this second "anyway" Thompson performs a second locating gesture (G2), indicating the same region of space but with an open hand rather than with an index-finger point. She then performs a two-handed presentation gesture into the central Interaction Space as she provides new information about the clip (lines 8-9, G3).

(94) Transcript 30: Emma Thompson
[UID:227f3a58-c570-11e9-a890-089e01ba0335,2100]

```
1  ET   She's a woman late night talk show host, so it's basically
2       science fiction.  And um (.)
3  --   (cheers)
4  ET   Did you see the way I slipped that in?  (.)  Sneaky sneaky
5       little political remark there from Dame Thompo.  Anyway *um
                                                          *G1>
6  SC   We don't need politics in late night
7  ET   Oh nooo, we really don't, good grief.  /Anyway* *so she's
       < - - - - - - - - - - - - - - - - - - - /- - - -* *G2- - ->
8       come into the writer's room* *to talk them and tell them,
       < - - - - - - - - - - - - -* *G3- - - - - - - - - - - - - >
9       give them a row* and say "you have got to make this better"
       < - - - - - - -*
G1:  right hand, index-finger point on desk, palm down, beat at
     "anyway", line 7
G2:  right hand, flat, palm down, held over desk
G3:  both hands, loose containment gesture toward Colbert
```

Figure 5.9: Gesture sequence corresponding to TRANSCRIPT 30: EMMA THOMPSON, ex. 94.

This sequence of two locating gestures followed by a presentation gesture expresses a complex reintroduction of the main discourse topic. The two locating gestures are performed while marking the initial return (G1) and summarizing previously provided information (G2). The topic being returned to, and the particular information about the topic that has already been described is already present in the Interaction Space, and so can be referred to with a locating gesture. Note that as in the Chalamet example (Section 5.5.1, ex. 83), Thompson uses a pronoun immediately after the return ("she", line 7). This again supports the maintained accessibility of the topic as a metaphoric object within shared space. Thompson then shifts to a presentation gesture as she presents the first new information since returning to the main discourse.

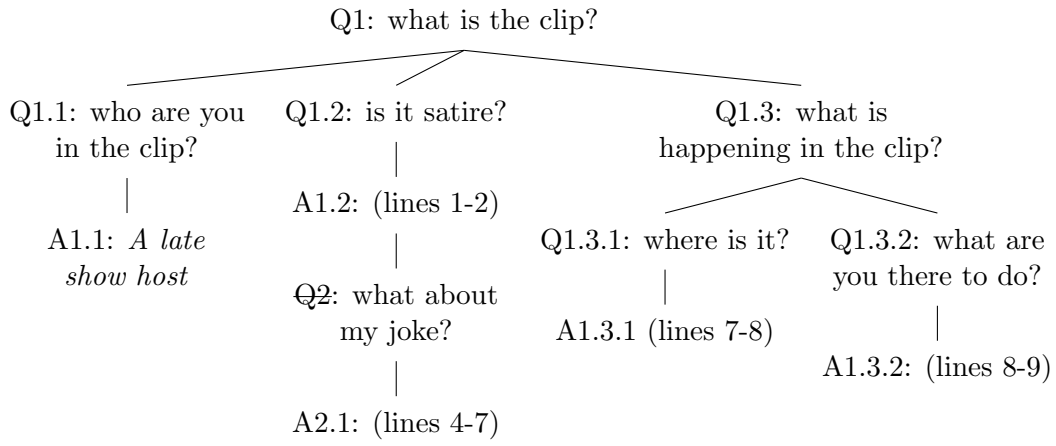
To better understand the embedding of discourse management actions performed by Thompson, consider the Interaction Space state representation in (95). G1 performs two LOCATE actions, once when initially performed (line 5), and once when re-articulated (line 7). These actions serve to draw joint attention to the ongoing topic (i.e. the movie clip) after each digression, as if to say "this is what we are talking about". G2 performs another LOCATE action into the same space while summarizing old information about the setting of the clip. The shared placement of these two gestures results in an embedded structure, as if to say "this is what we are talking about, and this is where we are at in talking about it". The assertion joint attention is also reinforced by her gaze, which is directed at her gesture (e.g. Gullberg & Holmqvist 2006). Having drawn mutual attention to the ongoing topic and relevant open question, Thompson then PRESENTS new information with G3. Because of the immediately preceding LOCATE actions, we again know to embed this new information under the established topic.

(95) **Management actions by Thompson, Lines 5-8, ex. 94**

$$IS1 = \begin{cases} \text{Participants: } \text{ET}_\text{S}, \text{SC}_\text{A} \\ \text{Content : Clip(Setting, +Action)} \\ \text{Management :} \\ \quad \textsc{Locate}(\text{G1}_1 \rightarrow \text{Clip}) \\ \quad \textsc{Locate}(\text{G1}_2 \rightarrow \text{Clip}) \\ \quad \textsc{Locate}(\text{G2} \rightarrow \text{Setting}) \\ \quad \textsc{Present}(\text{G3} \rightarrow \text{Action}) \end{cases}$$

This embedded structure is reflected in the discourse structure provided in (96). "Anyway" serves to close the digression (Q2), the fist locating gesture signals where in the structure we are returning to (Q1), the second locating gesture signals the sub-question we are currently addressing (Q1.3), and the presentation gesture signals a new partial answer to this sub-question (A1.3.2).

(96)



Though there are relatively few locating gestures in the present data set, the 8 that do occur conform to the predictions made in Section 5.3. As demonstrated in Table 5.20, no locating gestures occur with full shifts. This was predicted to be the case because full shifts introduce an entirely new topic to the discourse. As such, there should be no relevant metaphoric object present in the Interaction Space to locate.

	Full shift	Partial shift	Return	Total
Interview	0/10 (0%)	1/12 (8%)	5/33 (15%)	6/55 (11%)
Monologue	0/16 (0%)	0/4 (0%)	2/31 (6%)	2/51 (4%)
Total	0/26 (0%)	1/16 (6%)	7/64 (11%)	8/106 (8%)

Table 5.20: Rate of locating gestures by shift type

Locating gestures were also predicted to occur only in keep and take turn positions, contexts in which the gesturer makes the first contribution to the introduced topic. There is only one exception to this prediction, which proves unexceptional in the context in which it is performed. In the relevant clip, American actor Christine Baranski is providing some background information for a clip that Colbert is about to display. After describing one of the actors for some time, she begins to falter and performs a partial shift marked by

"anyway" to reorient the discourse toward the higher level discourse topic of the clip itself. After performing a locating gesture toward the screen on which the clip will appear, she cedes her turn and the clip is played.

5.6.3 Addressing

Addressing gestures are gestures that deictically refer to the addressee and do not have the physical affordances of presentation. These gestures are well-documented in the literature on interactive gesture, and are generally associated with eliciting backchannels and facilitating turn-transitions (e.g. Bavelas et al. 1992; Bavelas 1994; Mondada 2007; Trujillo & Holler 2021). As such, they are not predicted to signal topic-shifting specifically, but may nonetheless appear in topic-shifting contexts to independently perform one of the above mentioned functions.

In the present data set, addressing gestures occur with "anyway" in only 7 examples. 6 of these occur in monologues, and specifically take the form of an index-finger point. I suggest that this is indicative of an idiosyncratic construction associated with Colbert's monologues. An example of this performance is provided in (5.6.3). The one example during an interview is somewhat ambiguous – a small flat handed flick toward Colbert that could be interpreted as either an addressing gesture or an abbreviated locating gesture.[36]

In (97), Colbert is critiquing the ways in which right-wing media was manufacturing an immigration crisis by over-reporting on 'caravans' of migrants reaching the US border. He briefly digresses to make a joke about how "horrible" the show *Fox and Friends* is (lines 2-4). He then pauses, bringing his hands to rest and looking down in faux despair (line 4, G1). He then returns to his critique, marking the return with "anyway" and pointing toward the audience as he does so (line 4, G2). He then raises his finger upward slightly as he gets to the important details of his critique (line 4, G3).

(97) Transcript 31: Stephen Colbert
[UID:224513fa-e4bb-11e8-9135-089e01ba0770,380]

```
1  SC   In fact, I counted how many times they talked about the
2       caravan on Fox and Friends- uh well technically had my
3       footage department do it.  Uh- I can't watch Fox and Friends
4       cause my doctor says it's just horrible *(.)*
                                                *G1-*
5  --   (laughter)
6  SC   *Anyway*, *this is true, in the six days* *leading up to the
        *G2- -*  *G3 - - - - - - - - - - - - - - * *G4- - - - - - ->
7       election, they used the word* "caravan" an average of
        < - - - - - - - - - - - - - -*
8       twenty-one times per episode
```

[36]Example retrievable at UID:41c28826-101f-11ea-a09a-089e01ba0770,2942

```
G1:  hands to rest at side, gaze down
G2:  left hand, index-finger point toward audience, palm center
G3:  left hand, index-finger point toward audience, palm down
G4:  left hand, flat, palm center, large rising sweep up and left
```

Figure 5.10: Gesture sequence corresponding to TRANSCRIPT 31: STEPHEN COLBERT, ex. 97.

The hand shape and ambiguous orientation of G3 resembles an under-articulated *pausing* gesture. Similar gesture sequences of addressing-to-pausing gestures appear with specification moves in Chapter 7, particularly during monologues. Based on this similarity, I suggest that the addressing gesture in this example relates to a *specification* move rather than the topic-shift marked by "anyway". In return and partial shifts, these two moves are not mutually exclusive. The topic-shift moves to cancel a potential sub-question, close it's immediately dominate question, and pose a question related to a higher level of the discourse's structure. If the new question is a sub-question under an established discourse topic, then a specification move is also achieved. This is the case in the Emma Thompson example discussed above. In the Colbert example in (97), the sub-question that is returned to is *how many times was "caravan" mentioned on Fox and Friends?*, which was initially opened in line 1, immediately before the digression. If I am correct in this analysis, then the lexical discourse marker "anyway" and its accompanying addressing gesture are performing complementary discourse management functions.

5.6.4 Non-manual re-engagement

As stated at the beginning of his section, non-manual engagement strategies often accompany manual topic-introduction gestures. For example, when Lin-Manuel Miranda performs a presentation gesture into the central Interaction Space, he also signals re-engagement with the space by reorienting his body, head, and gaze toward Colbert (5.6.1, ex. 92). Similarly, Colbert re-engages with the audience as he performs the addressing gesture discussed in the previous section, reorienting his gaze to look directly at the camera.

Non-manual gestural strategies are also used unaccompanied by manual gestures in 41 of the 85 examples in which topic-introduction is gesturally expressed. One such example is provided in (98). We saw this discourse excerpt once before as an example of "anyway" in a 'pass' turn-position (Section 5.3.4, ex. 74).

In this excerpt, American Comedian Conan O'Brien is describing the shorter format of his new show *Conan*. He is humorously self-deprecating as he implies that his audience really doesn't like the show, and they could cut it even shorter (lines 1-2). As O'Brien looks down in dramatized self-pity (P2), Colbert tries to push him for more details (line 3). O'Brien declines, saying "anyway" and perking up slightly, shifting his head and gaze toward Colbert (line 4, P3). O'Brien passes his turn, leaving Colbert to open the new discourse topic. Instead of beginning an entirely new topic, Colbert completes the shift as a partial shift, addressing the brevity of O'Brien's new show from a different angle. At this point, O'Brien completes his re-engagement with the primary Interaction Space, fully reorienting his head and gaze toward Colbert.

(98) TRANSCRIPT 32: CONAN O'BRIEN
[UID:31292f26-3a62-11e9-928f-089e01ba0335,1809]

```
1  COB  I said, △let's do half an hour because we noticed that our
                △P1- - - - - - - - - - - - - - - - - - - - - - ->
2       crowd, after a half an hour, they were like "we're good"△
        <- - - - - - - - - - - - - - - - - - - - - - - - - - - △P2>
3  SC   Really?
4  COB    (.)  △ anyway △
          <- - -△P3- - - △P4-≫
5  SC   I think [in show business]
6  COB           [I- I got sad-]
5  SC   we call that "leave them wanting some"
7  COB  yeah (.)  no, even at half an hour, they're like you know
8       you could go to fifteen so (.)
9  SC   Now there's a- tonight, obviously, you're dressed beautifully
```

Figure 5.11: Gesture sequence corresponding to TRANSCRIPT 32: CONAN O'BRIEN, ex. 98.

This is exactly the type of context in which we expect to see non-manual re-engagement without accompanying manual gestures. O'Brien has initiated a topic-shift, but does not wish to make the first contribution to the new topic. This would make the use of a presentation gesture or locating gesture infelicitous. Instead, O'Brien signals his cooperation and continued engagement with the ongoing discourse non-manually.

As I discussed with non-manual *dis*engagement strategies, the class of non-manual engagement strategies is very heterogeneous in terms of both form and function. Re-establishing the speaker-hearer line through head and gaze orientation constitute re-engagement with the Interaction Space. So does leaning in or stepping toward the addressee. Without a more fine-grained classification of non-manual engagement behavior, it is difficult to meaningfully compare the observed behaviors with the general predictions made in Section 5.3.

5.7 Complex sequences

In this final section, I look at three full topic-shift sequences. The first considers the function of *shrugs* as a signal of transition between topics and an unwillingness on behalf of the gesturer to begin the new topic. The second and third look at the diversity of gestural strategies used during lengthy derailments. These three examples highlight the importance of considering full contexts when interpreting the contribution interactive gestures make to successful topic-shifts.

5.7.1 Shrugging away a topic

In (99), American actor Renée Zellweger reenacts the time that she met former president Jimmy Carter. It is a rather embarrassing story – she gets nervous and doesn't know what to say to the former president after waiting hours to see him for a book signing. After pantomiming taking the signed book back from an imagined Jimmy Carter (G1), Zellweger signals the end of her reenactment by silently placing the imaginary book onto the desk between her and Colbert (G2). Upon finishing her reenactment, Colbert and the audience clap. She expresses gratitude and, seemingly, embarrassment as she awkwardly looks down, performs two consecutive shrugs (G3 & G4), and readjusts her posture. Colbert then moves to end the interview, at which point Zellweger re-engages with the primary Interaction Space by looking directly at Colbert and shaking his hand.

(99) Transcript 33: Renee Zellweger
[UID:c45441ee-e027-11e9-ad79-089e01ba0335,2601]

```
1  RZ   And I was really ready, cause here it comes, I'm going to
2       have my thing that I say to Jimmy Carter, and he looked up
3       at me and he said "hello", and um and I said "um (.)  hey"
4       (.)  *(whispers) thank you* *(.)*
                 *G1- - - - - - - - - * *G2-*
5  SC   (claps)
6  RZ   *Anyway* thank you very much *yes, thank you* thank you
        *G3- - *                    *G4- - - - - - *
7  SC   *Renee, thank you so much for being here
        *G5- - - - - - - - - - - - - - - - - - - ≫
```

```
G1:  two hand pantomime bringing book to self
G2:  two hand pantomime placing book on table
G3:  two hand shrug, head tilted left, gaze down
G4:  one hand shrug, readjusts body
G5:  hand shake with Colbert
```

Figure 5.12: Gesture sequence corresponding to TRANSCRIPT 33: RENEE ZELLWEGER, ex. 99

There are two parts of this interaction that are particularly interesting for understanding the contribution of gesture to topic-shifts – (i) Zellweger's placement of an imaginary book on Colbert's desk, and (ii) her extended refusal to contribute to the discourse as expressed by shrugs and gaze aversion.

Zellweger's placement of the imagined book onto the desk (G2) can be considered an exceptional *removal* gesture. The imagined book metonymically represents the entire narrative. By 'shelving' the imagined book, Zellweger also metaphorically 'shelves' the narrative. This is exceptional in two ways. First, the 'removal' of the book is toward Colbert, directly into the shared Interaction Space. Given that one of the defining characteristics of removal gestures, as defined in this work, is movement *away* from the Interaction Space, the trajectory of this gesture seems problematic. I argue that despite its exceptional trajectory, this gesture can be still classified as a removal given the larger gestural context. Prior to the removal gesture, Zellweger is engaged in an imaginary interaction with Jimmy Carter, during which she is oriented away from Colbert. Because of this, the placement of the imagined book onto the desk between her and Colbert is, in fact, *away* from the relevant Interaction Space. It just so happens that the relevant Interaction Space in this case is an imagined one positioned perpendicular to the real space between her and Colbert.[37]

[37]See Chapter 4 for more extensive discussion of navigation between different real and imagined Interaction Spaces.

The second way in which this gesture is exceptional is that it appears to extend what begins as a pantomime into an interactive gesture. The placement of the imagined book onto the desk is not a part of Zellweger's re-enactment and does not convey information about what actually happened when she met Jimmy Carter. Instead, it serves as a clever and opportunistic way to convey discourse-structural meaning – to end the re-enactment and related narrative. Framed in the terms of *conceptual blending* (e.g. Fauconnier & Turner 1998), it appears that the 'book', which exists as a referent in the narrative space, functions as a discourse topic in the real space. Only after reorienting away from the narrative space (i.e. the imaginary Interaction Space), does the dual function of the 'book' become apparent. I do not know of significant work looking at the extension of pantomime gestures past their use in the pantomimed activity. This example suggests that such instances could provide unique insights into conceptual blending in multimodal discourse.

Zellweger's repeated use of shrugs and gaze aversion is also interesting in respect to the gestural expression of topic-shifting. If G2 does indeed constitute a removal action, then Zellweger has already performed a form of gestural topic-dismissal before saying "anyway". What then does "anyway" and its accompanying shrug communicate? Given the full gestural context, I suggest that Zellweger's removal gesture (G2) performs the initial topic-dismissal. The use of "anyway" subsequently reinforces this dismissal, while the shrug, coupled with her gaze aversion, serves to clarify that she does not intend to contribute to a new topic.

Because "anyway" simultaneously signals the dismissal and introduction of a topic, and because she has *already* signalled topic dismissal in the verbal mode, it is important that she clarifies her intentions. The shrug is a good candidate for doing so, and stands in direct contrast from the presentation gestures discussed elsewhere in this work. Shrugs have been associated with the inability or refusal to take an action (e.g. Debras 2017). This function is metaphorically related to the displaying of an open, empty hand (e.g. Müller 2004) – by exposing empty hands, one conveys that they have nothing to offer. The composition of the PUOH shape and orientation is disambiguated from its use as a presentational gesture through its composition with the shrug. Using the shrug where we might *expect* to see a presentation gesture serves as en effective way to cancel the expectation of a topic's introduction.

5.7.2 Readjusting to new structure

In (100), American actor Nick Kroll has just sworn, rather dramatically, on live television. This results in a full derailment of the discourse – Colbert and the audience are laughing, Kroll is turning red with embarrassment, and a series of confusing overlapping turns ensues until the floor is renegotiated and the discourse returned to. This example demonstrates the potential of *self-adaptors* to contribute to discourse management while also conveying the social and psychological information they are typically associated with (e.g. Ekman &

Friesen 1969; Neff et al. 2011; Waxer 1977).[38]

In this excerpt, Colbert and Kroll are discussing Kroll's recent experiences at Burning Man, a large music festival in the desert with a reputation for eccentric personalities and excessive drug use. Colbert asks if Burning Man is a cult. This results in a digression that eventually ends with Kroll making an inappropriate joke about the "gifts" a cult leader would ask for (line 1). This joke derails the discourse completely as Kroll realizes that the show is being aired live. During these the subsequent renegotiation of the floor (lines 2-5), Colbert remains oriented toward Kroll, while Kroll repeatedly shifts his head and gaze orientation from Colbert (P4) to the audience (P2, P5, P6) and then away from both (P3, P7). Through these re-orientations, Kroll's hands remain at rest in his lap except to perform two self adaptors. First, in his awkward silence immediately after the joke, he wipes his nose and looks away to his upper right (line 1, G2, P3). Second, he straightens his tie and straightens his back as he says the first of two "anyway"'s (line 6, G3). The discourse is finally returned to as Colbert presents a picture of Kroll at Burning Man, and Kroll confirms his re-engagement with the topic by repeating "yes" (line 6) and using two consecutive locating gestures (G4 & G5).

(100) Transcript 34: Nick Kroll
[UID:9fe08bd4-c79f-11e8-80e8-089e01ba0335,1868]

```
1  NK   △*White Nikes and a (bleeped) I guess △(laughs)* △ *(.)*
        △P1- - - - - - - - - - - - - - - - - -△P2- - - - △P3- ->
          *G3- - - - - - - - - - - - - - - - - - - - - -*   *G2-*
6       △Is this live?
        △P4- - - - ->
2  SC   uh-huh
3  NK   Yeah
4  SC   We're live, [we're live right now.  uh-huh.  uh-huh]
5  NK                [Yeah△ we're live, we're live, great, great]
        <- - - - - - - -△P5- - - - - - - - - - - - - - - - - ->
6       △*Anyway um* △(.)  anyway, △but- but- *yes and then- yes,
        △P6- - - - - △P7- - - - - △P8- - - - - - - - - - - - -≫
          *G3- - - -*                              *G4- - - - - - ->
7       we- uh* *the camp I stayed at was like* a grand hotel
        <- - -* *G5- - - - - - - - - - - - - -*
P1, P4:  head and gaze toward Colbert
P2:  head toward audience, gaze alternatively down or to audience
P3:  head turned right, gaze to upper right
P5:  head and gaze forward, between audience and Colbert
P6:  lean body and head toward Colbert, gaze toward audience
```

[38]A special thanks to my friend and colleague Kelly Jones (University of California, San Diego) for conversations about this particular example.

```
P7:  head forward, between audience and Colbert, gaze down
P8:  head toward Colbert, gaze at gesture
G1:  two hand shrug
G2:  wipes nose with back of hand
G3:  right hand adjusts tie
G4:  right hand, index finger point to picture, palm center
G5:  right hand, flat hand point to picture, palm center
```

Figure 5.13: Gesture sequence corresponding to TRANSCRIPT 34: NICK KROLL, ex. 100

It is clear that the locating gestures G4 & G5 are expressing topic introduction much in the same way as the locating gestures discussed in Section 5.6.2. In fact, the hand shapes of Kroll's two locating gestures exactly parallel those performed by Emma Thompson in (94). An index-finger point is used first as the main topic is reintroduced, and a flat hand point is used while situating the new contribution relative to this topic.

In contrast, it is initially unclear how Kroll's tie-straightening (G3) relates to the topic-dismissal marker "anyway" that it aligns with. As with Zellweger's shrug, Kroll's self-adaptor does not seem to relate directly to either the topic *dismissal* or topic *introduction* functions of "anyway". I suggest that the function of Kroll's self-adaptor more closely relates to the function of "um", which it also aligns with. Like "um", Kroll's self-adaptor signals his intention to continue. This particular performance also likely metaphorically signals Kroll's

intention to behave cooperatively by evoking the primary metaphors MENTAL STATES ARE PHYSICAL STATES and MORALITY IS CLEANLINESS (e.g. Dancygier & Sweetser 2014; Lakoff & Johnson 1980) – by physically tidying himself, he is also metaphorically 'tidying' his previously crude behavior.

Both of Kroll's self-adaptors (G2 & G3) *do* convey information about Kroll's psychological state, as is expected of self-adaptors – he is embarrassed and uncomfortable, and this is reflect in self-directed actions. However, I think it is also apparent that these movements, especially G3, convey discourse-structural information – he is preparing to make a cooperative contribution to the discourse.

5.7.3 Shifting strategies in repeated derailments

The final example I discuss in this chapter is one that we've already seen twice before in discussions of addressee-triggered returns (5.3.2, ex. 66) and presentation gestures (5.6.1, ex. 93). I return to it once more here in order to fully appreciate the diversity of verbal and gestural strategies Irish actor Saoirse Ronan employs as she attempts to recover from Colbert's repeated interruptions.

Colbert's first interruption occurs in line 9 as he sarcastically reminds Ronan that "there are ways to tell that a woman is pregnant". Ronan seems poised to incorporate this contribution into her storytelling as she dramatically raises her eyebrows, points enthusiastically at Colbert, and exclaims "yeah!" (line 10, G4). However, Colbert proceeds to immediately interrupt her again by evoking "the talk" (line 11) and euphemistically referring to sex as a "firm handshake" (line 13). In response to these two interruptions, Ronan awkwardly acknowledges Colbert's contribution with "I know" (line 12) and "right" (line 14). Through these two interruptions, Ronan's hands remain mostly at rest. After this, she begins to employ clearer lexical and gestural strategies.

She uses "anyway" for the first time in line 15, followed by the frame-setting phrase "7 months later". This is accompanied by the cyclic PUOH gesture discussed in Section 5.6.1. Here, the "anyway" signals the dismissal of the digressive topic and the continuation of the main discourse, "7 months later" specifies the part of the narrative that she intends to return to, and the large cyclic PUOH gesture signals her intention to continue contributing information. Despite all of this, Colbert interrupts again (line 17). Ronan again uses "anyway" to try to initiate a return shift, this time also employing "but" to reinforce the cancellation of the digression, and "so" to reinforce the intended continuation of her story. In the gestural mode, she performs a two-handed *clearing* gesture with a *blocking* hand shape and orientation (G9). The clearing movement serves to clear the space of digressive contributions, whereas the blocking shape and orientation signal a request for Colbert to stop his disruptive actions. Colbert interrupts one last time (line 20) before Ronan successfully takes the floor and continues her story.

(101) Transcript 35: Saoirse Ronan
[UID:92fedfb8-1b1f-11ea-9518-089e01ba0335,2060]

```
1  SC   I understand, uh, I understand that, uh, Greta was pregnant
2       for the entire shoot and you didn't know.
3  SR   None of us knew.  None of us knew-
4  SC   How did you not know?
5  SR   We didn't, so, we didn't know because she was just like
6       *amazing and like had this incredible like command over the
        *G1- - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
7       set.* *And we just didn't, it didn't even come into our
        <- -* *G2- - - - - - - - - - - - - - - - - - - - - - - - ->
8       head* *But also-*
        <- -* *G3 - - -*
9  SC   There are ways to tell that a woman is pregnant, by the way
10 SR   *Yeah*
        *G4- *
11 SC   I don't know if you've had the talk but
12 SR   I know- (mumbles) there- there were babies made.  (laughs)
13 SC   uh-huh, a firm handshake
14 SR   Right (laughs).  She, um, *so she, y'know, must've done
                                  *G5 - - - - - - - - - - - - - ->
15      that* *And she um she- I don't know (laughs)* *Anyway, 7
        <- -* *G6 - - - - - - - - - - - - - - - - - - * *G7- - - >
16      months later-*
        <- - - - - - *
17 SC   We don't know, we don't know
18 SR   *We don't know what happened.* *But s-* anyway-
        *G8 - - - - - - - - - - - - - -* *G9- - *
19 SC   It could have been immaculate, but-
20 SR   *There's a baby cooking in there.* *And it was like seven
        *G10 - - - - - - - - - - - - - - - - * *G11- - - - - - - - - ->
21      months in* *none of us knew*
        < - - - -* *G12- - - - - - *
G1:  both hands, loose open, variable orientation, small irregular
     cyclic movements
G2:  both hands, loose open, variable orientation, flick outward
G3:  right hand brushes hair, left hand begins presentational gesture
G4:  left hand index-finger point up, held toward Colbert
G5:  both hands, rotate to palm up orientation in lap
```

```
G6:  both hands, loose open, palm toward self
G7:  both hands, loose palm up, fast cyclic motion
G8:  both hands, palm up shrug
G9:  both hands, palm out, flat, arched movement outward
G10:  left hand, palm center, fist, rotates at wrist with small
      lateral movements
G11:  right hand, palm center, flat, moves laterally inward
G12:  both hands, loose containment gesture
```

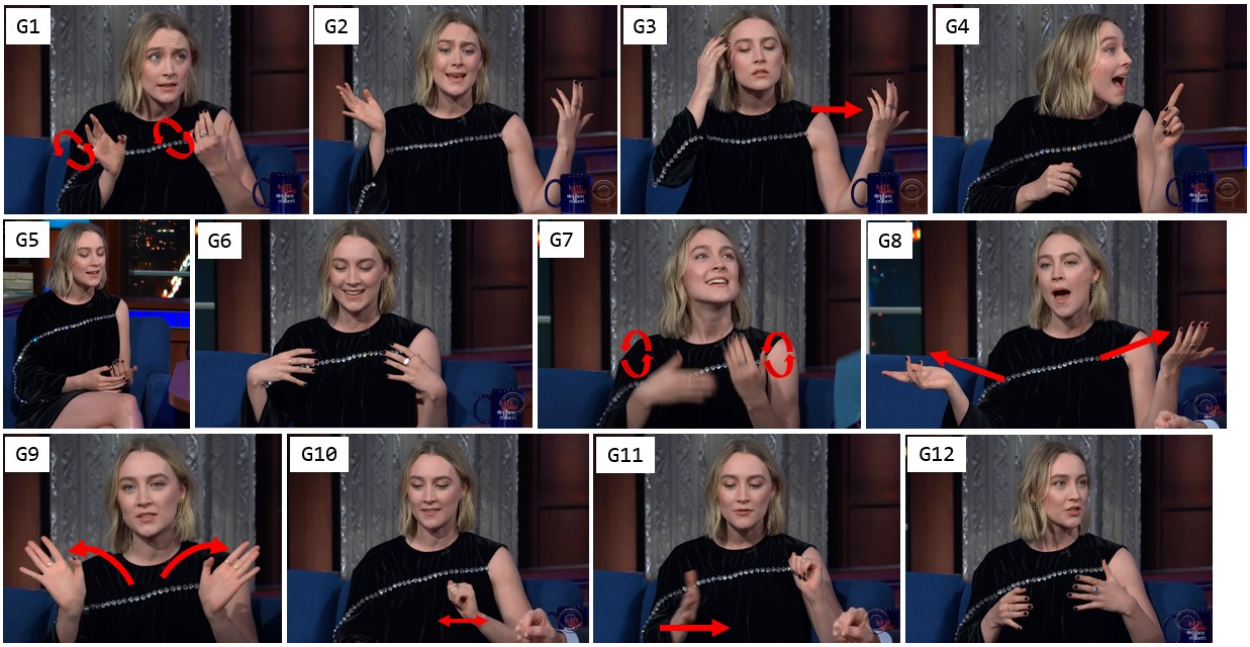


Figure 5.14: Gesture sequence corresponding to TRANSCRIPT 35: SAOIRSE RONAN, ex. 101

This particular discourse is an incredibly effective demonstration of the principles of *multiple strategies* and *optionality of expression*. Ronan attempts what is basically the same discourse move, a return shift, *five times*. Each time, she uses a different set of verbal and gestural strategies to express the move. As the the interruptions continue, she employs more strategies at once by including more lexical discourse and more complex interactive gestures. The shifting of strategies emphasizes the proposed *optionality of expression* principle – Ronan has access to a multitude of strategies to express her desired return to the narrative and chooses from these with each attempt, thus demonstrating that any single strategy is optional.

5.8 Summary

In this chapter, I have explored the ways in which interactive gestures convey discourse-structural meaning alongside the topic-shift marker "anyway". I argued that the observed variation in the gestural mode is a reflection of equal variation in the verbal mode. To demonstrate this, I considered discourse-structural factors that are left under-specified by "anyway", including shift-type, shift-trigger, additional lexical discourse markers, and the position of the shift relative to the turn-taking structure. I then discussed the ways in which different gesture performances correlate with different discourse-structural factors.

In general, the prediction made as to gestural expressions of topic-shifting were supported, particularly in cases of *removal* gestures for expressing topic dismissal and *presentation* gestures for signalling topic introduction and continuation. I also showed that seemingly idiosyncratic gestures nonetheless follow general patterns by evoking the same kind of action schema. I ended by emphasizing the importance of considering the full discourse-structural and social context when analyzing discourse management strategies in both the verbal and gestural modes.

Chapter 6

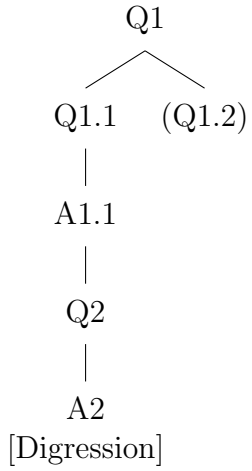
By the way, digression is multimodal

6.1 Introduction

Digression is a discourse structural phenomenon in which a discourse segment (one or more utterances) *relates* to the preceding discourse, but does not *contribute* to furthering current discourse goals. In a question-based approach to discourse structure, this means that the information provided in a digression may be *thematically* related to other information in the discourse, but does not provide an answer or partial answer to an open discourse question. In this chapter, I explore the gestural marking of digression as performed with the lexical digressive marker "by the way" (henceforth BTW). To do so, I consider the distribution of four classes of action schematic gestures introduced in Chapter 2 (Presentation, Referring, Stopping and Engagement) across use-contexts of BTW. I argue that *stopping* gestures in the present data set express a temporary stop in the pursuit of discourse goals and function as digressive markers along with BTW. The remaining three action schema can be used to signal digression by spatially differentiating the metaphoric location of digressive information from the main discourse, but are not digressive markers in and of themselves.

As proposed in Chapter 3, a *digression* is a discourse move that pursues a question embedded under an *answer*, rather than an open question (van Kuppevelt 1995; Riester 2019). This kind of contribution may be thematically related to a topic under discussion. However, it does not contribute to achieving discourse goals because it does not contribute an answer or partial answer to an open question. Importantly, the most recent still-open question is returned to once the digression ends. This is represented schematically in (102). The digression (Q2) is related in some way A1.1, as reflected by its attachment location, but it does not offer information that is helpful in resolving the higher Q1. The expectation of another partial answer to Q1 is maintained, despite the presence of the digression. This expectation is represented as the potential question "(Q1.2)".

(102)



Digression, under this analysis, is considered a minimally disruptive discourse move so long as (i) the digressive segment is relatively short, and (ii) the digression is coherently related in some way to what is happening in the interaction. The longer a digression lasts, and the less related it is to established discourse topics, the more disruptive the digression is to overall discourse coherence. Consider the three fictitious digressions in (103) as an illustration. The bracketed region is considered the digressive segment.

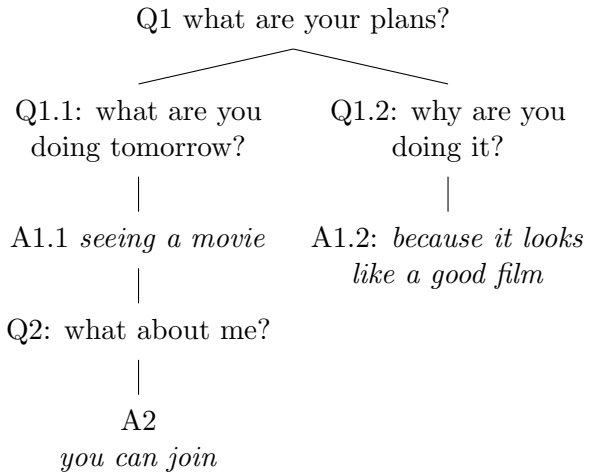
(103) a. Tori and I are going to see a movie tomorrow night at the Grand Lake Theater. [(*By the way*), you're welcome to join!] The movie is getting really good reviews, so I'm pretty excited.

b. Tori and I are going to see a movie tomorrow night at the Grand Lake Theater. [(*By the way*), you're welcome to join! Tickets are half off on Tuesdays. Plus Tori would love to meet you.] The movie is getting really good reviews, so I'm pretty excited.

c. Tori and I are going to see a movie tomorrow night at the Grand Lake Theater. [(*By the way*) your new puppy is really cute. She kind of reminds me of my childhood dog named Daisy. I got her when I was five and loved her so much.] The movie is getting really good reviews, so I'm pretty excited.

The digression in (103a) is relatively *un*-disruptive. It is very short and is serving a positive social function that is related to the ongoing discourse topic – I am conveying information about my plans and I am being friendly by inviting you to be a part of those plans. In fact, the disruption to the discourse is so minimal in this case that I can still reduce the first referent after the digression, "the movie", to the pronoun "it" without causing confusion. A representation of this discourse's structure is provided in (104).

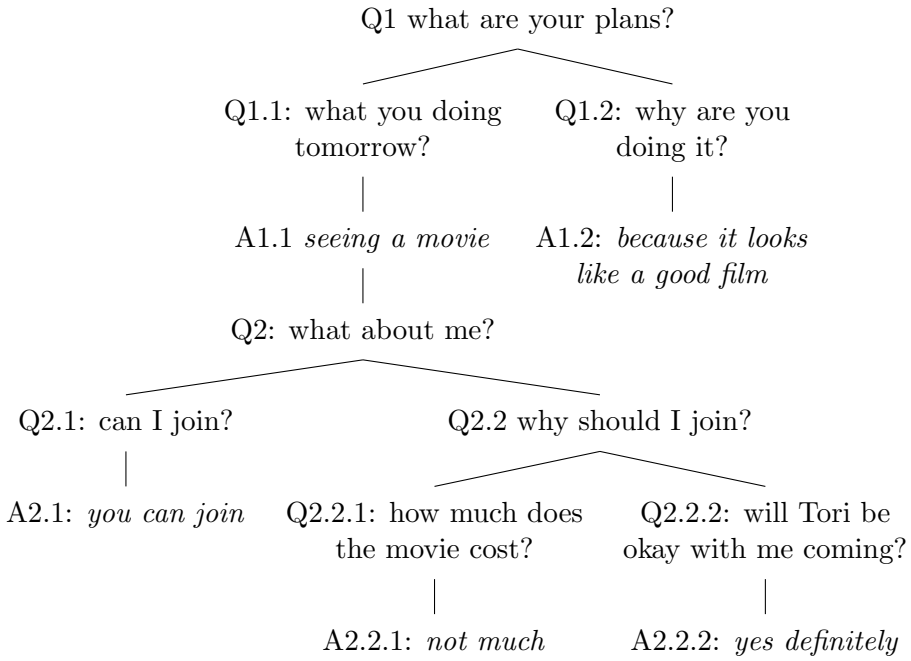
(104)



The digression in (103b) is also clearly related to the ongoing discourse topic and seems to serve a similar social purpose. However, it is quite long. By the second sentence in the digression, you might think that I've introduced a new discourse goal – we were talking about *my plans*, but now we're talking about what I think *your plans* should be. By the third sentence, you may assume that I have nothing else to say about my plans. When I finally do return to describing my plans, it might seem a bit jarring, and if I were to say "it" rather than "the movie", you would likely be momentarily confused.

To better understand why the longer digression in (103b) is more disruptive than the first, we can compare the complexity of the discourse structures in (104) and (105). With each additional sentence in the digression, the digression's structure gets more complex as more potentially open *digressive* questions are posed. By the third sentence of the digression, I could continue talking about *why Tori won't mind extra company*, *why else I think you should join*, or w*hat else you could do while I watch a movie*. Addressing these questions is, in a sense, more immediately relevant – these digressive questions were posed more recently than any non-digressive question.

(105)



The final digression in (103c) is even worse. It is difficult to recover any meaningful relationship between *my plans* and *your new dog*. You might be flattered by my compliment, but you're also probably confused about why I offerred it when I did. You might even become annoyed as I start talking about my old childhood dog, given the seeming irrelevance. If I replaced "the movie" with "it" after this digression, you would likely have not know what I was referring to.

Previous work on lexical discourse markers demonstrates that the more unexpected a contribution is, the more likely it is to be marked with a lexical discourse marker (e.g. Murray 1997; Sanders 2005). In a question-based approach to discourse structure, the "expected" contribution is always one that offers an answer to the most immediate open question, or suggests a better strategy for answering it. A contribution is *unexpected* if it fails to do either of these things. If a contribution is *too* unexpected, we risk derailing the discourse and having to openly renegotiate discourse goals. For example, in response to the digression in (103c), it would be reasonable to stop the discourse and say, "okay, why are you telling me this?" or "what does this have to do with the movie you're going to see?".

To alleviate risks of derailment, an interlocutor may 'warn' their addressee that they are about to do something unexpected. BTW is a lexical strategy for doing so. Though it is not 'obligatory' in any of the examples in (103), it's usefulness as a cue for upcoming unexpected structure is increasingly apparent as the digression gets more disruptive. In (103a), I could very well leave BTW out. If I still wanted to cue the digression, I could opt for a prosodic strategy instead – talking faster and lower to mark the information as somehow 'less' important to the ongoing discourse (e.g. Bolinger 1986; Local 1992).[1] For

[1]Note that this strategy relies on iconicity – by making speech less perceptually salient, we mark it as

longer digressions, it may be most efficient to preface the unexpected contribution with a lexical digressive marker like BTW, *incidentally*, or *parenthetically* (e.g. Mittwoch et al. 2002; Traugott 2020). Doing so will help keep discourse goals salient by overtly signalling that they will be returned to. If the digression is too long or too irrelevant, the end of the digression can also be signalled with a resumptive marker like "back to the point" or topic-shift marker like "anyway".

In order to better understand the contributions of lexical and gestural digressive markers in maintaining discourse coherence through a digression, I carefully consider variation in both the verbal (Section 6.3) and gestural (Section 6.4) modes. As in previous chapters, I use the Multimodal Discourse Management hypothesis, and the four proposed principles therein, to guide my discussion. These are reiterated for digression in particular in (106).

(106) **Proposed principles of Multimodal Digression**

a. *Multiple strategies*: There are multiple strategies, within both the verbal and gestural modes, for expressing the digressive status of an uttrance.

b. *Optionality of expression*: The use of each individual expressive strategy is optional and subject to contextual variation.

c. *Independence of contribution*: Strategies may be employed independently and may profile a different aspects of the digression.

d. *Compositional management*: The strategies employed in both modes are integrated systematically and predictably into a single coherent multimodal message.

Based on both existing literature, and previous discussion in this work, I make the following predictions for the gestural strategies used to mark digressions. No predictions are made for *addressing* gestures (a sub-type of the REFER action schema) because they are expected whenever action is requested of the addressee (e.g. Bavelas et al. 1992; Bavelas 1994).

(107) **Gestural expression of digression**

a. *Stopping gestures*: If BTW marks a temporary digression from discourse goals, we expect stopping gestures to be used to temporarily prevent the metaphoric forward motion of discourse participants toward those goals. This prediction is supported by work on "blocking" gestures which demonstrates the relationship between metaphorically holding an object away and a request to stop a line of action (e.g. Wehling 2017).

b. *Presentation & locating gestures*: If the information provided in the digression is not relevant to discourse goals and is not meant for further discussion, they may be located outside of the Interaction Space. This is because the Interaction Space is the space in which discourse topics are placed and organized. Locating a topic *outside* of this mutually accessible space may signal that it is not meant for further manipulation or management.

less worth our attention.

c. *Alternating engagement with the Interaction Space*: If BTW signals a temporary departure from discourse goals, we may expect to see patterns of *disengagement* and *re-engagement* with the Interaction Space, where the period of *disengagement* signals the presence of the digression. This is because interlocutors are expected to orient themselves toward the Interaction Space when participating in the main discourse. When they are *not* participating in the main discourse and the topics therein, they may signal this lack of participation with re-orientation away from the Interaction Space. Reorientation of the head and torso away from the addressee was previously observed in digressions by Hinnell (2020, Ch. 5).

These predictions are largely supported by the gestural behavior observed in the present data set. However, there is also significant variation in gesture performance, including in the articulators involved and in particular formal features, such as hand shape. Potential functional motivations for this formal variation are discussed throughout.

This chapter is structured as follows. Section 6.2 describes the data set and annotation methodologies used. Section 6.3 reviews variation in the form and function of BTW-marked digressions as observed in the present data set. In Section 6.4, I describe recurrent manual gesture patterns and discuss their relationship to other discourse-structural factors. Section 6.5 concludes.

6.2 Data & Methodology

All data comes from interviews and monologues on the talk show Late Show with Stephen Colbert, collected through the UCLA television news archive in collaboration with the Red Hen Lab. An initial search for the string "by the way" on the Late Show with Stephen Colbert between 2016 and 2019 provided 916 results (including repeated clips and commercials). The most recent 100 unique instances were annotated, resulting in a data set ranging from February 2019 to December 2019. Of the 100 annotated results, 68 occurred during interviews, and 32 during monologues.

All examples were coded for the following variables. Detailed definitions of variables are discussed where relevant throughout the paper.[2]

(108) a. Discourse type: *interview, monologue*[3]

b. Speaker: *Colbert, Guest*

c. Lexical unit: *by the way* and any accompanying discourse markers (e.g. *and by the way*)

d. Clause position: *beginning, middle, end*

e. Discourse relation: *background, elaboration, evaluation, meta-discursive, shift*

[2]Annotations are available upon request.

[3]For interview data, the name and demographic information of the guest were also recorded, including *occupation, gender, race*, and *nationality*.

All 100 examples were also coded for the presence of manual gesture. Where manual gestures did occur (84/100), the *handedness, hand shape, palm orientation,* and *movement* were annotated, as described in Chapter 2. These features were used to determine the gestural class as well as the alignment of the hand gesture relative to BTW which was annotated as *address, block, present, refer*, or *remove*, or was left blank if none of the 5 classes were suitable. See discussion in Chapter 2, Section 4 for deriving gesture class from morphological gesture features. Finally, each instance was notated with a 1-2 sentence qualitative description of gestural activity and communicative context.

The unique corpus ID (UID) is provided for each example discussed. Where only speech is of concern (Section 6.3), the speaker's name and relevant discourse fragment is presented as a simple transcript. Where gesture is also of concern (Section 6.4), a multi-tier transcript of the discourse fragment is provided, using conventions from Embodied Conversation Analysis (Mondada 2018) to highlight gesture-speech alignment.

6.3 Variation in the verbal mode

As is typical of lexical discourse markers, *by the way* serves different functions in different contexts (e.g. Fischer 2000; Fraser 1999). In line with Charolle's (2020) analysis of digressive markers in French, BTW appears to be *multifunctional* in that it both signals a particular type of discourse *move* (a digression) and is associated with particular discourse *relations*. In its first function, BTW sets expectations as to the type of contribution that is being made. In particular, it serves as a warning that the contribution is going to be *unexpected*, i.e. that it is not going to address an open discourse question. In its association with particular discourse relations, BTW helps to signal that the information conveyed in a contribution will relate to immediately preceding information in a particular way. The interaction between these two functions, and observed variation in both, is addressed in Section 6.3.1.

The discourse marker BTW is just one lexical strategy available to speakers for marking a digression in the verbal mode (e.g. Traugott 2020). Because it is but one strategy of many, BTW is also subject to significant structural variation as speakers employ different strategies in different orders to mark the digression. This structural variation is addressed in Sections 6.3.2 & 6.3.3.

6.3.1 Discourse-structural variation

To understand and categorize the connection between the digression and immediately preceding discourse structure, I use the conception of 'discourse relation' as established in Rhetorical Structure Theory (Mann & Thompson 1988; Taboada & Mann 2006). Under this approach, a discourse relation is an underlying proposition that conveys information about how two adjacent discourse segments relate to one another. These relations may or may not be expressed overtly in the verbal (or gestural) mode. This conception of discourse relation aligns with the understanding of digression as a discourse structural phenomenon, rather

than a product of overt linguistic expressions. It also allows for the *optionality of expression* inherent in my proposed principles of multimodal discourse management.

In the data set, I identified four distinct discourse relations that hold between the main topic and the host clause of BTW – *background, elaboration, evaluation, meta-discursive*. Three of these relations (*background, elaboration,* and *evaluation*) are taken directly from Mann & Thompson's (1988) proposed discourse relation inventory. The meta-discursive relation is particularly associated with spoken face-to-face discourses in which participants can react in real time to how the discourse is progressing. Because the the RST framework is based primarily on written data (Taboada & Mann 2006), there is no obvious RST correlate to the meta-discursive relation observed here. I treat the recurrence of these four relations in the data set as indicative of four *types* of digression.

In addition to these four digression types, I observe two other use-contexts of BTW – *shift* and an idiosyncratic *joke* construction unique to Colbert's monologue. The joke construction can be considered an idiosyncratic digression type. Shift contexts, however, do *not* constitute digressions – if the preceding topic is not returned to, the structure is not a digression under the present approach. I argue that BTW remains a 'digressive marker' in these cases, but serves a politeness function by marking a non-digression *as if it* were a digression. We will see an example of this in Section 6.3.1.5.

Simple definitions for the six recurrent use-contexts of BTW are given in Table 6.1.

Relation	Definition
Background	a digression that provides information necessary or helpful for the comprehension of an adjacent discourse segment
Elaboration	a digression that contributes information related to an ongoing discourse topic
Evaluation	a digression that provides a speaker's opinion or assessment of an ongoing discourse topic
Meta-discursive	a digression that comments on the discourse itself
Shift	a move in which a discourse goal is abandoned in favor of a newly introduced topic
Joke	an idiosyncratic construction in which Colbert prefaces a pun with "by the way"

Table 6.1: Use-contexts of BTW

The *background, elaboration,* and *evaluation* use-contexts are similar in that the BTW-

marked digression contributes information to the discourse that is directly related in some way to the topics already being discussed. Meta-discursive digressions and shifts, on the other hand, do not provide information about topics that are already under discussion. Meta-discursive digressions comment on *how* topics are being discussed by interlocutors, and shifts, as discussed in Chapter 5, introduce a new topic to the discourse.

Table 6.2 summarizes the distribution of the 6 use-contexts of BTW as observed in the present data set. In one use of BTW, the speaker is immediately interrupted, making the intended function of BTW unrecoverable, hence the count of 99. There are three noteworthy differences in the use-context distributions in interviews and monologues. As already noted, the 'joke' use-context is an idiosyncratic construction that is recurrent in Colbert's monologues, and, by definition, does not occur in interviews. Shifts and meta-discursive uses occur almost exclusively in interviews, suggesting particular social functions of BTW in these contexts.

	Background	Elaboration	Evaluation	Meta	Shift	Joke	**Total**
Interview	17 (25%)	12 (18%)	12 (18%)	12 (18%)	14 (21%)	- -	67/68 (99%)
Monologue	6/32 (19%)	4/32 (13%)	5/32 (16%)	2/32 (6%)	1/32 (3%)	14/32 (44%)	32/32 (100%)
Total	23 (23%)	16 (16%)	17 (17%)	14 (14%)	15 (15%)	14 (14%)	**99** (99%)

Table 6.2: Distribution of use-contexts with BTW

In the remainder of this section I provide examples of each use-context, and discuss the function of BTW in each case. Where relevant, I make predictions as to how the particular use-context may affect the gestural expression of digression.

6.3.1.1 Background

In a *background* context, the digression marked with BTW provides information that helps the interlocutor to understand some aspect of the previous discourse segment. In this way, these digressions preempt a possible clarifying question about the main line of inquiry, but do not contribute directly to the main discourse goal. For example, the discourse excerpt in (109) shows BTW as marking an overt clarifying question. In this discourse, Elizabeth Warren (American politician) and Colbert are discussing the impacts of a policy agenda which would increase taxes on the "rich" in order to fund universal access to healthcare in the United States. Colbert preempts a question about whether or not these taxes would affect middle class workers by asking Warren to clarify her definition of "rich" (line 2), marking the digression with BTW.

(109) **Background**

TRANSCRIPT 1: STEPHEN COLBERT TO ELIZABETH WARREN
[UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335,2121]

```
1  SC   This is a mild lightning round.  Mild lightning round.
2       How rich is rich for this two cent thing, by the way?
3  EW   Oh, it is people who have fortunes of above fifty million
4       dollars
```

This type of digression is considered minimally disruptive because the information provided is helpful in achieving discourse goals, even if it does not directly contribute to answering an open question. In fact, such digressions may help *prevent* future derailments by resolving points of misunderstanding before they arise. This type of digression also serves a particular social function. By embedding clarifying information in a digression, the speaker alleviates the risk of 'over-telling' while also ensuring that their interlocutor has all necessary background information.[4]

Because the information provided in these digressions is helpful for assuring that discourse questions are answered sufficiently, we expect *presentation* and *locating* gestures into the central Interaction Space – information that is relevant to the ongoing discourse should be mutually accessible within the space.

6.3.1.2 Elaboration

In an *elaboration* context, the BTW-marked digression provides information that is thematically related to an ongoing discourse topic, but is not helpful in achieving discourse goals. Unlike background digressions, which can be interpreted as 'for-the-addressee', elaboration digressions seem to be wholly 'for-the-speaker'. There is some information that the speaker wants to share, but recognizes that it is not particularly relevant for discourse goals, and so marks it as a digression.

An amusing, self-aware use of an elaboration digression is provided in (110). In this joke, Colbert is satirizing a recurring scandal during the Trump-presidency in which then-president Donald Trump would re-tweet something terrible and then argue that it was not a big deal because he didn't *write* something terrible.

(110) **Satirical elaboration**

TRANSCRIPT 2: STEPHEN COLBERT
[UID:a5309d94-be5d-11e9-a8ce-089e01ba0335,387]

```
1  SC   What?  I didn't stab you.  That was somebody else's knife.
2       I re-stabbed you.  And by the way the guy whose knife I used,
3       a very respected stabber
```

[4]See work in Enfield et al. (2007) for analagous findings in the articulation of pointing gestures.

The imagined discourse goal in this joke is presumably answering the question "who stabbed you?". After offering a nonsensical answer (lines 1-2), Colbert performs an elaboration digression which provides intentionally insensitive and unhelpful information about the knife's owner (lines 2-3). Though information about the imaginary "stabber" is thematically related to the topic under discussion, it certainly does not help Colbert prove his innocence.

In the present data set, many of the elaboration digressions seem to serve a face-saving function by providing additional unnecessary information intended to counter some perceived negative judgement. For example, in (111), Eddie Redmayne (British actor) is laughing while telling a story about an accident that happened on a recent movie set. Because his laughter may portray him as insensitive to another actor's pain, he clarifies through a BTW-marked digression that the actor was not actually hurt.

(111) **Face-saving elaboration**
TRANSCRIPT 3: EDDIE REDMAYNE
[UID:bf03ea3c-1731-11ea-a000-089e01ba0335,2085]

```
1  ER   I heard Felicity go 'I don't think I can move my neck'.  And-
2       and so, that's how we started this film.  And- and from that
3       moment, wh-wh-what was wonderful was that a lot of the film
4       was then shot on green screen.  But we had- and she was
5       totally fine, by the way.  No Felicities were hurt in the
6       making of this movie
```

This type of digression is variably disruptive, depending on its length and perceived relevance of contribution. The example in (111) ends up being particularly disruptive as Colbert interprets Redmayne's digression as revealing unnecessary "spoilers". We will return to this particular example in Section 6.4.1.

Because the information provided in an elaboration digression is related to ongoing discourse topics, locating gestures within the Interaction Space may occur. If the related topic is already established in the Interaction Space as a metaphoric object, it may be referred to via a deictic gesture while providing the elaboration, as if to say "this additional information is about *that* topic".

6.3.1.3 Evaluation

In an *evaluation* context, the digression marked with BTW expresses some attitude of the speaker toward a referent in the ongoing discourse. In (112), Colbert is introducing a hot topic in the news which concerns something called "diplomatic cables". He momentarily digresses from the main line of inquiry (summarizing the news item and providing commentary) to express an opinion about the term "diplomatic cables".

(112) **Evaluation**
TRANSCRIPT 4: STEPHEN COLBERT TO JOHN OLIVER
[UID:804021ac-aab8-11e9-8801-089e01ba0335,2090]

```
1  SC  There was a dustup kind of in the press and then in between
2      politicians because these cables, diplomatic cables.  I
3      love the term diplomatic cables, by the way
```

These digressions are minimally disruptive given that they are thematically related, often very short, and always have a clear social motivation – the speaker wants to share their opinion, but recognizes that it isn't necessary for achieving discourse goals.

6.3.1.4 Meta-discursive comments

In a *meta-discursive* context, the digression marked with BTW comments on the social interaction and the knowledge states of the interlocutors, rather than the discourse content.[5] In my data, this frequently takes the form of platitudes and assessments that the speaker makes about the state of the discourse, as in (113). For example, in (113), Colbert is talking to Andrew McCabe (American attorney) about a recent book he published. Colbert briefly interrupts McCabe's claims of cooperating with the FBI to jokingly say that McCabe is "under oath". That McCabe is "under oath" does not contribute to the main discourse topic (describing the writing of the book), and is not taken up by McCabe as a possible topic other than by briefly acknowledging the joke before returning to his story.

(113) **Meta-discursive**
TRANSCRIPT 5: STEPHEN COLBERT TO ANDREW MCCABE
[UID:UID:08beb46e-34e2-11e9-88b3-089e01ba0335,2242]

```
1  SC  you're under oath, by the way, you're under oath
2  AM  Thank you, okay.  Y'know, I worked very closely with the FBI
```

In the present data set, these digressions tend to be moderately disruptive, requiring at least one turn per participant before the main discourse is returned to. They can also be indicative of an already disrupted discourse – when the discourse is not progressing as expected, it is likely worth commenting on.

Because meta-discursive digressions are associated with the interaction as a whole rather than the discourse structure as such, we expect to see *addressing* gestures as interlocutors discuss each other's actions.

6.3.1.5 Shifts

As defined in Chapter 3, a *shift* is a discourse move that prematurely closes an open discourse question in favor of a new question. In such cases, the content of the discourse segments adjacent to BTW are, at most, tangentially related. I argue that when BTW is used to mark a topic-shift, performs an important social function – lessening the disruption of a topic-shift by marking it *as if it were* a digression. To illustrate this point, consider the following example.

[5] Pons Bordería & Estellés Arguedas (2009) identify similar digressions as 'interactional-based digressions'.

In (114), Colbert is talking to Thom Yorke, a well-know British musician, about a rumor that the Queen may remove then-prime minister Boris Johnson from office (lines 1-2). The discourse is humorously derailed as Yorke refuses to comment on the Royal Family's usefulness, or lack thereof, on live television (lines 3-6). Colbert attempts a recovery by joking about the Royal Family throwing "great parties" (line 7). This attempted recovery doesn't go well either. After an awkward pause, Colbert abruptly shifts the topic to Yorke being knighted. The only connection between these two segments is the presence of the Queen as a participant.

(114) **Shift**

TRANSCRIPT 6: STEPHEN COLBERT TO THOM YORKE
[UID:705c6f26-02c3-11ea-ba63-089e01ba0335,3329]

```
1  SC   It would be so exciting to see a member of the Royal Family
2       do something useful
3  TY   (laughs) You said that, not me.  I can't say anything.
4       I can't say- I- I'm merely commentating
5  SC   yeah
6  TY   observing
7  SC   I'm sure they throw great parties, I'm sure they throw great
8       parties
9  TY   the royal family?
10 SC   yeah
11 TY   (chuckles, nods, mouths "yeah")
12 SC   would you - would you want to be knighted by the way?
13 TY   nuh- I- I think I've blown that chance
```

With topic-shifts, we expect to see gestures similar to those that occurred with "anyway" in Chapter 5. The only important difference in predictions is that we expect to see *removal* gestures in which the dismissed topic, as a metaphoric object, is removed from the Interaction Space. *Removal* gestures are not expected for any other use-context of BTW.

6.3.1.6 Idiosyncratic joke construction

Of the 32 monologue examples, nearly half (14/32) are a recurrent joke construction unique to Colbert's monologues. In these constructions, Colbert digresses from a news story to make a pun, prefacing the pun with BTW. Once such example is given in (115).

(115) TRANSCRIPT 7: STEPHEN COLBERT
[UID:c7cce1dc-ff9e-11e9-9f3f-089e01ba0335,794]

```
1  SC   Cruz claims a quid pro quo is not illegal unless there is
2       "corrupt intent".  By the way, "corrupt intent" is also what
3       you call it when Trump goes camping
```

Because these appear to constitute an idiosyncratic, well-rehearsed performance, I do not make predictions as to what types of interactive gestures are expected. However, we can hypothesize that if this is a *multimodal* rehearsed construction, then similar gestures should be used across examples.

6.3.2 Clausal position

In addition to functional variation, BTW is also subject to structural variation in the data set. In this section, I demonstrate the observed variation in the clause position of BTW, which can occur at the beginning, in the middle, or at the end of its host clause. It is worth noting that the two other lexical discourse markers considered in this work (*anyway* and *here's the thing*) only occur clause-initially. The examples in (116) demonstrate the observed variation in clause position of BTW.

(116) a. Transcript 8: Chris Christie
[UID:23c24db2-c0b9-11e9-a047-089e01ba0335,2684]
1 CC By the way, I forgot Kirsten Gillibrand.

b. Transcript 9: Helen Bonham
[UID:42b7e700-0b68-11ea-8e77-089e01ba0770,2843]
1 HB She was tiny, by the way.

c. Transcript 10: Craig Ferguson
[UID:2d0e822a-a46f-11e9-80c4-089e01ba0335,1669]
1 CF Do you like this, by the way, the untucked shirt?

Of the 9 'clause medial' examples, 3 occur in *appositive relative clauses*. Appositive relative clauses serve as a *syntactic* cue for digression, independent of BTW (e.g. Loock 2007). This means that in examples like the one given in (117), the digression is (at least) doubly marked – both the lexical discourse marker BTW and relative clause head "who" signal the presence of a digression. This particular example is also marked with the lexical discourse marker "coincidentally" which provides additional information as to how the digression relates to the ongoing discourse topic (namely one of Trump's financial scandals).

(117) Transcript 11: Preet Bharara
[UID: 0269d444-a3a6-11e9-a2b4-089e01ba0335,2860]
1 PB But to call and try to cultivate a relationship with the
2 U.S. attorney in the southern district of New York, who,
3 by the way, coincidentally, has natural jurisdiction of
4 the Trump organization and the Trump foundation and
5 various other things, I didn't think it was appropriate

The distribution of the clausal position of BTW in the present data set is summarized in Table 6.3.

	Beginning	Middle	End	**Total**
Interview	32 (47%)	14 (21%)	22 (32%)	68 (100%)
Monologue	18 (56%)	7 (22%)	7 (22%)	32 (100%)
Total	50 (50%)	21 (21%)	29 (29%)	**100** (100%)

Table 6.3: Clause position of BTW

The clear plurality of clause-initial BTW in my data set aligns with previous work on discourse markers that argue for initial position as the default position whenever discourse coherence is at stake (Givón 1983). Because digressions, by definition, do not contribute to the current discourse goals, they threaten discourse coherence. If the addressee does not accurately identify the utterance as digressive, a derailment may occur. For example, an addressee may interrupt to ask "wait, why are you telling me this?". Such an analysis is further supported by work showing that discourses are, in fact, *incoherent* when lexical digressive markers are not used in certain contexts (e.g. Ariel 2010).

Given the functional variation of BTW, we may expect BTW to occur in different positions in different use-contexts at different rates. If Givón (1983) and Ariel (2010) are correct in saying that discourse markers occur more frequently when there is greater risk of incoherence, then BTW whould occur at the beginning of the clause more frequently in more disruptive use-contexts. In Section 6.3.1, I suggested that *shift* and *meta-discursive* use-contexts were the more disruptive given that the host clause of BTW contributes information that is unrelated to the discourse topic at the time. This does not hold for the present data set, as summarized in Table 6.4. Note that counts are out of 99 rather than 100 because the relation of one instance of BTW could not be determined due to the speaker being immediately interrupted.

There are two things to note about the distributions of BTW summarized above. First, *evaluation* digressions are the only use-context in which BTW appears clause-finally in the majority of cases. In the previous section, I suggested that *evaluative* digressions are minimally disruptive because they (i) tend to be short, (ii) are thematically related to the ongoing discourse topic, and (iii) serve a clear social function. The predominance of clause-final BTW in evaluations can thus be said to adhere to Givón (1983) and Ariel (2010) claims described above. Second, Colbert's idiosyncratic joke construction showed the least amount of variation in BTW clause position. This supports the proposal that these instances are rehearsed performances with a set form.

	Beginning	Middle	End	**Total**
Background	12/23 (52%)	4/23 (17%)	7/23 (30%)	23/23 (100%)
Elaboration	8/16 (50%)	5/16 (31%)	3/16 (19%)	16/16 (100%)
Evaluation	4/17 (24%)	3/17 (18%)	10/17 (59%)	17/17 (100%)
Meta-discursive	7/14 (50%)	2/14 (14%)	5/14 (36%)	14/14 (100%)
Shift	8/15 (53%)	3/15 (20%)	4/15 (27%)	15/15 (100%)
Joke	10/14 (71%)	4/14 (29%)	0/14 (0%)	14/14 (100%)
Total	**49/99** (49%)	**21/99** (21%)	**29/99** (29%)	**99/99** (100%)

Table 6.4: Clause position of BTW

These observations of digressive marking in the verbal mode can be extended to make predictions about marking in the gestural mode. In order to ensure that addressees accurately identify an utterance as digressive, speakers can take advantage of multiple strategies – lexical (e.g. Ariel 2010; Charolles 2020; Traugott 2020), syntactic (e.g. Loock 2007), prosodic (e.g. Bolinger 1986; Crystal 1969; Local 1992), and gestural (e.g. Bavelas 1994; Bavelas et al. 1995; Hinnell 2020). The more threatening to coherence a contribution is, the more clearly marked it should be. Furthermore, the less effective BTW is at marking the digression (e.g. when it doesn't appear until the end of the clause), the more we expect to see other strategies employed.

6.3.3 Adjacent discourse markers

Additional lexical discourse markers occur adjacent to BTW in 16 examples. In each case, the additional lexical discourse markers serve different discourse management functions, and indicate features of the discourse move that are left underspecified by BTW. Two of the most interesting examples are discussed below.

In (118), Colbert is asking American journalist John Dickerson about why he thinks the Democratic party isn't more openly critical of billionaires. In the except given, Colbert and Dickerson cooperate in framing the issue. Colbert suggests that redistribution of wealth is a deeply American notion (lines 4-7). Then Dickerson adds a pragmatic note about how those "607 billionaires" are probably not in politically important districts, so there is even less of a reason for Democratic politicians to care about their opinions (lines 8-11). Dickerson's BTW-marked contribution is considered a *background* digression in that it is providing additional information that is helpful for answering the open discourse question ("why do we care?"). Dickerson prefaces this contribution with a string of lexical discourse markers "uh well and by the way".

(118) Transcript 12: John Dickerson
[UID:465f22fa-01fa-11ea-8c13-089e01ba0335,1987]

```
1   SC   There are only, according to Fortune Magazine, 607
2        billionaires in the world.  Why do we care what they think?
3   JD   and-
4   SC   And because there's a lot more of us.  And, they've got
5        too much money.  Give your money to people who need it.
6        And collective- collective distribution of you wealth is
7        not anti-American, it's as American as Eisenhower
8   JD   uh well and by the way of those number of billionaires
9        there probably aren't many in the caucus states, in Iowa
10       and in New Hampshire.  So they're really even less
11       important
```

Each of Dickerson's four lexical discourse markers convey some discourse-structural information about his intended contribution. The "uh" serves as an initial signal for his intention to take the turn, or, perhaps more accurately, that he would like to continue the turn he started in line 3 (e.g. Clark & Tree 2002). The "well" acknowledges the multiple viewpoints being used to frame the issue (e.g. Aijmer & Simon-Vandenbergen 2003; Cuenca 2008) – Colbert is appealing to ideology, while Dickerson is appealing to pragmatism and reaches the same conclusion. The "and" again signals Dickerson's intention to continue (e.g. Schiffrin 1986), and may help to 'smooth' (Turk 2004) the jump he is trying to make back to Colbert's point about the rarity of billionaires (lines 1-2). An interactive gesture accompanying this sequence may co-express the function of any of the four lexical discourse markers, or may profile an entirely different aspect of the contribution.[6]

In (119), Colbert and American actor Jennifer Aniston are discussing the time that their friend, Tig Notaro, and Colbert tried to text Aniston while on live television. Colbert asks Aniston if she knew what was happening at the time. After a series of overlapping turns (lines 1-4), Aniston begins to explain what she thought at the time. As she begins the next

[6]The gestures used during this particular segment are discussed in Chapter 4, Section 4.

line in her recounting ("and then– "), she restarts her sentence twice, and eventually provides background information as to just how crazy the text message was (line 6).

(119) Transcript 13: Jennifer Aniston
[UID:eafcb39a-1343-11ea-a550-089e01ba0335,1824]

```
1  SC  Did you have any idea what was going on?  [What did you-]
2  JA                                            [I sure did not]
3  SC  What did you think when that- [when that came through?]
4  JA                                   [I thought]
5      Tig was having a moment.  And then- I also- 'cause
6      by the way, not much of that was spelled correctly
```

As usual, BTW signals Aniston's intent to digress momentarily from her story, in this case to give background information that may help her addressee better understand her reaction. The immediately preceding "'cause" signals that the contribution will offer some kind of explanation.

The occurrence of additional lexical discourse markers adjacent to BTW is summarized in Table 6.5. Given the relative infrequency of adjacent lexical discourse markers in this data set, it is difficult to draw any meaningful conclusions from the distribution. However, there are two things worth noting. First, "and" occurs in more use-contexts than any other accompanying discourse marker, aligning well with Schiffrin's (1986) claim that "and" is a kind of default connective, signalling the continuation of an action, whatever that action may be. Second, five different additional lexical discourse markers occur in background contexts, suggesting heterogeneity in discourse-structural features.

	Background	Elaboration	Evaluation	Meta	Shift	Joke	**Total**
and/also	1	4	1	-	3	-	9
because	1	-	-	-	-	-	1
coincidentally	1	-	-	-	-	-	1
oh	-	1	-	-	-	-	1
um	2	-	-	-	1	-	3
well	1	-	-	-	-	-	1
Total	6	5	1	0	4	0	16

Table 6.5: Lexical discourse markers accompanying BTW

Given the principle of *independence of contribution*, we expect the presence of other lexical discourse markers to effect gestural expression. The use of multiple discourse markers reflects complexities in the underlying discourse structure, and each marker can profile a different aspect of that structure. This means that a gesture accompanying, for example, "and by the way" may co-express the continuation indicated by "and", the digression signalled by

"by the way", or some other aspect of the structure that is not overtly expressed verbally. This means that when we consider variation in gestures accompanying BTW, we must also consider the presence of adjacent lexical discourse markers.

6.4 Variation in the gestural mode

Of the 100 instances of BTW annotated, 84 occurred with manual gestures. Of those 84 manual gestures, 77 were judged as performing interactive functions.[7] The remaining 7 manual gestures were judged as representational gestures, performing a primarily semantic rather than discourse management function. These 7 gestures are not considered for further analysis. The occurrence of gesture with BTW in the present data set is summarized in Table 6.6.

	Interactive	Representational	No manual gesture	**Total**
Interview	50/68 (74%)	6/68 (9%)	12/68 (18%)	68/68 (100%)
Monologue	27/32 (84%)	1/32 (3%)	4/32 (13%)	32/32 (100%)
Total	77/100 (77%)	7/100 (7%)	16/100 (16%)	100/100 (100%)

Table 6.6: Recurrent gestures aligned with BTW

The 77 interactive gestures assessed varied in form and timing, but did so systematically, with variants corresponding to structural and functional differences in the signalled digression. I argue that this systematicity indicates that interactive gestures contribute directly to discourse coherence, just as *spoken* discourse markers do. Given this, *gestural* discourse markers can be used to better understand the nature of digression and variations in its function across discourse contexts.

Each gesture example is presented as a transcript, followed by a series of relevant screenshots, each co-indexed with particular points in the transcript. Following Embodied Conversation Analysis conventions (Mondada 2018), each example is presented as a two- or three-tiered transcript, with a tier for speech, a tier for manual gesture, and, where relevant, a tier for postural shifts. The duration of manual gestures is denoted by asterisks

[7]Gestures were marked as interactive *unless* a clear semantic lexical affiliate was identified in the accompanying speech.

aligned across the two tiers.[8] Postural shifts are denoted by delta markers (△). For clarity, only relevant gestures are annotated. Additional un-annotated transcript is given to provide sufficient discourse context.

In Section 6.4.1, I discuss variation in the timing of interactive gestures relative to the clausal position of BTW. Section 6.4.2 summarizes the observed formal variation of gestures accompanying BTW based on the five action schematic gesture classes defined in Chapter 2 (Presentation, Removal, Stopping, Referring, and Engagement). Sections 6.4.3-6.4.4 discuss the four gesture classes that appear recurrently in the present data set, beginning with the two classes that I argue serve as gestural digressive markers (*stopping* and *disengagement*), followed by the remaining two more multifunctional classes (*presentation* and *referring*).

6.4.1 Temporal alignment with BTW

Each manual gesture was coded for its alignment with BTW. The possible values were *BTW only*, *digression only*, or *both*. I argue that these patterns of alignment can help us better understand the capacity of interactive gestures to contribute discourse structural information independent of spoken discourse markers. If, for example, gestural digressive markers consistently align with BTW and only BTW, we may argue that the gesture is co-expressing the function of BTW, and is, at most, *adding* information to our interpretation of the spoken discourse marker. If, on the other hand, gestural digressive markers regularly align with the entire digression, we can argue that the gesture is independently expressing information about the underlying digressive structure, rather than affiliating only with the lexical expression. What we actually find is a mixture, summarized in Table 6.7.

	BTW only	Digression only	Both	Total
Interview	12/50 (24%)	14/50 (28%)	24/50 (48%)	50/50 (100%)
Monologue	9/27 (33%)	7/27 (26%)	11/27 (41%)	27/27 (100%)
Total	21 (27%)	21 (27%)	35 (45%)	77/77 (100%)

Table 6.7: Temporal alignment of gestural discourse markers with lexically-signaled digressions

[8]Following the standard analysis of gesture structure (e.g. Kendon 2004), "gesture duration" includes preparation, strokes, and pre-/post-stroke holds. Gesture phases are not distinguished in transcription as only overall gesture alignment is relevant to the current research question.

To demonstrate what these alignments look like, and how they can provide evidence for the expressive capacity of the gestural discourse marker, a detailed example is given for each type of alignment.

We have seen the example in (120) once before in Section 6.3.3. In the excerpt, John Dickerson (American journalist) and Stephen Colbert are discussing the relationship between extreme wealth and the rhetoric around the economy regularly espoused by the Democratic party. Dickerson begins a digression to elaborate on an earlier sub-question regarding the number of billionaires. As he marks the beginning of the digression, he points toward Colbert's cue card that is on the desk to his right (Fig.6.1, G2). During the digression itself, Dickerson performs two more interactive gestures, first bringing a pinched hand down to the desk as if hold down an object (G3), and then flattening his hand to perform a small lateral sweep right (G4). Colbert then begins to negotiate the end of the digression through a lexical backchannel ("right", line 12), overlapping with Dickerson's summation marker "so" (line 11).[9] As Dickerson ends the digression, summarizing its purpose, Colbert begins to ask a follow-up question, which returns the discourse to the main topic under discussion.

(120) Transcript 14: John Dickerson
[UID:465f22fa-01fa-11ea-8c13-089e01ba0335]

```
1   SC   There are only, according to Fortune Magazine, 607
2        billionaires in the world.  Why do we care what they think?
3   JD   and-
4   SC   And because there's a lot more of us.  And, they've got
5        too much money.  Give your money to people who need it.
6        And collective- collective distribution of you wealth is
7        not anti-American, it's as American as Eisenhower
8   JD   uh *well* *and by the way* of those number of billionaires
           *G1-* *G2- - - - - - *
9        *there probably aren't many in the caucus states, in Iowa*
         *G3- - - - - - - - - - - - - - - - - - - - - - - - - - - *
10       *and in New Hampshire*
         *G4 - - - - - - - - - *
11       *[so* they're] really even less [important]
         *G5*
12 SC    [right] (.)                    [why is] it a bad idea to
13       attack billionaires, I suppose
G1:  containment gesture toward Colbert, lowers to desk
G2:  finger point to lower right
G3:  pinch shape, press down onto desk
G4:  flat hand, lateral sweep right, press down onto desk
```

[9]Overlapping speech is demarcated with brackets, again following Embodied Conversation Analysis conventions.

G5: containment gesture toward Colbert

Figure 6.1: Gesture sequence corresponding to TRANSCRIPT 14: JOHN DICKERSON, ex. 120.

Dickerson performs interactive gestures throughout the digression, spatially organizing topics as metaphoric objects on the desk in front of him. However, because there is a distinct interactive gesture that aligns only with the lexical discourse markers "and by the way" (G2), it is this gesture that is considered for its status as a digressive marker, and it is this alignment that is reported. As I argue in detail in Section 6.4.4.1, this pointing gesture, oriented toward a particular region within the Interaction Space, signals the *background* relation underlying this particular digression. By *locating* a topic that is present in the Interaction Space as a metaphoric object,[10] Dickerson is directing attention toward the topic he intends to contribute to during his digression, namely *the number of billionaires*.

In (121), Elizabeth Banks (American actor) is telling Colbert about acting in the Christmas film *Fred Claus* (2007), a turning point in her career that inspired her to search for more 'serious' roles. Upon realizing that her statements may make her sound ungrateful, Banks self-corrects, and clarifies that working on the film was "very fun" (line 1). As she says this, she reaches toward Colbert with a flat down-turned palm, leaning her head to the side and scrunching her eyebrows in a face-saving gesture, as if to say "don't get me wrong" (Fig. 6.2, G1). To further avoid judgement, Banks goes on to specify a part of the job she enjoyed using an evaluative digression (line 1-2). Throughout this digression, Banks is turned away from Colbert, toward the audience. She performs two consecutive *clearing* gestures, with her palms facing outward in a *blocking* orientation, as if to keep objects from entering her immediate space (G2 & G3).

(121) TRANSCRIPT 15: ELIZABETH BANKS
[UID:f1dc45d4-0067-11ea-ac64-089e01ba0335,1888]

1 EB *it was very fun.* by the way *the North Pole set was*
*G1- - - - - - - - * *G2- - - - - - - - - - *
2 *one of the greatest sets I've ever* worked on
*G3- - - - - - - - - - - - - - - - - *

[10]In this case, the topic as a metaphoric object is metonymically associated with Colbert's physical cue card, which has the number of billionaires written on it.

```
G1:  palm down reach toward Colbert
G2:  two hands, palm out, sweep up and out
G3:  two hands, palm out, sweep out (abbreviated repetition of G2)
```

Figure 6.2: Gesture sequence corresponding to TRANSCRIPT 15: ELIZABETH BANKS, ex. 121.

The lexical digressive marker BTW is uttered *between* two distinct interactive gesture sequences, as Banks retracts her right hand toward her lap and subsequently prepares for her next gesture. Because there is no gesture performed *during* BTW, the first gesture to occur during the host clause ("the North Pole set was one of the greatest sets I've ever worked on") was considered for its status as a gestural digressive marker. Both the outward facing blocking orientation of Banks' manual gesture and Banks' turn away from Colbert are considered gestural digressive markers, discussed in Sections 6.4.3.1 and 6.4.6 respectively.

In (122), Eddie Redmayne (British actor), is discussing a mishap that happened on the set of his recent film. During this, Redmayne goes on a brief elaborating digression to reassure his audience that the actor was not actually hurt during the mishap. Beginning with the underspecified discourse connective *and*[11] (line 1), Redmayne holds an upward pointing index finger as if to say "wait a second", directing the gesture first toward Colbert (G1a) then sweeping across his body twice to direct the gesture toward the audience (G1b-c).[12] Colbert interrupts this digression to say "no spoilers", while performing a one-handed blocking gesture, holding an outward facing palm toward Redmayne, as if to physically stop an incoming object. The functional connection between this blocking gesture and that described in the previous example will be further explored in Section 6.4.3.1. In response, Redmayne apologizes, bringing his hand up to his forehead in a gesture of embarrassment (line 3, G3). Redmayne then returns to the main discourse, turning his head back toward Colbert and raising an upturned open hand, as if to present an object for inspection (G4).

[11]For an in depth discussion of the range of discourse structural uses of *and* see Schiffrin (1986).

[12]This gesture is analyzed as a single gesture (G1) with three strokes (a-c) rather than as three distinct gestures. This analysis strays from more traditional analyses which hold that a single gesture can only have one stroke (e.g. Kendon 2004). However, I argue that the consistent handshape, fluidity of movement across changing trajectories, and shared discourse function justifies the non-standard analysis presented here.

(122) TRANSCRIPT 16: EDDIE REDMAYNE
[UID:bf03ea3c-1731-11ea-a000-089e01ba0335,2085]

```
1  ER   *a lot of the film was then shot on green screen but we had*
        *G1- - - - - - - - - - - - - - - - - - - - - - - - - - - - -*
2       *and she was totally fine/ by the way no Felicities /were
        *G2a- - - - - - - - - - -/b - - - - - - - - - - - - - /c ->
3       hurt in the making of [this movie* *um ]      sorry*
        <- - - - - - - - - - - - - - - - - * *G3- - - - - - -*
4  SC                          [no spoilers no spoilers]
5  ER   uh *but it meant we had the* um we had the memory of that
           *G4- - - - - - - - - - -*
G1:  loose containment gesture in central gesture space
G2:  index finger pointed up, held toward Colbert, 2x lateral sweeps
G3:  left hand to forehead expressing embarrassment
G4:  loose palm up open hand toward Colbert
```

Figure 6.3: Gesture sequence corresponding to TRANSCRIPT 16: EDDIE REDMAYNE, ex. 122.

In this case, the gesture occurring with the lexical digressive marker BTW is held for the full duration of the digression, which both precedes and follows the lexical marker. In all three examples discussed, both the gestural and spoken digressive markers can be said to 'take scope' over the digression. By this I mean that digressive markers, regardless of their position, signal the digression in its entirety as a discourse unit. Because of the temporal

and linear limitations of speech, the scope of BTW is abstract – BTW cannot co-occur with the entirety of the digression that it signals. Instead, interlocutors must use other contextual cues to deduce the end of the digression and the return to the main discourse. With gestural discourse markers, the scope of the digression can be fully embodied – the gestural digressive marker can co-occur with the entire digression because they are expressed in different modes. The end of the digression is marked by the end of the gesture, and the return to the main discourse is marked by the return of Redmayne's attention to Colbert.

The variation in temporal alignment of interactive gesture with BTW is partially accounted for by considering the clausal position of BTW. As demonstrated in Table 6.8, interactive gestures that align with BTW only, and not the digression itself, occur when BTW is uttered at the beginning or middle of the digression. When BTW is uttered at the *end* of the digression, the interactive gesture always occurs during the digression itself.

Position	BTW only	Digression only	Both	Total
Beginning	19/40 (48%)	8/40 (20%)	13/40 (33%)	40/40 (100%)
Middle	2/14 (14%)	2/14 (14%)	10/14 (75%)	14/14 (100%)
End	- -	11/23 (48%)	12/23 (52%)	23/23 (100%)
Total	21/77 (27%)	21/77 (27%)	35/77 (45%)	77/77 (100%)

Table 6.8: Temporal alignment of interactive gesture with signaled digression

This latter point is particularly striking as it indicates the capacity for gesture to bear the burden of maintaining discourse coherence in the absence of a spoken correlate. In these cases it is the *gesture* that serves as the first overt cue for the presence of a digression, only to be co-expressed later by the lexical discourse marker.[13]

This section has shown the ways in which gestural discourse markers are subject to temporal variation relative to the discourse structure, just as lexical discourse markers are. However, there are key differences between the observed verbal and gestural alignment variations that reflect the differing affordances allowed to the modalities. Most obviously, lexical markers are linearly bound, and cannot co-occur with the discourse segment they are take scope over. Gestural markers, on the other hand, can physically manifest the scope of the

[13]These results align well with Harrison's (2010) study of gestures occurring with negation, which they argued occur as early as is non-contradictory in order to "prepare" the listener for the negation.

discourse segment, as seen in examples (121) and (122). This is a powerful expressive tool because it can serve to disambiguate the beginning and end of a digressive segment without having to resort to deductive analysis based on implicit relations identified by the addressee (or researcher). Finally, the timing of the lexical and gestural discourse marker appear to be relatively independent. This further suggests that the interactive gesture is conveying discourse structural meaning independently from the co-occurring lexical discourse marker.

6.4.2 Formal variation in manual gestures

In this section, I discuss the observed formal variation in gestures accompanying BTW-marked digressions. In doing so, I identify two classes of gestural digressive markers; one manual, which I call *stopping* gestures, and one postural, which I call *disengagement*. In the present data, *stopping* gestures take two forms: (i) *blocking* in which the speaker holds up a flat out-turned palm as if to stop an object from entering their immediate body space, and (ii) *pausing*, an emblematic gesture paraphrasable as 'wait a second', in which the speaker holds an upward pointing index finger toward their addressee. *Disengagement* gestures involve head and body movements away from the Interaction Space and are used to signal disengagement from the current discourse state. These two sets of gestural digressive markers will be discussed in Sections 6.4.3 and 6.4.6 respectively. Three other classes of hand gestures (*presentation, locating,* and *addressing*) also occur recurrently in the data set. I argue that though these are performing discourse management functions, they do not serve as gestural digressive markers. Their independent contributions to discourse management are discussed in Sections 6.4.4 & 6.4.5. Examples of the four recurrent hand gesture classes observed with BTW are given in Figure 6.4.

Figure 6.4: Four classes of hand gestures associated with BTW

Of the 77 interactive gestures in the data set, 67 were classified as belonging to one of these four gestural classes, as summarized in Table 6.9. Of the remaining 10, 2 were *removal* gestures[14], and 8 exhibited formal features that did not justify the inclusion in any of the action schematic classes defined in Chapter 2. For example, there were two instances in which

[14]See Chapters 2-5 for extensive discussion of the form and function of *removal* gestures.

the speaker brings their hands forcefully down to the desk. Though these seem to convey interactive meaning associated with emphasis, they do not exhibit the physical affordances of any of the four defined action schema (PRESENT, REFER, REMOVE, STOP).[15]

	Presentation	Locating	Addressing	Pausing	Blocking	**Total**
Monologue	3/32 (9%)	2/32 (6%)	1/32 (3%)	10/32 (31%)	8/32 (25%)	**24/32** (75%)
Interview	13/68 (19%)	9/68 (13%)	11/68 (16%)	1/68 (1%)	9/68 (13%)	**43/68** (63%)
Total	**16/100** (16%)	**11/100** (11%)	**12/100** (12%)	**11/100** (11%)	**17/100** (17%)	**67/100** (67%)

Table 6.9: Gestures accompanying BTW by interaction type

I will argue here that *blocking* and *pausing* gestures serve as digressive markers by metaphorically *stopping* the main discourse. Blocking gestures do so metaphorically, representing a barrier to physically stop incoming actions and objects. Pausing gestures may be said to do so metonymically, such that the upward pointing index finger represents the numeral one, evoking a "wait one moment" request. Additionally, locating gestures that are directed outside of the primary Interaction Space between the Colbert and his guest are also considered digressive markers, metaphorically signalling the utterance as non-central to the current discourse. Addressing gestures, on the other hand, are *not* digressive markers, as such, but rather perform parallel discourse management.

Each of these four gesture classes occur with different digression types at different rates, suggesting that the gesture is expressing some aspect of the digression beyond that expressed by BTW. The occurrence of each gesture class with each use-context of BTW is summarized in Table 6.10. There is significant variation in gesture performance in all use-contexts, but there are also clearly identifiable trends.

[15]This is not to say that the unclassified gestures don't evoke *any* action schema. They surely do, just not ones defined or pursued in this work.

	Presenting	Locating	Addressing	Pausing	Blocking	**Total**
Background	3/23 (13%)	4/23 (17%)	1/23 (4%)	1/23 (4%)	3/23 (13%)	12/23 (52%)
Elaboration	2/16 (13%)	3/16 (19%)	1/16 (6%)	**4/16 (25%)**	1/16 (6%)	11/16 (69%)
Evaluation	3/17 (18%)	1/17 (6%)	1/17 (6%)	1/17 (6%)	**7/17 (41%)**	13/17 (76%)
Meta	2/14 (14%)	1/14 (7%)	**5/14 (36%)**	- -	3/14 (21%)	11/14 (79%)
Shift	**6/15 (40%)**	1/15 (7%)	4/15 (27%)	- -	- -	11/15 (73%)
Joke	- -	1/14 (7%)	- -	5/14 (36%)	3/14 (21%)	9/14 (64%)

Table 6.10: Gestures accompanying BTW by discourse type

If we consider the recurrence of each gesture class in different discourse-structural contexts, we find that there are clear *associations*. A given gesture class is considered *associated with* a given use-context if the following two conditions hold:

A. Gesture X is the most frequent gesture observed in context Y

B. The rate at which gesture X occurs is context Y is higher than the rate at which gesture X occurs in any other context

For example, evaluation digressions are most frequently accompanied by blocking gestures *and* the rate at which blocking gestures appear with evaluation digressions is higher than the rate at which blocking gestures appear in any of the other defined use-contexts. This means that blocking gestures are considered to be *associated with* evaluation digressions. The form and associated use-contexts for each gestural class are summarized in Table 6.11.

There are two further things worth noting about the distributions of gesture classes in the present data set. First, background digressions are not associated with any one gesture class, occurring at similar rates with *presenting, locating* and *blocking* gestures. Background digressions are also the least frequently accompanied by manual interactive gestures of all defined use-contexts. I argue the relative infrequency and variability of interactive gestures with background digressions suggests that background digressions are minimally disruptive – they are not regularly marked because they do not pose a significant threat to discourse coherence, and thus do not *need* to be regularly marked.

Second, *locating* gestures are not associated with any one use-context, and are the only gesture class to occur in *all* defined use-contexts. This makes sense given the formal and functional variability of locating gestures. As abstract deictic gestures (e.g. McNeill et al. 1993), locating gestures are used to direct joint attention toward some region of space within or outside of the Interaction Space. In section 6.4.4.1, I discuss the ways in which variation in the trajectory of locating gestures reflects functional difference in the marked digressions.

Class	Form	Relation
Presenting	gesture as if to present an object	Shift
Locating	deictic gesture toward (metaphoric) referent	–
Addressing	deictic gesture toward addressee	Meta-discursive
Pausing	index-finger pointed upward, fingers closed	Elaboration
Blocking	flat open palm, turned away from speaker's body	Evaluation

Table 6.11: Forms and functions of gesture classes associated with BTW

In the remainder of this section, I discuss each of these four classes and the ways in which they systematically reflect properties of the discourse structure in concert with, but independently from, the lexical discourse marker BTW. I will focus most attention on *blocking* (6.4.3.1), *pausing* (6.4.3.2), and and a subset of *locating* (6.4.4.1) gestures in their capacity as digressive markers. I then return briefly to addressing gestures (6.4.4.2) to discuss their status as *non-digressive* gestural discourse markers. The section concludes with a discussion of non-manual postural shifts during digression (6.4.6).

6.4.3 Stopping

Stopping gestures are gestures that exhibit the physical affordances of stopping an object from moving closer to the speaker's body. These gestures occur in slightly over one-quarter of the present data set (28/100). I identify two functionally distinct types of stopping gestures: *blocking* and *pausing*. *Blocking* gestures are performed by holding up an out-turned open palm in front of the body, as if to stop a medium-to-large object from entering the immediate bodily space. In the present data set, these gestures are associated with evaluation digressions and frequently compose with a CLEARING action schema to produce a semiotically complex gesture. *Pausing* gestures are performed by holding an index-finger point, directed upward, with the palm facing out, as if to ask the addressee to "wait a moment". This gesture is also compositional, combining the iconic palm-away orientation of blocking gestures, with an emblematic hand shape signalling the numeral '1'. I discuss the function of each in turn.

6.4.3.1 Blocking

Blocking gestures occur when one or both hands are held up with a flat open palm facing outward toward the addressee, as if to stop an object from entering the immediate personal space. These gestures are well-documented in the gesture literature as performing discourse management functions associated with stopping action, be it an utterance, a line of inquiry, or a potential undesired inference.[16] In line with Wehling's (2017) work on discourse management gestures in political debate, I use the term 'blocking' to evoke the force dynamics central to the gesture's meaning and function: the open outward facing palm enacts a physical barrier that metaphorically stops the main discourse from moving forward.

In this section, I present one close analysis of a discourse excerpt with Nancy Pelosi to fully demonstrate the function of a blocking gesture in context. I will then discuss the compositionality of blocking gestures observed more generally in the data set, and their capacity to profile multiple aspects of the utterance, namely (i) its digressive status, and (ii) its evaluative contribution.

In (123), Nancy Pelosi (American politician) is describing recent legislation passed by the Democratic party. After completing a list of successes, she goes on a brief evaluative digression, claiming that these pieces of legislation were not only successful, but also "popular" (lines 2-3). She then returns to the main topic of discussing the Democratic party's goals, marking the return with the discourse marker "so" (line 3-4).

(123) Transcript 17: Nancy Pelosi
[UID:6a02fce6-159f-11ea-8a36-089e01ba0335,2870]

```
1  NP   just to name a few things that we have sent over *(.)  but
                                                        *G1- - ->
2       the list goes on and on* *and by the way* *all of them*
        <- - - - - - - - - - - - * *G2 - - - - - * *G3 - - - - *
3       *very popular* in the public.  So we weren't looking for a
        *G4- - - - - *
4       fight we were looking for bringing people together which I
5       think is our responsibility
G1:  right hand, palm up cyclic lateral sweep right
G2:  two hands, palm away, small push forward
G3:  containment gesture toward Colbert, lowers to desk
G4:  right hand, palm away, arching sweep right
```

[16] See, for example, Kendon's work on 'Vertical Palm (VP)' gestures in Southern Italian speakers (2004, chapter 13), and Bressem & Müller's work on 'holding away' gestures (2014; 2017) in German speakers, and Wehling's (2017) work on 'blocking' gestures in American political debates.

Figure 6.5: Gesture sequence corresponding to TRANSCRIPT 17: NANCY PELOSI, ex. 123.

Just prior to the digression, Pelosi performs a cyclic gesture, her right hand rotating in small circles while moving out and to her right. This is a recurrent gestural representation of repetition (Ladewig 2011, 2014a), semantically paralleling the expression "goes on and on" in her speech (line 2, G1). Pelosi then performs a two-hand blocking gesture aligned with the lexical digressive marker "and by the way" (line 2, G2). While holding the blocking orientation, both hands perform two small emphatic forward movements at"and" and "by".[17] Pelosi then turns toward Colbert, performing a containment gesture as if to locate "all of them" on a region of the desk in front of her (line 2, G3). For the evaluation itself, claiming the legislation to be "very popular", Pelosi performs a sweeping gesture with a loosely outward facing palm (line 2, G4). The clear blocking gesture directed toward the audience at BTW may perform two simultaneous interactive functions, momentarily stopping the main discourse topic and stopping any inference on behalf of the audience that what the Democratic party was doing didn't matter or was ill-received.

In addition to the characteristic hand shape and orientation, 5 of the 14 blocking gestures in the present data set include outward lateral movement, such as flicks (lateral movement of the hand, with flexion at the wrist), sweeps (lateral movement of the arm, with flexion at the elbow or shoulder) and shaking. I argue here that these gestures are *compositional* – the hand shape and orientation stop a line of action, whereas the movement expresses some other aspect of the communicated message. We have already seen two examples of blocking gestures above that exhibit this compositionality that are worth revisiting here. In the first (ex. 121), actor Elizabeth Banks performs a two-handed blocking gesture for the entire duration of an evaluative digression. While maintaining the blocking hand shape and orientation, she performs two lateral sweeps outward, as if to clear a vertical surface. In this compositional gesture, the blocking shape and orientation serves to temporarily stop the main discourse topic and, possibly, an inference that Banks is being ungrateful. The outward lateral movements express the superlative evaluation in "one of the *best* sets".

In the second previously seen blocking gesture (ex. 122), Colbert performs a one-hand blocking gesture, holding an outward facing palm toward actor Eddie Redmayne, while

[17]These small movements, called 'beats', are known to align with prosodic accents and primarily express emphasis (e.g. Leonard & Cummins 2011; Loehr 2012).

saying "no spoilers", shaking both his head and his outstretched hand as he does so. In this compositional gesture, the blocking gesture represents Colbert's request to stop Redmayne's speech and his current line of inquiry, evoking the metaphor SPEAKING IS FORWARD MOTION (e.g. Wehling 2017). The lateral shaking of both Colbert's head and hand co-expresses the spoken negation.[18]

Three further examples are given in (124) to demonstrate formal variation in open-hand gestures accompanying evaluative digressions. In (124a-b) we see two blocking gestures. The first shows Andrea Savage (American actor) performing a very similar gesture to the Banks' blocking gesture described above – Savage moves both hands, palms facing outward, up to chest height then outward in time with the evaluation (and pun) "it crushed". The second shows Colbert performing a one hand blocking gesture, flicking his hand to the left while describing "amazing" hypothetical ratings. The third (124c), shows Colbert performing a similar lateral movement, with a similar open hand shape, but a distinct orientation, with his palms turned downward toward the desk while claiming a show "went great". Note that in all three cases the *blocking* gestures is aligned with the evaluative digression, *not* BTW.

(124) a. TRANSCRIPT 18: ANDREA SAVAGE
[UID:67f2c08a-f628-11e9-920c-089e01ba0335,3371]

```
1  AS   he leaps into the glass and throws glass in his face,
2       by the way *it crushed*
                   *G1- - - -*
G1:  two hands, palm out, lateral sweep out
```

b. TRANSCRIPT 19: STEPHEN COLBERT
[UID:2d0e822a-a46f-11e9-80c4-089e01ba0335,1065]

```
1  SC   *the ratings would be amazing* by the way
        *G2- - - - - - - - - - - - -*
G2:  left hand, palm out, flick left
```

c. TRANSCRIPT 20: STEPHEN COLBERT
[UID:135ad8d4-f496-11e9-9c8d-089e01ba0770,1226]

```
1  SC   I'm already watching the show on DVR. *It went great*
                                              *G3- - - - - -*
2       by the way
G3:  both hands, palm down, small lateral sweep out
```

[18]This migration of a gesture from one body part (the shaking of the head) to another (the shaking of the hand) is understudied, but see Debras (2017) for a discussion of this phenomenon in 'shrugs'.

Figure 6.6: Gesture sequence corresponding to TRANSCRIPTS 18-20, ex. 124

Work specifically on palm-down gestures like that in (124c), associate lateral sweeps with negation and intensification (Harrison 2010; Kendon 2004), connecting the two functions by arguing that the lateral sweep, when accompanying a positive assessment, is negating *alternatives*, as if to say "this is nothing other than great".[19] However, these studies fail to engage seriously with the possibility of the lateral sweep combining compositionally with other gestural features. This step is crucial for formalizing the contribution of interactive gesture to discourse management – each gestural feature can contribute independent action-schematic meaning. In the examples above, the palm out orientation stops a line of inquiry by metaphorically stopping forward movement, while the lateral sweep, regardless of palm orientation, negates alternatives, by metaphorically clearing a surface of objects.

6.4.3.2 Pausing

Pausing gestures are a sub-type of *stopping* gesture, performed by holding an index-finger pointing upward, as if to ask the addressee to "wait a second". These gestures exhibit the same palm orientation as *blocking* gestures, partially evoking the same BLOCKING action schema, and compose with the conventional numeral '1' hand shape. These gestures occur 11 times in the data set, almost exclusively in two use-contexts: elaborative digressions (4 instances) and Colbert's idiosyncratic joke construction (5 instances). All but one example occur during monologues, suggesting that, at least in the context of BTW-marked digressions, *pausing* gestures may be a quirk of Colbert's performance style. The one example of *pausing* gesture not performed by Colbert in the present data set was already discussed in Section 6.4.1. I will present one additional example here.

In (125), Colbert is making a joke about Donald Trump claiming that *re*-Tweeting something offensive is okay because he wasn't the one to originally write it. I discussed this excerpt as an example of a meta-aware elaboration digression in Section 6.3.1. In the verbal mode, Colbert marks this elaboration digression with the lexical discourse markers "and by

[19] Work on similar gestures in German speakers (Bressem & Müller 2014, 2017), claims that these gestures are associated with negative assessment, grounding the gesture in metaphoric removal of unwanted objects from the speaker's immediate personal space. This is clearly not the function of the gestures described here where the lateral movement is recurrently associated with *positive* assessment.

the way", where "and" is signalling the elaboration relation, and "by the way" is signalling the status of the contribution as digressive. In the gestural mode, Colbert performs a pausing gesture directly aligned with the lexical discourse marker sequence (Fig. 6.7, G2). As he does so he raises his eyebrows dramatically, a facial gesture typically associated with emphasis (e.g. Flecha-García 2010).

(125) TRANSCRIPT 21: STEPHEN COLBERT
[UID:a5309d94-be5d-11e9-a8ce-089e01ba0335,387]

```
1   SC    what I didn't stab you that was somebody else's knife
2         *I restabbed you (.)* *and by the way* *the guy whose knife
          *G1- - - - - - - - - -* *G2 - - - - - * *G3 - - - - - - - ->
3         I used* *a very respected* stabber
          <- - -* *G4- - - - - - -*
4         (piano sound)
G1:  right hand fist, stabbing motion; left extended down, palm up
G2:  right hand fist; left hand index finger point up
G3:  right hand fist; left hand index finger point back
G4:  right hand fist; left hand index finger point to audience
```

Figure 6.7: Gesture sequence corresponding to TRANSCRIPT 21: STEPHEN COLBERT, ex. 125.

After performing the *pausing* gesture, Colbert maintains the index-finger point hand shape for two more gestures, pointing backward toward an imagined "stabber" (G3), and then toward the camera in an *addressing* gesture (G4). This is a good demonstration of the use of a gesture 'catchment' (McNeill et al. 2001). The initial pausing gesture serves to signal a forthcoming digression, and the hand shape of this marker is maintained through the digression, in a sense 'taking scope' over the digression, as was discussed in the Redmayne example above.

6.4.4 Referring

Referring gestures are deictic gestures that direct joint attention to a region of space relative to the Interaction Space, and do not exhibit the physical affordances compatible with the PRESENT or REMOVE action schema. This class of gesture has two functionally distinct sub-types – *locating* gestures which direct joint attention toward a topic as a metaphoric object through abstract deixis (McNeill et al. (1993)), and *addressing* gestures which direct joint attention toward the addressee in order to elicit a response (e.g. Bavelas et al. (1992); Bavelas (1994)). Referring gestures occur in approximately one quarter of the present data set (23/100) and occur in a diverse set of use-contexts with BTW. I will look at each sub-type in turn, but focus more closely on exploring the observed functional variation in locating gestures.

6.4.4.1 Locating

The *locating* class of gestures consist of deictic gestures, usually one-finger points, toward empty space. These gestures can be considered examples of *abstract deixis* which are used to negotiate the location of topics as metaphoric objects in shared space (Azar & Özyürek 2015; McNeill et al. 1993). These gestures are extraordinarily common and multi-functional, organizing topics along different spatial parameters to express particular discourse relations such as sequences (e.g. Calbris 2008) or contrasts (Hinnell 2019). Given this multifunctionality, it is unsurprising that locating gestures, as a class, do not appear to be associated with a particular use-context in the present data set. However, more careful inspection of particular locating gestures in particular use contexts reveal functionally motivated formal variation.

In this section I present two examples of locating gestures, one performed with an elaboration and one performed with a background digression. These two gestures differ primarily in their trajectory relative to the Interaction Space – the point is *away* from the Interaction Space in the elaboration context, and *into* the Interaction Space in the background context. I argue that the first formal variant, locating gestures *away* from the Interaction Space, can function as digressive markers.

In (126), Susan Rice (American diplomat) is discussing the reorganizing of government jobs under the Trump administration. She contrasts the reported disorder with what she did in her tenure as National Security Advisor under the Obama administration. Rice performs an elaboration digression (lines 3-4) highlighting that it was her that made the decisions, not Obama, before continuing the main discourse by explaining how she reorganized government positions. This is digression is minimally disruptive, as it is brief and transparently related to ongoing discourse topics. This particular elaboration digression also seems to serve a social function. Rice appears to be proud of her actions and wants her addressee to know that she was the one responsible for them. However, asserting this in the main discourse may come off as bragging. An effective solution then is to embed the information in a digression such that the desired information is still conveyed, but is done so *as if* it were not particularly

important. During the digression Rice performs a loose index-finger point, moving her hand out from the center of the Interaction Space to its periphery.

(126) TRANSCRIPT 22: SUSAN RICE
[UID:c1d66448-e995-11e9-9e68-089e01ba0770,2769]

```
1  SR   it's one thing *to rationally and* *wisely reallocate staff
                       *G1- - - - - - - -* *G2- - - - - - - - - ->
2       and* do a a thoughtful downsizing.  that's what I did under
        <- *
3       the uh Obama admini*stration by the way not directed by the
                          *G3 - - - - - - - - - - - - - - - - - ->
4       president of the United States* *b- I did it myself because*
        <- - - - - - - - - - - - - - -* *G4 - - - - - - - - - - -*
5       I thought we could rebalance some jobs here, put more there,
6       take some from here
G1:  containment gesture toward Colbert
G2:  palms center, alternating up and down movement
G3:  loose index finger point, arching sweep right
G4:  loose hand, palm down, small push downward
```

Figure 6.8: Gesture sequence corresponding to TRANSCRIPT 22: SUSAN RICE, ex. 126.

Rice begins this sequence by metaphorically presenting a discourse topic (*the reallocation of staff*) to Colbert, bringing her hands into the central interaction space as if to hold an object up for inspection (line 1, G1). Rice then performs a representational gesture aligned with "reallocate", moving her hands up and down in alternation, iconically representing the shuffling of positions (line 1, G2). Rice then retracts her hands to a rest position in her lap until performing the digression, at which point she points loosely toward empty space to her right (line 3, G3). As she returns to the main topic to describe what she actually did as National Security Advisor, her hand returns and once again indicates a location in the central interaction space (line 4, G4). The locating gesture metaphorically separates the information presented in the digression from information conveyed in pursuit of ongoing discourse goals by locating the two in different regions of space. Importantly, the information that is most

relevant to the ongoing discourse topic is located centrally between her and Colbert, whereas the digressive information is located in a relatively 'inaccessible' position outside of shared space. In this way, this locating gesture marks the information in the digression as not immediately relevant to the ongoing discourse.

The second example demonstrates the use of a locating gesture during a background digression. In contrast with the previous example, the locating gesture in this case indicates a space directly in the center of the Interaction Space. This example was seen once before in Section 6.3.1.

In (127), Colbert is asking Elizabeth Warren (American politician) to participate in a kind of rapid-fire interview in which Colbert presents a policy issue and Warren must respond briefly and as quickly as possible. Before actually beginning the game, Colbert asks a clarifying question about Warren's proposed 'wealth tax' on the ultra-wealthy (line 2). This question is asked, presumably, for the benefit of the audience – Colbert temporarily digresses from the game to ensure that the audience has the background knowledge necessary to understand Warren's answers. As Colbert asks the clarifying question, he brings his middle finger down to the desk, indicating a position along the speaker-hearer line and relatively close to Warren's body. Once she has sufficiently answered the question, and after a brief joke, Colbert returns to the main discourse to begin the game, signalling the return with the lexical discourse marker "moving on" (line 15).

Because of the length of the digression, and that both speakers are involved, the transcript below marks the relevant gestures of both Warren and Colbert. Colbert's gestures are denoted by asterisks (*) and listed as CGx. Warren's gestures are denoted by plus signs (+) and listed as WGx.

(127) Transcript 23: Elizabeth Warren
[UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335,2121]

```
1  SC   this is a *mild lightning round* *mild lightning round uh*
                  *CG1- - - - - - - - -* *CG2- - - - - - - - - -*
2       *how rich is rich for this two cent thing by the way?
        *CG3 - - - - - - - - - - - - - - - - - - - - - - - - ->
3  EW   +oh it is people who have fortunes of above fifty million
        +WG1 - - - - - - - - - - - - - - - - - - - - - - - - ->
4       dollars+ +so your first fifty million dollars is free and
        <- - - + +WG2 - - - - - - - - - - - - - - - - - - - - >
5       clear+
        <- - +
6  SC   excellent
7  EW   +but your fifty million and first dollar+ +ya gotta pitch in
        +WG3 - - - - - - - - - - - - - - - - - -+ +WG4- - - - - - >
8       two cents every dollar after that+ +[just so you know  ]+
        <- - - - - - - - - - - - - - - - + +WG5- - - - - - - - -+
```

```
9  SC     *                                  *[but then I wouldn't]
          <*                                 *CG4- - - - - - - - - ->
11 SC    have a dollar I would only have 98 cents
          <- - - - - - - - - - - - - - - - - - >
12 EW    +oh boo [hoo+ ]
          +WG6 - - -+
13 SC          [you're] a monster*
          <- - - - - - - - - - -*
14 EW    (laughs)
15 SC    *movin' on, okay Iran versus the Saudis*
          *CG5- - - - - - - - - - - - - - - - - - -*
16 EW    yes
17 SC    *what would convince president Warren that the Iranians did
          *CG6 - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
18        this*
          <- -*
CG1:  right hand, ring hand shape, held center right
CG2:  right hand, loose blocking gesture
CG3:  right hand, point with middle finger down to desk
CG4:  right hand, loose palm up open hand (mostly off screen)
CG5:  right hand, index finger pointed up
CG6:right hand, ring hand shape, several beats, then down to desk

WG1:  both hands, loose containment gesture with several beats
WG2:  right hand, palm down, reach toward Colbert
WG3:  right hand, index finger point toward Colbert, small cyclic motion
WG4:  right hand, two finger point down ward, 2x lateral sweep
WG5:  right hand, two finger point, reach toward Colbert
WG6:  both hands, flat hands held to chest
```

Figure 6.9: Gesture sequence corresponding to TRANSCRIPT 23: ELIZABETH WARREN, ex. 127.

Colbert introduces the game while holding a ring gesture in the central interaction space (line 1, CG1). This hand shape, in which the thumb meets the other fingers as if to pinch a small object, has been associated with precision (Kendon 2004, ch. 14). As described in Chapter 2, this gesture evokes a particular sub-type of the PRESENT action schema in which a the presented object requires close and careful inspection. Colbert then partially releases the ring hand shape and performs a loose blocking gesture as he repeats "mild lightning round" (line 1, CG2). This blocking gesture may serve to 'stop' any inferences of the game being any more threatening than "mild".

Instead of immediately beginning the game, Colbert then digresses to ask for clarifying information (line 2) that will be relevant for one of the upcoming questions. Note that the lexical digressive marker in this case appears at end of his question, whereas his pointing gesture into the central interaction space begins at the beginning of his question. Colbert holds this position as Warren provides an answer to the question. In Kendon's (1995) observations of the *mano a borsa* gesture in Southern Italian speakers, he suggests that speakers may hold a hand gesture past the end of a question in order to mark when the question has been sufficiently answered – the release of the gesture signals that the provided answer has been accepted. Such an analysis would also make sense for Colbert's locating gesture in this example.

After briefly participating in a joking display of outrage at Warren's answer (lines 9-13), Colbert ends the digression with the lexical marker "moving on" and performs a pausing gesture as he does so (line 15, CG5). Finally, when Colbert returns to the main discourse to ask his first question in the "mild lightning round", he repeats the ring hand shape that he held immediately *before* the digression. This type of hand shape repetition to index a topic is discussed in detail in previous work (e.g. Laparle 2021; McNeill et al. 2001).

Importantly, the two locating gestures discussed above differ in their trajectory, as shown in the side-by-side comparison in Figure 6.10. In the short, elaborative digression in (126), the pointing gesture indicates a location off to the side, away from the addressee and the Interaction Space. In the multi-turn, clarifying digression, the pointing gesture indicates a location in the central Interaction Space between interlocutors.

Figure 6.10: Trajectories of locating gestures with BTW

This formal difference between the two gestures makes sense when we consider (i) the function of the Interaction Space in organizing topics as metaphoric objects for discussion, and (ii) differences in the communicative functions of these two digressions. Given the Interaction Space model in Chapter 4, we expect topics that are relevant for ongoing discourse goals to be located within the Interaction Space, where they can be perceived and managed by both participants. We expect topics that are *not* relevant for ongoing discourse goals to be located outside of the Interaction Space, inaccessible to the addressee for further interaction or management. The information provided in a background digression, such as that in (127), is immediately relevant to ongoing discourse goals, as it provides necessary information for answering open questions. For this reason, we expect background information to be located within the Interaction Space. The information provided in elaboration digressions is relevant, but not necessary, for pursuing open questions. For this reason, we do not expect elaborative information to be located in the Interaction Space. These expectations are directly reflected in the two locating gestures discussed.

As with the blocking gestures discussed above, we see evidence of compositionality. In this case the pointing *hand shape* locates some metaphoric object in space, whereas the *trajectory* of the point indicates the status of the topic as either immediately relevant to the discourse topic or digressive. This suggests that locating gestures are not in and of themselves digressive markers, but that the trajectory and location of the metaphoric object

can be manipulated to mark a digression. The capacity of abstract deictic gestures to indicate the discourse-structural status of particular contributions is worth further study in larger data sets.

6.4.4.2 Addressing

Addressing gestures are hand gestures that deictically refer to the addressee and lack the physical affordances of *presenting* or *removing* an object. Three formal variants of *addressing* gestures are shown in Figure 6.11. In the present data set, addressing gestures are associated with meta-discursive digressions, i.e. digressions that comment on the state of the interaction, and also occur four times with topic-shifts. It is also worth noting that of the 12 addressing gestures in the data set, Colbert performs 8 of them. This is likely a repercussion of the social dynamics of the interview format – Colbert is in the more appropriate position to decide on the topic and make general statements about the interaction. The expectations of who is asking questions and who is answering them is also highly asymmetrical in the interview context – Colbert, as host, asks his guests questions. This makes addressing gestures regularly appropriate for Colbert to use, but not his guests.

I argue here that addressing gestures accompanying BTW are *not* digressive markers, and instead perform more general discourse management functions, especially eliciting responses from the addressee (e.g. Bavelas et al. 1992; Bavelas 1994; Mondada 2007). This means that in cases of addressing gestures with BTW, the lexical and gestural discourse marker are performing complementary functions. I will briefly discuss three examples to demonstrate this proposed analysis.[20]

Figure 6.11: Addressing gestures with BTW

The first example, corresponding to Figure 6.11, G1, was discussed as an example of the topic-shift use-context in Section 6.3.1. In this discourse segment, Colbert is talking to Thom Yorke (British musician) about the Royal Family and how they don't seem to do many useful things. Though Yorke seems to find this amusing, he overtly refuses to engage with the topic (ex. 128, lines 3-6). Colbert then attempts a joke about the Royal Family "throwing

[20]Similar uses of addressing gestures appear throughout the three case studies in this work. However, given the above described biases in this particular data set, it is important to test this analysis in more symmetrical interactions.

good parties" (line 7), which Yorke acknowledges but also declines to engage with (lines 2 & 4). After a brief but awkward pause, Colbert attempts to recover from the derailment by initiating a new topic – being knighted – which is tangentially related via the involvement of the Royal Family (line 11). At this point, Colbert leans in toward Yorke and extends a flat vertical palm (fig. 6.11, G1). Note that this gesture is performed through the question and retracted at BTW, which occurs clause finally. Yorke accepts this topic-shift and goes on to discuss how his music has likely offended the Royal Family, thus making knighthood unlikely.

The addressing gesture directs joint attention toward Yorke who is expected to make a subsequent contribution – everyone's attention, including Colbert's and the audience's, is now on him. The use of an addressing gesture is particularly interesting in interactional context – Yorke has refused to fully engage with the discourse through several turn cycles. As described in Chapter 4, addressing gestures, especially when accompanied by postural disengagement, are associated with REQUEST actions, such that the addressee is expected to assume the role as speaker. Given Yorke's recent refusals to do so, Colbert's addressing gesture may play a particularly important role in finally eliciting a full response from Yorke.

(128) TRANSCRIPT 24: THOM YORKE
[UID:705c6f26-02c3-11ea-ba63-089e01ba0335,3329]

```
1  SC   It would be so exciting to see a member of the Royal Family
2       do something useful
3  TY   (laughs) You said that, not me.  I can't say anything.
4       I can't say- I- I'm merely commentating
5  SC   yeah
6  TY   observing
7  SC   I'm sure they throw great parties
8  TY   the Royal Family?
9  SC   yeah
10 TY   (chuckles, nods, mouths "yeah")
11 SC   *would you - would you want to be knighted* by the way?
        *G1- - - - - - - - - - - - - - - - - - - -*
12 TY   nuh- I- I think I've blown that chance
G2:  right hand, palm vertical, reach toward addressee
```

In (129), Colbert is discussing the current political climate in America with Rahm Emanuel (American politician and diplomat). After Emanuel lists some of the most pressing problems (lines 1-3), Colbert offers a new related topic, asking for Emanuel's opinion about two previous US presidents, Bill Clinton and Barack Obama (lines 4-6). This topic-shift is interesting because it partially returns to a previous topic – discussing former presidents – but specifies a topic not yet discussed – how previous presidents debate. As Colbert suggests this topic,

he performs an index-finger point with both hands, oriented toward Emanuel (fig. 6.11, G2).[21] Emanuel accepts this topic-shift, and the discourse moves on.

(129) Transcript 25: Rahm Emanuel
[UID:f6394af0-c95d-11e9-a3fc-089e01ba0770,2056]

```
1  RE   oh yeah it was a recession, a financial crisis, auto industry
2       but outside of that - two longest wars, we kinda cruised
3       right through it
4  SC   so- so- um- *by the way* between those two guys, between Bill
                    *G3- - - - *
5       Clinton and Barack Obama, both sharp as a tack, who's the
6       better debater in your opinion?
G3:  both hands, index finger point, toward addressee
```

The final example in (130) was discussed as an example of meta-discurve digression in Section 6.3.1. In the clip, Colbert and Andrew McCabe (American attorney) are talking about McCabe's recent book publication. As previously described, Colbert interrupts McCabe's publishing story by jokingly reminding him that he is "under oath" – since the book he is discussing regards politically sensitive information. As Colbert says this, he leans toward McCabe, with a casual smirk, and reaches out with a down-turned open hand (fig. 6.11, G3). Colbert holds this gesture for the full duration of the digressive statement. After briefly acknowledging this, McCabe returns to his story, lexically marking the return with "you know".

(130)

(131) Transcript 26: Andrew McCabe
[UID:08beb46e-34e2-11e9-88b3-089e01ba0335,2242]

```
1  AM   as an interviewer I appreciated that question
2  SC   yes yes
3  AM   um (.)  [so I-]
4  SC           [*you're under] oath, by the way, you're under
                 *G3- - - - - - - - - - - - - - - - - - - - ->
5       [oath right now*]
        <- - - - - - -*
6  AM   [thank you okay] you know I worked very closely with the FBI
7       in the process of getting the book published
G3:  right hand, palm down, reach toward addressee
```

Unlike in the other two examples of addressing gestures, Colbert's gesture in this example is not primarily meant to elicit a response, though the acknowledgement McCabe makes is

[21] The hand shape here may preempt the immediately following gesture in which Colbert uses the pointing hand shape to locate the referents, Clinton and Obama, in separate regions of the Interaction Space.

appropriate. Instead, Colbert is directing joint attention toward McCabe as the participant for which a particular communicative imperative (being "under oath" holds). This analysis is supported by the fact that Colbert *interrupts* McCabe to make the meta-discursive comment and perform the accompanying gesture. In the other two examples, Colbert's addressing gesture is produced during his turn.

In all three cases, the addressing gesture directs joint attention to the addressee. However, the function of the gesture – as either eliciting a response or asserting a condition of the interaction – is dependent on the discourse-structural context in which it is preformed. In this data set, addressing gestures occurring with topic-shifts contribute to eliciting a response from the addressee. Addressing gestures occurring with meta-discursive digressions seem to function more simply to gain the attention of the addressee in order to comment on the interaction.

6.4.5 Presentation

Presentation gestures are present in all three case studies in this work. However the formal variation in the observed presentation gestures differ. In Chapter 2, I introduced three sub-types of presentation gestures: (i) *PUOH* gestures, in which a participant enacts the presentation of a metaphoric object held on an open hand (e.g. Müller 2004), (ii) *containment gestures*, in which participants enact the presentation of metaphoric box-shaped object for inspection and 'unpacking', and (iii) *precision grip* gestures, in which speakers enact the presentation of a small delicate object for careful inspection. The diversity of presentation gestures in the current data set is relatively small; 16 presentation gestures occur, 11 of which are PUOH gestures, 4 are containment gestures. Only one precision grip gesture was observed in the data set. In the present data set, presentation gestures are associated with shift contexts. This aligns with the findings in Chapter 5 that demonstrate the use of presentation gestures for topic introduction in during topic-shifting.

Three examples of presentation gestures are presented. The first occurs during a monologue and demonstrates a PUOH presentation gesture during BTW, followed by a series of containment gestures. The following two occur during interviews and are used to elicit responses, much like the addressing gestures described above.

In (132), Colbert is talking about a baseless claim then-president Donald Trump made about politician Beto O'Rourke changing his name to "indicate Hispanic heritage". The Tweet Colbert quotes to introduce this topic ends with a non-sequitor about respecting law enforcement. After reporting on this Colbert begins a humorous critique. He begins this critique with BTW and a contradiction of Trump's claims, then provides an explanation (categorized here as *background*) for why Trump's claims are ridiculous.

(132) Transcript 27: Stephen Colbert
[UID:f56ce17a-ce14-11e9-b211-089e01ba0335,294]

```
1  SC   Nation, *heal.  By the way, Mister President,* *Beto isn't
                  *G1- - - - - - - - - - - - - - - - - - - * *G2- - - ->
```

```
3       pretending/ to be Hispanic./ If he was,* *the part of his
        <- - - - /- - - - - - - - - /- - - - - - * *G3- - - - - ->
3       name he would have changed* *would have been O'Rourke.*
        <- - - - - - - - - - - - - -* *G4- - - - - - - - - - - -*
3       *Okay, Beto/ O'Rouke* *does not sound/ like a Hispanic name*
        *G5- - - - / - - - - -* *G6- - - - - -/ - - - - - - - - - - *
G1:  two hands, flat, trace half circles to form "whole"
G2:  right hand, PUOH, extended rightward, repeated 3x; left hand
     hold, loose, palm up
G3:  two hands, flat, palm center/up forming "v" shape, two beats
     while moving laterally to center body space
G4:  two hands, flat, palm center, sweep up then left and down
G5:  two hands, flat, palm center, two beats wile moving rightward
     across body space
G6:  two hands, flat, palm down quick outward movement from center,
     repeated 2x
```

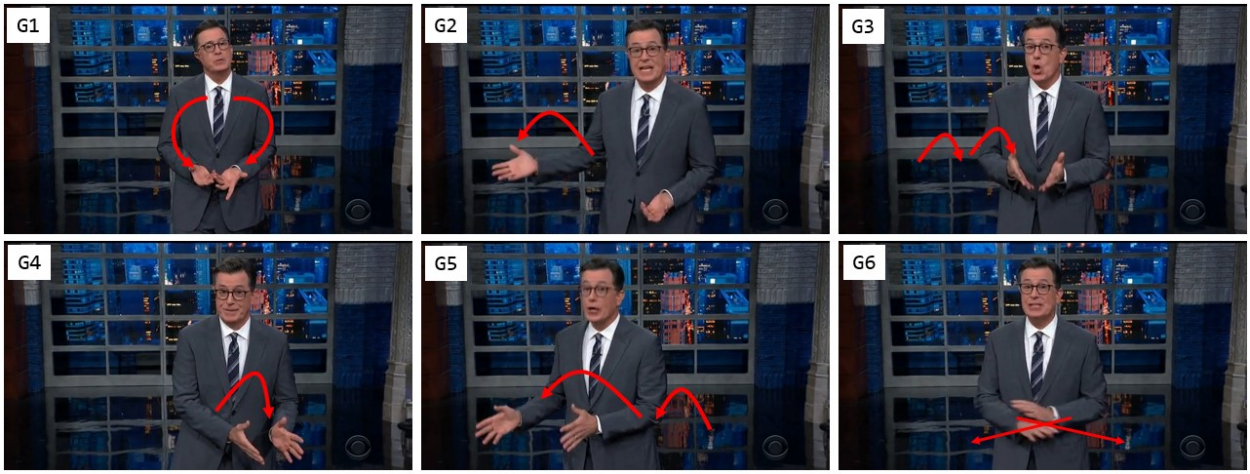


Figure 6.12: Gestures corresponding to

While prefacing the background digression with BTW and "Mr. President", Colbert holds his hands in the ending position of his preceding representational gesture (line 1, G1). He then folds his arm outward for the initial contradiction, as if to present an easy-to-perceive object.[22] This initial presentation gesture is to the side. He then brings his hands to the center of the Interaction Space as he begins his explanation, presenting the topic as centrally

[22]This use of the PUOH gesture is reminiscent of Jehoul et al.'s (Jehoul et al. 2017) analysis of two-handed PUOH gestures as markers of obviousness in Dutch speakers.

located, and maximally accessible in the space. He then performs a series of containment gestures expressing contrast via physical separation in space (e.g. Hinnell 2019, 2020).

The function of this digression is somewhat ambiguous, and seems to change as it progresses. I argue that it is initially introduced as a background digression, in which Colbert offers relevant information in order for the addressee (in this case, an imagined Trump) to better understand the issue he has raised. However, as the 'digression' goes on, it comes to look more like a shift. This is supported by the accompanying gestures. The initial presentation is to the side, where we expect digressive information to be located, but the immediately following gesture physically brings this topic, as a metaphoric object, into the center of the Interaction Space. As an accessible topic in the Interaction Space, Colbert than continues to move and manipulate the topic through presentational gestures, further indicating that the topic has shifted status from a digressive topic to a discourse topic proper.

The following two examples demonstrate the use of PUOH gestures in meta-discursive contexts.

Figure 6.13: Addresse-directed presentation with BTW

In (133), Colbert sarcastically thanks Howard Stern (American radio personality) for "stopping by on the way to the funeral" (lines 4-5), an underhanded comment on Stern's all-black attire. This false platitude serves to shift the topic away from Stern's complaints of the temperature in the studio. Stern accepts this topic shift by defending his attire and then telling a story about how his style changed to mostly black. Colbert's gesture (G1), a two-hand palm-up open-hend gesture, begins at "thank you" and extends through both BTW and most of the remaining utterance. This gesture serves to offer both gratitude and a new topic to Stern, presented as a metaphoric object on open hands.

(133) TRANSCRIPT 28: HOWARD STERN [UID:d88d7758-a2dc-11e9-973b-089e01ba0335]

```
1  SC   this is comedy weather
2  HS   I need a throw blanket or something.  it's unbelievable.
3       why so cold?  [I know --]
4  SC          *[thank you] by the way thank you for stopping by* on
               *G1- - - - - - - - - - - - - - - - - - - - - - - - - *
5       the way to the funeral
```

```
6  HS   I know well I thought I looked good
G1:  two hands, palm up, reach toward addressee
```

This digression appears to by multifunctional, especially when the broader interaction contect is considered. On the one hand, the 'digression' does result in a topic shift – prior to BTW Stern is addressing a question regarding the temperature of the theater, and after BTW he is addressing an unrelated question about his clothing style. However, the move is also clearly meta-discursive offering a (false) platitude and commenting on the present state of the discourse, namely that he is happy to have Stern on as a guest. Note that this example occurs at the beginning of the interview, before discourse goals are solidly established.

In (134), Colbert is talking to Cynthia Erivo (English actor and singer) about her recent successes. Prior to the digression, Colbert is discussing her recent role as Harriet Tubman and the possibility of her being nominated for an Oscar because of it. He then continues this subtopic of Erivo potentially winning an Oscar[23], at the end of which he apologizes for possibly "jinxing" her. This digression, marked with "by the way" between a sarcastic "you're welcome" and genuine "sorry" (line 11), is a comment on his own statement and does not contribute to the main discourse topic, namely discussing Erivo's successes. Colbert performs a two-handed palm up gesture during the sarcastic "you're welcome" as if to offer the "jinx" as a gift presented on his open hands. Colbert than returns to the main discourse to discuss another successful role, lexically marking the return with the marker "now".

(134) Transcript 29: Cynthia Erivo
[UID:90e134a6-fba8-11e9-b915-089e01ba0335,2716]

```
1  SC   if you were so lucky as to win an Oscar, you would be the
2       youngest EGOT, ever Emmy Grammy Oscar, Tony
3       (laughter)
4  CE   just sounds weird doesn't it
5  SC   EGOT?
6  CE   it's weird innit
7  SC   it does yeah it does sound weird (.)  it sounds fun though
8  CE   yeah
9  SC   sounds fun
10 CE   kind of fun
11 SC   *you're welcome for the jinx* by the way [sorry about that]
        *G2- - - - - - - - - - - - *
12 CE                                          [thank you (laughs)]
13 SC   now um uh you're also playing- you're playing another icon
G2:  two hands, palm up, toward addressee
```

[23]Note that this subtopic is temporarily derailed by Erivo commenting on the strangeness of the phrase "EGOT" (lines 3-10).

The presentation gestures discussed in this section do not appear to function as digressive markers. In most cases, they seem to do the opposite by presenting a topic as a metaphoric object into the center of the Interaction Space.

6.4.6 Postural shifts

In this final section, I discuss the ways in which reorientation away from the primary Interaction Space is used to signal digression, where 'primary Interaction Space' refers to the physical space being engaged during the main discourse. This gestural strategy occurs with and without accompanying manual gestures. I discuss three examples in depth. In the first two, a head turn is the only gestural digressive marker in the proximity of BTW. These two examples, though similar in strategy, differ in trajectory. In the first, the digression is marked by a turn *away* from the primary addressee, whereas in the second the digression is marked by a turn *toward* the addressee. Though contradictory when considered in isolation, the gestural context of each demonstrates an important consistency: digressive postural shifts are marked by a disengagement from whatever space is primary to the interaction *at the time of digression*. The third example demonstrates the employment of multiple gestural strategies, disengagement followed by manual stopping, in marking digression.

In (135), Rachel Maddow (American journalist) is answering a question Colbert posed about notable events that haven't made it into the news. Maddow responds by briefly describing a controversy around the then president Donald Trump that had not been extensively reported on. The digression marked with BTW in this example is particularly complex and atypical because it is being attributed to another speaker, namely a spokesperson of the "ways and means committee". As Maddow marks the reported speech as digressive (line 3), she turns her head away from Colbert, purses her lips, and markedly changes the volume and tenor of her voice, all cues that she is performing an impersonation. When she gets to the information that pertains directly to the main discourse, namely the nature of the unreported scandal, she turns her head back to toward Colbert and returns both her facial expression and voice to normal.

(135) Transcript 30: Rachel Maddow
[UID:705c6f26-02c3-11ea-ba63-089e01ba0335]

```
1  RM   just before that we got very quiet word in a court filing
2       that nobody put out a press release about um △that from the
                                                    △P1- - - - - >
3       ways and means committee that said △'oh by the way we've had
        <- - - - - - - - - - - - - - - - - △P2 - - - - - - - - - >
4       a whistle blower come forward to say' there's been △improper
        <- - - - - - - - - - - - - - - - - - - - - - - - - - △P3 - ->
5       influence △by the administration on the handling of Trump's
        <- - - - -△
5       tax returns at the IRS
```

```
P1:  head oriented toward Colbert
P2:  head & gaze oriented away, pursed lips
P3:  gaze oriented toward Colbert
```

Figure 6.14: Gesture sequence corresponding to Transcript 30: Rachel Maddow, ex. 135.

We can analyze this digression in three distinct ways. First, we can fully attribute the digression to the imagined discourse. Under this analysis, Maddow is expressing an evaluation of the actions of the ways and means committee: they attempted to downplay the importance of the news, and an efficient way to express this evaluation is to report the information as a digression, foregrounding the use of digressions to mark information as somehow 'less important'. Traugott (2020:127) refers to these as 'transgressive' uses of BTW, in which "oh BTW" is employed sardonically by the speaker to highlight a particularly shocking reveal. Second, we can also attribute the digression to Maddow herself, who is providing supplementary information, namely her opinion of the conduct of the ways and means committee. A compelling reason to opt for this analysis is that she turns back to Colbert to report the information that relates directly to the main discourse (line 4, P3), even though it is still grammatically presented as reported speech. Last, we can consider this a kind of *meta*-digression, in which Maddow is digressing from her role as interlocutor to perform an impersonation. This analysis is also compelling given that the head turn depicted in Fig. 6.14, P2, is characteristic of viewpoint shifts in both spoken and signed narratives (Stec 2012; Sweetser & Stec 2016).

Instead of deciding between these analyses of the digression, we can accept all three as available to the interpreter. The complexity of this digression and its performance in both the verbal and gestural modes provides the perfect opportunity to analyze the particular ways in which individual digressive signals are profiling distinct aspects of the digression. In the discourse, it *does* seem that Maddow wants to negatively assess the conduct of the ways and means committee, and the lexical marker BTW coupled with the silly voice she puts on during her impersonation, can convey this. The posture shift, on the other hand, profiles

the other two functions of the digression – namely to mark her evaluation as supplementary to the main discourse, and to mark the information as reported speech.

In (136), John Lithgow (American actor), is preparing to read a satirical poem he wrote about then-president Donald Trump. Before beginning, he clarifies a term that Colbert used, *dumpty*, as a name he calls Trump. This clarifying digression (categorized as *background*) is marked with the lexical marker BTW and a postural shift away from the poem he is about to read. The closing of the digression is also verbally marked with the acknolwedgement that "that's another poem" (line 2) and gesturally marked with a prototypical 'away' gesture (Bressem & Müller 2014, 2017; also Chapter 5 of this work).

(136) Transcript 31: John Lithgow
[UID:c6a732b6-56a3-11e9-8c00-089e01ba0335,2058]

```
1  JL   uh Dumpty△ by the way is is my name for our pre△sident (.)
        ≪P1- - -△P2- - - - - - - - - - - - - - - - - -△P3- - - >
2       as in- △as in Trumpty Dumpty *but △that's another poem*
                                      *G1 - - - - - - - - - - -*
        <- - - △P4- - - - - - - - - - - -△P5- - - - - - - - ->
3       and it's uh (mumbles) it's called it's called Afterward
        <- - - - - - - - - - - - - - - - - - - - - - - - - - -≫
G1:  left hand, palm down, lateral sweep left
P1, P3, P5:  head & gaze oriented toward poem in hand
P2, P4:  head & gaze oriented toward Colbert
```

Figure 6.15: Gesture sequence corresponding to Transcript 31: John Lithgow, ex. 136.

Prior to the digression, Lithgow is looking directly at a print out of the poem he is preparing to read (P1). He then turns his head away from the poem and toward Colbert to clarify the term *Dumpty* (line 1, P2). He appears to finish this clarification as he pauses and turns back to the poem (line 1, P3), but then adds additional information to the clarification, turning once again toward Colbert (line 2, P4). Lithgow then unambiguously closes the digression by stating "but that's another poem" and performing an 'away' gesture as if to move an unwanted object out of his immediate space. As he does so, he turns his head back to the poem to again prepare to perform the main discourse task – recitation.

These two examples demonstrate that *disengagement* occurs relative to the space in which the main discourse is taking place at the moment of digression. It is not enough to say that turning away from one's addressee is *disengagement*, though that is surely frequently the case. Instead, it is necessary to carefully consider discourse and gestural context in order to ascertain which position is primary to the interaction at the time, and thus what change should count as disengagement.

The final example revisits a performance of the blocking gesture by Andrea Savage discussed in Section 6.4.3.1. In this discourse excerpt, Savage is describing a stunt that an actor made on set in which the actor broke through glass and threw it into his face. Savage then makes a brief evaluative digression, in which she makes a pun about how good the stunt was (i.e. that it "crushed" much like the glass literally did). As previously described, the blocking gesture aligns only with the evaluative digression (line 2, G1). During the lexical marker BTW, Savage is transitioning between a representational gesture she performed when describing the stunt and the subsequent blocking gesture. However, a postural shift, in which she turns her head, gaze, and shoulders away from Colbert, aligns exactly with BTW (line 2, P2).[24]

(137) Transcript 32: Andrea Savage
[UID:67f2c08a-f628-11e9-920c-089e01ba0335,3371]

```
1  AS   he leaps into the glass and throws glass△ in his face
                                                  △P1- - - - - ->
2       △by the way△ *it crushed*
                     *G1- - - -*
        △P2- - - - △P3- - - - -≫
G1:  two hands, flat, palm away, lateral sweep out
```

[24]Note that this performance is nearly identical to that of Elizabeth Banks in Fig. 6.2, ex. 121.

Figure 6.16: Gesture sequence corresponding to TRANSCRIPT 32: ANDREA SAVAGE, ex. 137.

This example demonstrates the ways in which different gestural strategies can be used independently in marking a digression. The lexical strategy BTW and the gestural strategy of disengagement are employed at the same time to mark the beginning of the digression. The subsequent blocking gesture then takes scope over the digressive utterance, signalling a temporary stopping of the main discourse. Intensification of the evaluative statement is expressed by the outward movement.

6.5 Discussion

This chapter has argued for the the inclusion of certain gestural markers as digressive markers on par with lexical digressive markers like BTW. In particular, I identified two gestural strategies – *manual stopping* and *disengagement* – for marking an utterance as digressive. *Manual stopping*, which can take the form of *blocking* and *pausing* gestures, serves to mark a temporary halting of the main discourse. Frequently, these gestures take scope over the entire digression, only ending once the digression is over and the main discourse is allowed to continue. *Disengagement*, which takes the form of reorienting away from the main interaction space, serves to signal a disengagement from the main discourse topic and can occur independently of manual gestures. I also discussed two other classes of manual gestures, *referring* and *presentation* gestures, which I argued were gestural discourse markers that perform discourse management functions *other than* digression. *Referring* gestures draw joint attention toward a participant of metaphoric object in order for that participant or object to be attended to in the discourse. *Presentation* gestures, as discussed in Chapter 5, present a topic as a metaphoric object into the Interaction Space for further inspection and manipulation.

The complex relationship between the lexical digressive marker BTW and co-occurring gestures problematizes approaches to discourse structure that only allow for simplex underlying relations (as in RST; Mann & Thompson 1988; Taboada & Mann 2006) or single questions (as in QUD; Roberts 1996, 2012). Instead, these findings suggest a richness and multifunctionality, where different expressive strategies profile different features of the complex discourse move. Though there are theories that already center this kind of multi-

functionality, such as Halliday's (1985) Systemic Functional Grammar, I believe that other popular approaches to discourse structure can be modified, to the benefit of both the theories themselves and our broader understanding of multimodal compositionality.

A full account of digressive marking requires us to consider not only multiple modes, but also multiple strategies within each mode. Lexical, prosodic, and syntactic constructions all work in concert with manual gestures and postural shifts to express a diversity of disruptive discourse moves, ranging from brief evaluative digressions to full topic shifts.

Chapter 7

Here's the thing about multimodal specification

7.1 Introduction

Specification is a discourse structural phenomenon in which an utterance contributes to answering an open discourse question. This can be considered the default, maximally cooperative, discourse move in any question-based model of discourse structure – *every* utterance should contribute to answering an open discourse question or setting a strategy for achieving discourse goals. In this light, the previous two chapters on multimodal topic-shifting and digression explored exceptional discourse structural phenomena in which a given utterance in some way does *not* contribute to established goals. In these cases, the utterance threatens discourse coherence – *why say something that doesn't help us achieve our goals*? To maintain coherence the intention and relevance of the speaker's contribution must be made clear. The previous two chapters discussed some ways in which this is done through lexical and gestural discourse markers. The specification discussed in this final case study does not put discourse coherence at risk. One might think that specification moves are thus not particularly worth expressing overtly – *if an utterance contributes to discourse goals as expected, why bother expressing that it does so?* I argue that lexical and gestural markers occurring with specification moves serve to draw attention to the proceeding utterance, emphasizing the importance of a particular contribution to the discourse's structure.

In this chapter, I explore the gestural expression of specification moves aligned with the lexical discourse marker *here's the thing* (henceforth HTT). HTT signals that the proceeding utterance adds important information to an ongoing discourse topic. In overtly expressing an expected relation (specification), the lexical discourse marker serves as emphasis, drawing the participants' attention to the perceived importance of a particular specification in achieving discourse goals. Though, to the author's knowledge, there is not significant work on HTT itself, this interpretation of HTT aligns with work on functionally similar lexical discourse markers, including *the thing is* (Wang 2016) and *the fact is that* (Aijmer 2007). This

interpretation is also supported by the spatial component in the phrase itself. The proximal deictic expression "here" signals importance and relevance by metaphorically locating the topic in the participants' immediate space.

Given the function of the Interaction Space in managing discourse goals and the accessibility of discourse topics, we can make predictions as to what types of interactive gesture we expect to co-occur with HTT. These predictions are summarized in (138).

(138) **Gestural expression of (emphasized) specification**

a. *Consistent engagement with the Interaction Space*: If the utterance marked with HTT contributes directly to ongoing discourse goals, we expect the speaker of HTT to maintain engagement with the Interaction Space. They must do so in order to perform relevant discourse management within the Interaction Space and add to the discourse structure therein.

b. *Presenting gestures*: If the utterance marked with HTT contributes new information deemed particularly important to the discourse, we expect presentation gestures into the central Interaction Space. Such gestures serve to make the new information maximally accessible to both participants for further management.

c. *Locating gestures*: If the utterance marked with HTT contributes particularly important new information to an ongoing discourse topic, we expect locating gestures to refer to established topics already in the Interaction Space. This is supported by the presence of the proximal deictic phrase "here" in the lexical discourse marker itself.

As with the previous case studies, these predictions are mediated by other discourse structural and interactional factors, such as speaker attitude and turn position. For example, whether a presenting or locating gesture is used may depend on the relative accessibility of the introduced subtopic – particularly surprising contributions may be more likely to be *presented* rather than *located* because their presence as a part of a topic already in the Interaction Space may be harder to accommodate.[1]

The data largely supports these predictions. Speakers remain engaged and heighten their engagement when expressing specification, and frequently perform presentational and locating gestures into the central Interaction Space. However, a more interesting picture emerges by considering particular gestural forms with particular use contexts. For example, Kendon's (1995; 2004, ch. 12) 'precision grip', in which the fingers are bunched together as if to carefully hold a small object, appears recurrently. As in Kendon's analysis, these gestures help to convey the particular importance of the specification, as if to say "it is this single thing that we should pay the most attention to". There are also recurrent unpredicted

[1] An analogy can be made between this and presupposition accommodation of so-called 'inferrables' in the verbal mode (e.g. Prince 1981). For example, when a character walks into a room in a narrative, unsurprising entities can be referred to with a definite article (e.g. "they walked into *a* side room and gazed out *the* window"), whereas unexpected ones cannot (e.g. #"they walked into a side room and gazed into the wormhole").

gestures, namely stopping gestures, in which speaker's request the stopping of some action through blocking and 1-second gestures. I will discuss how these 'exceptional' specification gestures unexceptionally profile other aspects of the discourse move, such as qualification and contradiction.

The present chapter is structured as follows. Section 7.2 gives an overview of data and annotation procedures. In Section 7.3, I review the observed structural and functional variation of HTT in the present data set and discuss the ways in which this variation affects our predictions as to gestural discourse management strategies. I then review the observed gestural variation in Section 7.4, and discuss the ways in which the data aligns with and diverges from the predictions outlined above. I focus not only on the gestures immediately aligned with HTT, but also those aligned with the proceeding specification. Section 7.5 concludes.

7.2 Data & Methodology

All data comes from interviews and monologues on the talk show Late Show with Stephen Colbert, collected through the UCLA television news archive in collaboration with the Red Hen Lab. An initial search for the string "here's the thing" on the Late Show with Stephen Colbert between 2015 and 2019 yielded 619 results (including repeated clips and commercials). The most recent 100 unique instances were annotated, resulting in a data set ranging from December 2018 to December 2019. This includes 61 examples from monologues, and 39 examples from interviews. 9 of the 61 instances during monologues involve interactions between Colbert and his band leader, Jon Batiste, on stage. The unique corpus ID (UID) is provided for each example.

All examples were coded for the following variables. Detailed definitions of variables are discussed where relevant throughout the chapter.[2] Each example was also notated with a 1-2 sentence qualitative description of gestural activity and communicative context.[3]

(139) a. Discourse type: *interview, monologue, band interaction*
b. Speaker: *Colbert, Guest*[4]
c. Lexical unit: *HTT* and any accompanying discourse markers (e.g. *but HTT*)
d. Phrase type: *independent, extended* (e.g. "here's the thing about dissertations"), and syntactic blend (e.g. "here's the thing is that dissertations are long")
e. Relation: *background, circumstance, contradiction, elaboration, shift, qualification, return*

[2]The reader is directed to Chapter 2 for a full discussion of all gestural variables.

[3]All annotations, including for those examples excluded from the present study, are available upon request.

[4]The name and demographic information of all guests was also recorded, including *occupation, gender, race*, and *nationality*.

f. Topic structure: 1-3 word description of discourse (sub-)topic before and after HTT (*topic 1* → *topic 2*); where the speaker is immediately interrupted, the topic structure is annotated as *interrupted*

g. Turn structure: for interviews – *cede, HTT, keep, take*; for monologues – *NA, take, end impersonation, IS shift*

Where only speech is of concern (Section 7.3), the speaker's name and relevant discourse fragment is presented as a simple transcript. Where gesture is also of concern (Section 7.4), a multi-tier transcript of the discourse fragment is provided, using conventions from Embodied Conversation Analysis (Mondada 2018) to highlight gesture-speech alignment.

7.3 Variation in the verbal mode

The lexical discourse marker HTT appears in a range of use contexts and exhibits structural variation. Though I argue that HTT marks a specification move in all use contexts, it is underspecified for the discourse relation connecting the two adjacent discourse segments (Mann & Thompson 1988; Taboada & Mann 2006). Seven distinct discourse relations were identified in the present data set. Four of these *returns, elaboration, circumstance* and *background* relations account for over three-quarters of the data set (76/100). Examples of each observed relation and the effects on predictions for gestural expression are discussed in Section 7.3.1. These relations correlate with other lexical discourse markers that recurrently appear adjacent to HTT in the data set, especially *but*. This lexical variation is summarized in Section 7.3.2. Section 7.3.3 discusses a second source of structural variation, which I call *phrasal* variation, in which the HTT phrase can optionally be extended, as in "here's the thing *about interdisciplinary dissertations*". The final source of structural variation discussed is the location of HTT relative to turn structure. Turn-medial and turn-initial HTT are both relatively common, but play distinct discourse structural and social roles. This is discussed in Section 7.3.4. An interim summary of the verbal variation observed with HTT and its impact on predictions of gestural expression is then provided in Section 7.3.5.

7.3.1 Discourse-structural variation

To understand and categorize the connection between the specification prefaced by HTT and the immediately preceding discourse segment, I again use the conception of 'discourse relation' as established in Rhetorical Structure Theory (Mann & Thompson 1988; Taboada & Mann 2006). Under this approach, a discourse relation is an underlying proposition that conveys information about how two adjacent discourse segments relate to one another. These relations may or may not be expressed overtly in the verbal (or gestural) mode. This conception of discourse relation aligns with the understanding of specification as a discourse structural phenomenon, rather than a product of overt linguistic constructions, and allows for the optionality of expression inherent in my proposed principles of multimodal discourse management.

In the present data set, I identify seven distinct discourse relations underlying the specification move marked by HTT: *background, circumstance, contradiction, elaboration, qualification, return* and *full shift*. A simple definition of each is given in Table 7.1.[5]

Relation	Definition
Background	a specification that provides information necessary or helpful for the comprehension of an adjacent discourse segment
Circumstance	a specification that provides information against which to evaluate an adjacent discourse segment's relevance to current discourse goals
Contradiction	a specification that provides information that requires a reassessment of the truthfulness of an adjacent discourse segment
Elaboration	a specification that contributes information deemed particularly important to answering the current discourse question
Qualification	a specification that contributes information against which to judge the speaker given an adjacent discourse segment
Return shift	a topic-shift in which an open question is returned to after a digression
Full shift	a topic-shift in which a discourse goal is abandoned in favor of a newly introduced topic

Table 7.1: Discourse relations underlying HTT

Background, circumstance and *qualification*, as defined here, are closely related; all three flavors of specification add information that is important to the interpretation of an adjacent discourse segment. However, the level at which the specification operates is distinct. The *background* relation operates primarily at the semantic level, helping an interlocutor understand what is being said. The *circumstance* relation operates primarily at the discourse structural level, helping an interlocutor understand why something is being said. The *qualification* relation operates primarily at the social level, helping an interlocutor decide how they should feel about the speaker given their contribution. The *elaboration* relation is also similar to these three in that the specification contributes important information about an

[5]Three of the identified relations, *background, circumstance,* and *elaboration*, are taken directly from Mann & Thompson's (1988) proposed discourse relation inventory. The *contradiction* and *qualification* relations are similar to Mann & Thompson's *antithesis* and *justify* relations respectively, but are modified to reflect their particular functions in dialogue. *Return* and *full shift* relations refer to varieties of topic-shifts, defined and discussed in Chapter 5 on *anyway*, and do not have obvious equivalents in the RST framework.

adjacent discourse segment, but does not specify a particular level at which this information should be considered important. All four of these specification types contribute directly to established discourse goals, though they use different strategies to do so.

The remaining three relations, *contradiction, return shifts*, and *full shifts*, are relatively atypical as specification moves in that each contributes to a change in discourse goals. The contradiction relation cancels some line of inquiry posed by another interlocutor, real or imagined, but maintains a higher level discourse goal. Return shifts and full shifts are types of topic shifts introduced in Chapter 5 on *anyway*. Return shifts close a digressive topic in order to return to a higher level open topic. Full shifts close an entire discourse topic in favor of a new one. In these three functional contexts, two forms of discourse management seem to happen simultaneously. First, a topic or strategy for pursuing a topic is ended. This management is not considered a part of the specification move. The HTT-marked specification, as in all cases, signals information as particularly important to achieving goals. The atypicality arises from this contribution also being the first made to the discourse topic.

These seven relations account for 94 of the 100 annotated instances of HTT. There are two examples in which HTT is immediately repeated, which I count as a single example. There are an additional four instances in which the speaker is immediately interrupted after HTT, leaving the discourse relation indeterminable. Table 7.2 summarizes the distribution of observed discourse relations for each interaction type. Because each interaction type has unique spatial and social affordances, different discourse structural and gestural patterns are expected. As previously stated, there are 9 examples during the "monologue" segment of the show in which Colbert is interacting directly with his band lead. Because these also demonstrate unique spatial and social affordance, these 9 examples are presented as separate from "monologue" counts, even though they occur during the monologue segment of the show.

	Background	Circumstance	Contradiction	Elaboration
Monologue	3	7	5	10
Band interaction	3	3	-	2
Interview	8	6	5	4
Total	**14**	**16**	**10**	**16**

cont.	Qualification	Return Shift	Full Shift	**Total**
Monologue	1	22	3	**51**
Band interaction	-	1	-	**9**
Interview	2	7	2	**34**
Total	**3**	**30**	**5**	**94**

Table 7.2: Discourse relations underlying HTT by discourse type

There are two noteworthy differences between monologues and interviews. First, despite the fact that interview data is less frequent than monologue data in this data set, the *background* relation occurs most in interviews. I hypothesize that this relates to navigation between different Interaction Spaces, and thus between different shared knowledge states. Consider a conversation between three friends, two of whom are housemates. The two housemates likely have more shared knowledge as to their daily happenings than the third friend does. This asymmetry in shared knowledge will likely result in the third friend asking for and being offered background information that would not be included in a conversation between the two housemates (e.g. the fact that a third housemate has messy habits and has been causing strife in the household). In an interview, each participant has at least two active Interaction Spaces available to them – the primary space between interviewer and interviewee, and a secondary space between each participant and the audience. Because the knowledge shared by each set of participants is different, it is likely that background information will have to be added to one or the other space in cases of perceived knowledge asymmetry. Second, return shifts are disproportionately represented across interaction types, appearing at a higher rate in monologues (43% of instances) than in interviews (21% of instances). This distribution is similar to that seen for *anyway* and reflects an idiosyncratic joke construction in which Colbert briefly digresses from a discourse topic to make a joke (usually a pun) before returning to continue the topic.

In the remainder of this section, I present examples of each relation and discuss the predicted impact on the gestural expression of specification. I begin with the four specification varieties that can be considered typical – *background, circumstance, qualification* and *elaboration.* I then turn to the atypical functional contexts in which the specification aligns with some strategy or topic shift – *contradiction, return shift, full shift.*

7.3.1.1 Background

The *background* relation expresses that a particular discourse segment is helpful in understanding the content of an ongoing topic (Mann & Thompson 1988:273). Where HTT aligns with a background relation, the proceeding utterance adds information that is deemed necessary for achieving the current discourse goal. A paraphrase of this could be "here's the thing you need to know first". In (140), Colbert and American musician Pharrell Williams are having a disagreement as to the possibility of time travel. Williams is convinced time travel is possible, while Colbert remains skeptical. In response to Colbert's skepticism (lines 1-4), Pharrell tries to establish important common ground before attempting to convince Colbert of his viewpoint (lines 5-6).

(140) **Background**

TRANSCRIPT 1: PHARRELL WILLIAMS

[UID:13f49f6c-18c4-11ea-b152-089e01ba0335.1850]

```
1  SC   Do we know if it's possible?  I'm not sure that we know that
2       it's possible.  My understanding is that the arrow of time,
```

```
3        y'know, could go in either direction.  But how would we
4        actually go?
5   PW   Okay so here's the thing, like time and space are
6        interlocked, right
7   SC   I- I've heard.
```

The information prefaced by HTT ("time and space are interlocked"), contributes directly to the ongoing discourse topic regarding Williams' belief in time travel. Specifically, it contributes information that Williams thinks will make Colbert more amenable to his position. Colbert has indicated to Williams that he is unlikely to agree with Williams' belief, leaving Williams to convince him otherwise. HTT helps to achieve this by drawing attention to important background information which is (i) necessary to Williams' argument, and (ii) information that both interlocutors can agree upon. Once both Williams and Colbert agree on the premise that "time and space are interlocked", Williams can explain why he thinks time travel may be possible. In this case, the background relation is being exploited for argumentation, but this is not always the case.

Because the background relation marks the introduction of new necessary information, we predict presentation gestures into the central Interaction Space. This is because information that is necessary for achieving discourse goals should be maximally accessible for reference and management. There are no differences between the predictions made for the gestural expression of a background relation and those made for the more general specification move (ex. 138).

7.3.1.2 Circumstance

The *circumstance* discourse relation, according to Mann & Thompson's original formulation, expresses that a particular discourse segment is meant to provide a framework against which to interpret another discourse segment (1988:272). Where HTT aligns with a circumstance relation, the specification proceeding HTT comments on the relevance and intended impact of an adjacent discourse segment. A paraphrase of this could be "here's the thing that makes what was just said / is about to be said relevant and interesting". This is distinct from the *background* relation because the information provided in the utterance prefaced by HTT is not necessary for understanding *what* is being said. It is not intended to affect the interlocutor's beliefs. Instead, the HTT-marked utterance in a circumstance relation provides information that is helpful in assessing and evaluating some information provided. It is intended to affect the interlocutor's understanding of *why* something is said, and how they should feel about it.

When the HTT-marked specification occurs *before* the assessed information, the specification "sets the stage" on which proceeding utterances should be evaluated. This is exemplified in (141). When the HTT-marked specification occurs *after* the assessed information, you get a kind of "punch-line" effect, in which an earlier utterance gets reassessed in light of the new specifying information. This is exemplified in (142).

The transcript in (141) is taken from a highly animated segment of dialogue between Colbert and Indian-American actor Mindy Kahling. In this clip, Colbert is trying to recover after an automated call from a Marriott hotel rewards program interrupts the interview. Colbert is embarrassed and apologetic. Kahling is insulted at having been interrupted and confused as to why someone as wealthy as Colbert would need a rewards program (lines 3-5). Colbert attempts to explain, but is cut off by Kahling (lines 6-7). After a series of overlapping utterances, Kahling secures her turn with HTT. Her use of HTT in this context marks the relevant circumstance: that Colbert has the "number 1 show" on television, and can thus be assumed to be quite wealthy. She then puts on an exaggerated voice while speaking "as Colbert" to the audience (lines 9-10).

(141) **Circumstance 1**

TRANSCRIPT 2: STEPHEN COLBERT & MINDY KAHLING
[UID:fc6004ce-bb38-11e9-afb5-089e01ba0770,2182]

1	SC	This is how you get rich, you go for the rewards. Exactly.
2		I also steal rolls from au bon pain.
3	MK	I don't even- y'know, I don't even know. I have so many
4		more questions about your financial situation. I don't use
5		a rewards program. He has-
6	SC	Here's the thing. [Hold on. Here's the thing. hold on]
7	MK	[No. He has- no, no. Here's the thing]
8		he has the- he has the number one show on t.v. and he's like
9		"ah, I gotta make sure that if I travel, that me and my wife
10		and my three children are being hosted in an economical way"

In this example, the audience is able to understand the semantic content of Kahling's impersonation without the prefacing HTT-marked specification. The specification, coupled with Kahling's exaggerated voice during the impersonation, helps the audience understand that they should assess the information conveyed during the specification negatively – despite being rich, Colbert has the audacity to sign up for the reward program that has just interrupted the interview.

The repetition of HTT by both interlocutors (lines 6 & 7) indicates a disagreement in whose contribution is more important to the ongoing topic. In this case, Colbert has inadvertently derailed the interview. He hopes to earn forgiveness by providing a sufficient excuse. His attempted excuse is rejected by Kahling who asserts that her interpretation of the derailment is both more accurate and more important.[6]

The discourse structural and social complexity of this example highlights the importance of considering all contextual dimensions when predicting and analyzing gestural expression. The circumstance relation itself marks the addition of important information for assessing a

[6]A nearly identical exchange occurs one other time in the data set (UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335), as Colbert calls out American politician Elizabeth Warren for attempting to avoid answering a question about taxes.

particular aspect of an ongoing discourse topic. Based on this discourse structural function, we may predict locating gestures to be relatively common, such that the speaker refers to a particular region of the Interaction Space to draw attention to the topic being assessed. However, other dimensions of the context may override the efficacy of expressing this relation in the gestural mode. For example, Kahling interrupts Colbert's attempted contribution, a violation of the cooperative principle of 'one speaker at a time' (e.g. Sacks et al. 1978). This relatively unexpected violation of turn-structure may be worth marking more than the relatively expected discourse move of specification. If this is the case, then we might expect a stopping gesture in this context as Kahling requests Colbert to cede his turn. We will return to this example in Section 7.4.3 to see how these competing communicative demands play out in the gestural mode.

A second example of the circumstance relation is given in (142). Here, Colbert is recounting recent ridiculous actions taken by then-president Donald Trump (a recurrent discourse topic on his show at the time). He makes a relatively unremarkable statement about Trump using the presidential motorcade and displays a video clip of the motorcade (lines 1-2). He then provides the circumstance against which to assess the given information – Trump used the motorcade, which consists of several armored cars and many personnel, just to cross the street (lines 3-4).

(142) **Circumstance 2**

TRANSCRIPT 3: STEPHEN COLBERT

[UID:c1c63da2-08e0-11e9-89e1-089e01ba0770,324]

```
1  SC   Last night, President Trump met with the Bush family at
2       Blair House.  Here he is arriving by presidential motorcade.
3       Um here's the thing, Blair House is literally across the
4       street from from the White House.  And Trump used the
5       motorcade to travel 250 yards.
```

In this example, the information conveyed in lines 1-2 consists of a simple recounting of events. Comprehension of the event, or agreement that the event happened, is not at risk. However, the relevance of the statement to Colbert's typical critique of Trump's actions is not immediately clear – every president uses the presidential motorcade, so why would Colbert present this fact as noteworthy? It is the specification following HTT that provides the circumstance under which Trump's use of the motorcade is relevant and noteworthy. Delaying the specification until *after* the relevant discourse segment serves to build suspense, and results in laughter from the audience as they reassess Colbert's previous contribution and realize the unexpected ridiculousness of an otherwise unexceptional event.

For examples like (142) in which the circumstance *follows* the information being assessed, we expect locating gestures to occur. This is because the topic under discussion should already be present in the Interaction Space and thus available for reference via a locating gesture. Locating gestures in this context would serve to direct joint attention toward the metaphoric object that needs reassessment.

7.3.1.3 Qualification

A *qualification* preempts, and attempts to prevent, a negative judgement made about the speaker as a result of a proceeding contribution. This is similar to a *circumstance* relation in that the specification provides information that helps the interlocutor perform an assessment. However, the assessment here is of the speaker's actions rather than the relevance or value of their contribution to achieving discourse goals. In other words, when aligned with a qualification relation, the segment prefaced by HTT conveys important *social* information, rather than discourse structural or semantic information.

In (143), Colbert is interviewing American actor Will Ferrell who is presenting as Rob Burgundy, the somewhat seedy character he plays in the movie Anchorman (2004). Colbert poses a morally-charged question, "did you ever sleep with a subject to get an interview?" (lines 1-2). Burgundy's actual answer is given in line 4 ("yes", "about fifteen times"). Knowing that this is the morally incorrect answer, Burgundy preempts the negative judgement by specifying that he share's Colbert's (and presumably he audience's) moral stance. This segment prefaced by HTT does not contribute actually answering the question at all. Instead this segment adds a qualification, signalling that Ferrell is willing to offer an answer, but only after specifying that he knows it is immoral.

(143) **Qualification**

TRANSCRIPT 4: WILL FERRELL

[UID:d129e820-ba6f-11e9-a245-089e01ba0335,2886]

```
1  SC   Did you ever and would you ever sleep with a subject to get
2       an interview?
3  WF   Well, here's the thing, it is completely unethical, it is
4       abuse of power.  But yes, I did it, about f-, about fifteen
5       times.
```

As in the *circumstance* examples discussed above, the interlocutor's understanding of Burgundy's answer is not at risk at the point of the HTT-prefaced specification. Unlike the *circumstance* examples, the HTT-prefaced specification does not encourage the interlocutor to reassess Burgundy's contribution, "yes" is clearly the answer. Instead, it helps the interlocutor assess Burgundy's moral standing against his immoral answer.

Qualifications serve a very particular social purpose – to stop a negative assessment of the speaker by the addressee. Stopping gestures, gestures that signal a request to stop some action through iconicity (*blocking* gestures) or convention (*pausing* gestures), are well suited for expressing this social function. Note that stopping gestures are not predicted for specifications generally since they do not express contribution of information, the primary function of specifications.

7.3.1.4 Elaboration

An *elaboration* adds information to an ongoing discourse topic. This relation is used for segments with HTT that can be considered specifications (i.e. that contribute to answering

an open question), but that do not exhibit the particular functional properties associated the *background, circumstance, contradiction* or *qualification* relations. Elaborations marked by HTT seem to emphasize information that is in some way surprising rather than crucial to achieving perceived discourse goals. Often, this involves including additional information when the topic may otherwise be ended, as is the case in (144).

In this clip, Colbert is discussing an incident in which American actor and public figure Gwyneth Paltrow reportedly injured an elderly man while skiing recklessly. Colbert digresses from this main topic to talk about the ski resort, Deer Valley, where the incident happened. Colbert notifies the audience of the digression by saying "I'll get back to the story in just a minute" (line 1). Colbert precedes to humorously compliment the resort (lines 2-4), opening a digressive question regarding the quality of the resort. He then pauses briefly, signalling a likely return to the main topic. Instead of returning, however, he further elaborates on the digressive topic, opening a new sub-question regarding the best ski trails. He marks this unexpected elaboration with "here's the deal" (line 5) and "here's the thing" (line 6).

(144) **Elaboration**

TRANSCRIPT 5: STEPHEN COLBERT

[UID:8bbb4f6a-2918-11e9-97a6-089e01ba0335,1252]

```
1  SC   I'll get back to the story in just a minute, but first I
2       want to talk about Deer Valley.  They have the best service
3       of all the ski resorts, and they don't allow those
4       snowboarding hoodlums on the mountain, okay.  It's absolute
5       paradise.  Try the turkey chili (.)  But here's the deal, a
6       lot of the ski runs there are really tricky.  Here's the
7       thing, okay, uh, you wanna come in here, you wanna get on
8       Silver Lake Express right away.  That will take you all the
8       way up to Silver Lake Lodge.
```

Unlike in the three relations already discussed, this elaboration does not seem to serve a particular purpose other than continuing an ongoing topic. The elaboration does not provide information necessary for comprehension of an adjacent segment, nor does it provide information helpful for assessing an adjacent contribution. Instead, the function of an elaboration is to emphasize the presence of another open sub-question that needs to be addressed. In the data set, these examples are often humorous, intentionally drawing attention to a topic that may otherwise be accepted as closed, such as the example above.

As with specification moves more generally, we expect to see presentation and locating gestures into the Interaction Space where they are maximally accessible for inspection and management. However, there may be discourse-structural factors that independently impact the trajectory of presentation and reference. In cases such as the example above, where the elaboration adds to a digressive topic rather than a main discourse topic, we may expect these gestures to occur in the periphery of the Interaction Space. As discussed in the previous chapter on *by the way*, this location on the periphery of the Interaction Space serves to

decrease the accessibility of the topic as a metaphoric object, signalling that it is not available for further management actions.

7.3.1.5 Contradiction

The *contradiction* relation signals a segment that refutes another interlocutor's actual or reported speech.[7] In interviews, this is typically done at the start of a turn and serves to refute some aspect of the other participant's contribution. In monologues, this relation typically occurs after an impression or video clip. Though the semantic contribution is the same (some information is being refuted), the social contribution is distinct. When the contradiction occurs during a monologue, the reported speech is presented, in a sense, *to be contradicted*. In an interview the contradiction refutes another participant's genuine contribution and can result in a derailment as discourse goals are renegotiated.

In (145), Colbert is asking American lawyer and politician Adam Schiff when the public will know whether or not then-president Donald Trump had taken illegal actions during his presidency (lines 1-8). In asking this, he is presupposing that the answer is not currently available. Schiff contradicts this presupposition by saying that we actually have the answer, which is that there is already sufficient evidence of the president's misconduct (lines 9-11). The initial contradiction "we already know what happened" (line 9) is prefaced with "well here's the thing". Schiff then proceeds to describe the cases of misconduct we already know about.

(145) **Contradiction (actual speech)**
TRANSCRIPT 6: ADAM SCHIFF
[UID:12908332-1d7b-11ea-903c-089e01ba0335,2819]

```
1  SC   If this is really about patriotism, if this is really about
2       finding the truth, if it's really about holding a president
3       to account for their misdeeds, how can you, as much as you
4       don't want to drag this out, how can you go forward without
5       knowing what the truth is?  And if we don't find out now, how
6       many years will it be before these things are unsealed?  How-
7       how many vegetables do I have to eat?  Do I have to do cardio
8       to live long enough to know what really happened here?
9  AS   Well here's the thing, we already do know what happened.
10      The evidence of the president's misconduct is already
11      overwhelming.
```

Schiff's contradiction of Colbert's contribution does not result in derailment, but does nonetheless shift discourse goals. Colbert sets a discourse question – *when will we know*

[7] The *contradiction* relation as defined here is roughly equivalent to the *antithesis* relation in a traditional RST framework (Mann & Thompson 1988:253-4). However, RST's antithesis relation was conceived for monologic, written discourses and thus does not accommodate differences and negotiation in discourse goals that occur in dialogues.

what really happened? This is rejected by Schiff and replaced with something like – *what do we already know really happened?* The HTT-marked specification provides the primary reason for this shift.

In (146), Colbert is reporting on a protest that occurred in the US capital in which Republican law makers stormed the impeachment hearings of then-president Donald Trump. The reported speech (line 1) presents Republicans as suggesting that they do not know what is happening in the impeachment trials, the supposed justification of their actions – why else would they dramatically storm a governmental hearing? The HTT-marked specification (lines 2-6) contradicts this – Republicans did have access to the impeachment trial, and thus their actions are not justified.

(146) **Contradiction (reported speech)**
TRANSCRIPT 7: STEPHEN COLBERT
[UID:94aa7702-1668-11ea-9503-089e01ba0335,527]

```
1  SC   The GOP is saying, "we need to know what's going on inside
2       of there".  But here's the thing, 47 Republicans on the
3       committees leading the investigation have access to the
4       closed door depositions and Republican lawyers are given the
5       same amount of time to question witnesses as Democratic
6       counsels.  In fact, of the Republicans who RSVP'd for the
7       room storming, 12 of them are allowed to sit in on all
8       depositions.
```

In the context of this monologue, the contradiction is likely expected – the main focus of Colbert's monologues is the critique of current events. The HTT-marked specification draws attention to the shift from reporting the events as presented in the news (lines 1-2) to Colbert's critique of those events (lines 2-7).

The contradictions in examples (145) and (146) are similar in function to specifications with an underlying circumstance relation – both flavors of specification provide information against which to assess an adjacent discourse segment. Unlike the circumstance relation, however, the contradiction relation requires a reassessment of truth-value rather than relevance to discourse goals.

I further argue that contradictions involve a kind of implicit topic dismissal – a discourse topic pursued by one interlocutor (real or imagined) is in some way refuted and rejected by another. For this reason, we may expect to see gestures similar to those observed in topic shifts. The rejection of the previous contribution may be expressed through *stopping* gestures, signalling a request to end a line of inquiry, or *away* gestures, signalling the removal of the rejected topic. As with topic-shifts, we also expect to see the presentation of the new topic into the Interaction Space. Because a higher-level ongoing discourse topic is maintained, these examples are most closely analogous to the *partial shifts* discussed in Chapter 5.

7.3.1.6 Topic-shifts

The four relations discussed above are typical specification moves in that they present new information about an ongoing discourse topic. Of the four, the *contradiction* relation is the most atypical in that it also involves the rejection of a line of inquiry. However, the HTT-marked segment still contributes to achieving a discourse goal that the preceding segment also contributes to. As such, its role as a specification move is still apparent. It is much less obvious that the remaining two relations observed in this data set, *returns* and *full shifts*, are types of specification moves at all. Indeed, both are discussed in Chapter 5 as being types of *topic-shift*. I argue here that these instances of HTT still mark a specification, refocusing the discourse onto the (re)introduced topic by emphasizing the importance of the first contribution. Topic-shifts and specification are thus not mutually exclusive discourse moves. Topic-shifts are related to *questions* – they signal a premature closure of an established question and the posing of a new question. Specifications are related to *answers* – they signal that a contribution offers an answer or partial answer to an open question.

The vast majority of topic shifts aligned with HTT (30/35, 86%) are *return* shifts in which a still-open discourse question is returned to after a digression. In this case, the emphasis expressed by HTT is multifunctional – it both draws attention to reestablishing a particular discourse goal, *and* highlights the importance of the particular contribution to achieving that discourse goal. In other words, the importance of the contribution, as expressed by HTT, is that it establishes a discourse goal distinct from that of the immediately preceding discourse segment. Of the 30 instances of return shifts identified in the data set, 23 occur during Colbert's monologues. As was the case with return shifts aligned with *anyway*, these most often occur after Colbert performs a short joke or pun. One such examples is given in (147).

In this clip, Colbert is describing a recent "culture war" scandal in which the state of Alabama refused to air an episode of an animated children's show called Arthur because it featured a marriage between two male characters. Colbert introduces the topic by describing what happened in the episode (lines 1-2). He then briefly digresses from the story to make a sexually-charged "don't google it" joke (lines 3-4). Colbert returns to the main topic, marking the return with a HTT-marked specification (line 4), and describing the scandal, namely Alabama's unreasonable response (lines 5-7).

(147) **Return (monologue)**

TRANSCRIPT 8: STEPHEN COLBERT

[UID:10ce4d02-2232-11ea-9af5-089e01ba0770,798]

```
1  SC   Last week, PBS aired an episode where Arthur's teacher Mr.
2       Ratburn got married to a chocolatier named Patrick the
3       Aardvark.  Quick tip:  if you're wondering how a rat and an
4       aardvark get it on, don't google it.  Here's the thing, your
5       kids may have watched it last week.  Unless your kids live in
6       Alabama, because Alabama public television refused to air an
```

7 Arthur episode with a gay wedding.

The discourse segment before the digression provides background information necessary for understanding Alabama's reaction as homophobic, which is the point of reporting on the story. Interestingly, it is not the main point, expressed in lines 5-7, that is prefaced by HTT. Instead, HTT marks the initial return to the topic.

Example (148) demonstrates a functionally similar return during an interview. In this clip, American actor Bradley Whitford and Colbert are discussing the increased polarization in American politics. Whitford begins to argue that this is primarily the fault of the Republican party which has, in his words, "gone bat shit crazy" (lines 2-3). He then begins contrasting the current state of the Republican party with previous generations, mentioning how George W. Bush, who seemed incredibly conservative at the time of his presidency, now appears liberal in comparison (lines 5-6). Colbert then interrupts Whitford's turn, acknowledging the humorous presentation of the comparison (line 7), encouraging its extension (line 8), and providing a comparison of his own (line 10). Whitford then regains the turn and returns to elaborate on his claim that the Republican party has taken a dangerous turn, marking both the initiation of his turn and the return to the main topic with HTT (line 12).

(148) **Return (interview)**

TRANSCRIPT 9: BRADLEY WHITFORD

[UID:68b70a26-1573-11e9-a95d-089e01ba0335,2635]

1 BW I do not think uh that the party is swinging too far uh to
2 the left, I think the right has gone absolutely bat shit
3 crazy. Uh and uh
4 -- (*cheers*)
5 BW I mean, y'know, you l- (*laughs*) uh George Bush looks like
6 Abbie Hoffman now, he's like y'know-
7 SC I know isn't it crazy
8 BW Yeah, he's painting his toes
9 [y'know. Yeah, yeah]
10 SC [Exactly. Nixon, Nixon] for Pete's sake. Nixon started the
11 EPA
12 BW Here's the thing is they have become the party of
13 accumulation for the few and corruption. And I firmly
14 believe that any Democrat who I can see possibly running is
15 what this country's all about, which is trying to get to the-
16 fulfilling the unfulfilled promises of opportunity for all,
17 and I think the Democrats will do that

Unlike the previous example, the HTT-marked specification here does mark the specification that is most important to Whitford's initial claim – the Republican party is "crazy" (line 2) because they are now "the party of accumulation for the few and corruption" (lines 12-13).

This means that the contribution that reintroduces the ongoing topic after the digression is one in the same as the specification that provides a particularly crucial argument.

If the digressions were removed, the HTT-marked specification in both cases would be categorized as an *elaboration*. The segments prefaced by HTT provide additional information about an ongoing topic, but not in a way that clearly aids in the (re)assessment of an adjacent segment. I argue that this suggests a dual function of HTT in these contexts, simultaneously emphasizing the return to the main discourse, and the contribution as being particularly important for ongoing discourse goals.

In addition to return shifts, I judged five topic-shifts aligned with HTT as constituting full shifts. Full shifts are distinct from return shifts in that the introduced topic is completely new to the discourse. I've included one particularly interesting example in (149). In this example, Colbert has just reintroduced American actor Jon Hamm to the audience after a commercial break. Colbert begins by returning the topic they had been discussing before the break – all of the movies Hamm has been in recently (lines 1-2). Initially, it seems that this is the topic that will be continued. Hoever, after a brief back and forth in which the discourse seems to stall (lines 3-5), Colbert changes both the topic and the structure of the interaction, marking the shift with HTT. The statement "I just found out that we're in these movies together" (lines 6-7) is not actually true, instead it serves as a segue into a cooperative improvisation in which Colbert and Hamm talk about a series of imaginary movies, each corresponding to a fake movie poster.

(149) **Full shift**

TRANSCRIPT 10: STEPHEN COLBERT TO JON HAMM
[UID:42d09d7e-e73a-11e9-be1d-089e01ba0335,2541]

1	SC	You're always working, and I understand you have a whole
2		bunch of other movies coming out right now too
3	JH	We kinda shoot 'em like, y'know, back to back to back-
4	SC	You got to
5	JH	Uh- I- well, y'know, I'd rather work
6	SC	But here's the thing that alarms me, is that I just found
7		out that we're in these movies together

This is considered a full shift because, though the two topics are *thematically* related, they contribute to distinct discourse goals. The segment prior to HTT contributes to a discourse goal of understanding Hamm's busy acting schedule. The segment after HTT contributes to a discourse goal of achieving cooperative improvisation. It is not possible to identify a single question that would unite these two topics as sub-goals under the same ongoing discourse topic, aside from something as general as "what are we doing in this interview?".

In each of these cases HTT serves to emphasize a particular contribution. In (148), this emphasis is placed on a contribution that is, indeed, important to an open discourse question. This looks very much like the other types of specification discussed. In the remaining two, however, the emphasis signalled by HTT serves to draw attention only to a shift in discourse goals rather than the actual information provided in the specification.

When aligned with topic-shifts, we may expect to see gestures that signal both a dismissal and initiation of a topic, as in Chapter 5. Topic dismissal gestures would include *stopping* gestures and *away* gestures. These are distinct from the gestures expected for specifications. Topic initiation gestures include presentation gestures and engagement with the Interaction Space, both of which are expected for specifications. In Chapter 5 we saw that gestures aligned with topic shifts can express dismissal, initiation, or both in sequence. In topic-shifts with HTT, we may expect more initiation gestures as the discourse marker serves to emphasize the topic initiation rather than the topic dismissal.

7.3.2 Lexical variation

In the majority of the data set, HTT is preceded by at least one other lexical discourse marker (62/100). 13 distinct lexical discourse markers were recorded, sometimes occurring in combination with each other (e.g. "but y'know here's the thing"). This degree of variation is unsurprising given the functional variation discussed in the previous section. All accompanying lexical discourse markers occur immediately before HTT with the exception of *though*, which occurs immediately after. *But* is by far the most common, occurring in approximately one quarter of the data set (26/100)[8]). Table 7.3 gives the distribution of all HTT-adjacent discourse markers that occur more than once in the data set according to the specification type. The six lexical discourse markers that appear only once in the data set (*alright, because, look, so, um,* and *y'know*) are combined under "*other*" in the table.

In discussing the effects that these additional discourse markers may have on our predictions of gestural expression, I will focus on *and, but* and *well* because of their frequency in the data set and the amount of attention they have been paid in the literature on discourse markers.

And, as a lexical discourse marker, is very frequent and has proven difficult to analyze, especially in relation to its use as a logical connective. Schiffrin (1986) calls for an underspecification account, arguing that *and* serves as a default connective in narratives and signals the continuation of a speaker's actions, whatever they may be.[9] Given this underspecification, we may expect *and* to occur most frequently with the most underspecified of our observed specification types, *elaboration*. This is indeed the case, but not overwhelmingly so. We also expect it *not* to occur with contradictions, qualifications, or topic-shifts which all call for a *dis*continuation of some action. This also proves to be true in this data set, with the exception of one instance in which *and* occurs with a return. Because of its continuative function, we may expect presentation gestures to co-occur with *and*, with each presentation corresponding to the addition of some idea. This aligns with the predictions for gestural expressions of specification more generally.

[8]Three of these 26 occur with indeterminate relations, hence the count of 23 in Table 7.3.

[9]Nevile (2006) and Dorgeloh (2004) reach similar conclusions. Though also broadly in agreement, Turk (2004) further argues that *and* in naturalistic conversation serves a 'smoothing' function, reinforcing cohesion when it is for some reason at risk. Turk's analysis aligns with the use of *and* in specification in which the speaker is asserting the importance and relevance of a particular contribution to the ongoing discourse.

	Background	Circumstance	Contradiction	Elaboration	Qualification	Return	Full Shift	**Total**
N/A	8	2	2	3	1	13	3	**32**
though	-	1	-	1	-	1	-	**3**
and	2	1	-	4	-	1	-	**8**
but (y'know)	-	8	4	5	-	5	1	**23**
no (but)	-	3	1	-	-	-	-	**4**
now	1	-	-	-	-	4	1	**6**
okay (so)	1	-	-	-	1	3	-	**5**
well	2	-	3	-	1	1	-	**7**
other	-	1	-	2	-	3	-	**6**
Total	**14**	**16**	**10**	**16**	**3**	**30**	**5**	**94**

Table 7.3: Co-occurring lexical discourse markers by relation

But, in its role as a lexical discourse marker, serves a cancellative function (Bell 1998). As such, we might expect *but* to occur most frequently with contradictions, which serve to refute an earlier contribution. Though *but* does occur in 40% of contradictions (4/10), it is most frequent, proportionately, with circumstance relations. This is unsurprising given that circumstance specifications often call for a reassessment of a previous discourse segment. *But* in the context of a circumstance relation can thus be said to reinforce the need for reassessment – an initial interpretation must be cancelled and replaced. The cancellative function of *but* would not be expressed by the presentation and locating gestures predicted for specifications more generally. If a gesture aligned with "but here's the thing" expresses the cancellation of *but* rather than the specification of HTT, we would expect to see movements associated with dismissal. This would include *stopping* gestures, *away* gestures, and disengagement from the Interaction Space.

Though *well* is among the most well-studied English discourse markers (Cuenca 2008:1373), there is little consensus as to what extent it can be said to have a single coherent set of functions. Those studies that *do* attempt to identify an underlying unifying function often argue for a 'heteroglossic' analysis in which *well* signals the recognition of multiple viewpoints and asserts the relevance of the proceeding statement in light of those multiple viewpoints (e.g. Aijmer & Simon-Vandenbergen 2003; Cuenca 2008; Le Lan 2007). This analysis is supported by the fact that all seven instances of *well HTT* occur in interviews in turn-initial position. This function also aligns well with the emphasis signaled by HTT and may serve to profile the move as renegotiating some aspect of the discourse structure, as is the case in *circumstance*, *contradiction*, and *qualification* relations. Though it does occur three times with a contradiction, it does not appear with circumstance relations in the data set. Additionally, it occurs twice with a *background* relation, which may serve to signal that the information was being taken for granted, but is now considered primary in light of different viewpoints and knowledge states. If *well* does serve as an acknowledgement of multiple viewpoints, we may expect *addressing* gestures, such as reaching toward the addressee, to occur in order to reinforce that acknowledgement gesturally.

7.3.3 Phrasal variation

In an overwhelming majority of instances (89/100), HTT appears as an independent phrase, prosodically separated from the utterance it prefaces. Exceptionally (6/100 instances), HTT appears with an extension, such that "the thing" is modified by a relative clause (ex. 149, Transcript 10), an infinitival clause (ex. 150a, Transcript 11), or some other adverbial phrase (ex. 150b, Transcript 12).

(150) a. Transcript 11: Nancy Pelosi
[UID:6a02fce6-159f-11ea-8a36-089e01ba0335,1991]

```
1  NP   My view was that it was perfectly wrong,
2       [and- but- but- but]
3  --        [(cheers)]
```

```
4   NP   here's the thing to remember, they make a big thing of
5        saying "well if it was so wrong, why would the president
5        put out the notes from the meeting".  It was a whistle
6        blower.  We would never have known about this absent the
7        whistle blower coming forward.
```

b. Transcript 12: Stephen Colbert to Scott Pelley
[UID:108c951e-7937-11e9-8bd9-089e01ba033,2416]

```
1   SP   You've got the best audience in television
2   SC   So important.  So few- here's the thing about you, so
3        few newsmen know the value of pandering
```

There appears to be a functional distinction between independent and extended HTT uses. As argued in this work, independent HTT is primarily used as an attention marker, signalling to the addressee that the proceeding specification is particularly important for achieving discourse goals. Extended variants also work in this way, but additionally convey particular information about how the utterance relates to the discourse's structure. This is especially apparent in (150b) where the phrase "here's the thing about you" clarifies the relevance of Colbert's utterance to the interaction. In the discourse, American journalist Scott Pelley has just complimented the audience. Colbert begins to make a general statement about the importance of "pandering" in journalism, but interrupts himself to specify that this statement is meant to compliment Pelley and the effort he just made to pander to the audience.

Lastly, there are five instances in which HTT appears in a grammatical blend (Fauconnier & Turner 1998) with *the thing is*, as is the case in (151) below. One may argue that these are speech errors. It is possible that the structural and functional similarity of the two phrases is conducive to such confusions. I do not take a stance here as to whether these are speech errors or 'legitimate' grammatical blends. I mention them only as a structural variant observed in the data that demonstrates the relationship between HTT and "the thing is".

(151) Transcript 13: Stephen Colbert
[UID:d6941018-a793-11e9-902d-089e01ba0335,1392]

```
1   SC   That's a summer flounder right there.  Unbelievable.  Just
2        as- just as speckled- just as speckled as the night sky right
3        there.  And here's the thing is that I- I wear a bandana when
4        I'm fishing
```

There are too few instances of extended HTT phrases in the present data set to meaningfully identify differences in gesture performance. However, it is possible that particular phrasal continuations may result in exceptional gestures. For example, in (150b), Colbert may perform an addressing gesture (e.g. pointing an up turned hand toward his addressee), to express the source of the topic ("you"), rather than the specification move itself.

7.3.4 Turn regulation

As we have seen throughout the examples so far, HTT can occur both turn-initially (ex. 140,143,145,148,150a) and turn-medially (ex. 142,144,146,147,150b,151). Out of the 38 instances of HTT during interviews, HTT is used at the beginning of the turn 22 times. This suggests that HTT can serve as a turn-entry marker during dialogues (Sacks et al. 1978; Tao 2003). One may argue that emphasizing the importance of one's contribution with HTT also emphasizes one's role as speaker and contributor. This is especially apparent in contexts of overlapping turns where the speaker role is being actively negotiated, such as in (141).

Perhaps counter-intuitively, HTT can also function as a turn-entry device during monologues. For example, HTT is frequently used immediately after ending an impersonation (5 instances), imaginary interaction (6 instances), or a video display (3 instances). I argue that in all these instances HTT is functioning *as if* it is turn-initial. Consider, for example, the end of the impersonation in (152). In this clip, Colbert has just displayed a video of Trump claiming that he doesn't hold the opinion that Mexico is run by "the cartels and the drug lords", a problematic claim he had made or suggested many times in the past. When the camera cuts back to Colbert on stage, Colbert highlights this hypocrisy through a humorous impersonation (lines 1-4). Colbert than returns to the main topic regarding the negative impact of increased tariffs on Mexican imports, marking this return with HTT. In this example, HTT can be said not only to emphasize the return to the main topic, but also the return of Colbert to his role as show host.

(152) Transcript 14: Stephen Colbert
[UID:d3618b7c-b5b8-11e9-9749-089e01ba0335,502]

```
1  SC   "Yeah, a lot of people are saying that.  For instance, me
2       just now saying that.  Who else?  The guy I shave in the
3       morning in the mirror, he always says it, then he calls me
4       bald and paints me orange" H- Here- here's the thing, Mexico
5       is America's number one trading partner, and the economic
6       consequences of these tariffs could be huge
```

As discussed in Chapter 3, turn-structure is expected to have a direct impact on gesture behavior. Engage actions are expected at turn-beginnings to mark the participant's role as speaker and main contributor. Thus, we expect postural shifts toward the central Interaction Space to occur in cases where HTT marks the beginning of a turn. These actions include turning and leaning toward the interlocutor. It is also possible that placement within the turn-structure will affect the likelihood of presentation versus locating gestures with HTT. Where HTT occurs at turn-beginnings, presentation gestures are independently more likely, as the speaker is making their first contribution to the Interaction Space for that given turn. Where HTT occurs within a turn, on the other hand, we may expect more locating gestures as the speaker is more likely to have already introduced the topic related to the specification into the Interaction Space.

7.3.5 Interim summary

I identified seven functional contexts in which HTT-marked specifications occur. Some of these do not change our predictions as to gestural expression. In particular, presentation and locating gestures are expected for *background*, *circumstance* and *elaboration* relations, just as they are for specification moves more generally. However, the remaining functional contexts do change our predictions. First, stopping gestures are predicted for contexts in which the speaker either ends a possible line of inquiry (*contradiction*) or requests the prevention of a negative assessment (*qualification*). Second, *away* gestures would be unsurprising in the context of topic-shifts, given the discussion of topic-dismissal gestures in Chapter 5 on *anyway*. However, because HTT still functions as a specification marker, even in these contexts, we might expect *fewer* dismissal gestures than in other topic-shifting contexts, such as in the presence of the discourse marker *anyway*. This is because the (re)introduction of a discourse topic, rather than the dismissal of one, is emphasized.

Immediately adjacent lexical items may also effect gestural expression. Where there is an immediately preceding discourse marker (e.g. "but here's the thing") or proceeding discourse marker (e.g. "here's the thing though"), the accompanying gesture may profile a part of the discourse move more closely related to the other discourse marker. For example, a gesture occurring with "but here's the thing" may profile the cancellative aspect of the discourse move, co-expressed by *but*, rather than the specification aspect expressed by HTT. Where the HTT phrasal unit is extended (e.g. "here's the thing that alarms me"), the accompanying gesture may co-express some aspect of the expression, such as a self-directed deictic gesture, rather than the specification move.

Finally, turn-structure is predicted to influence gesture behavior. Because HTT occurs turn-initially and turn-medially, but not turn-finally, the influence of turn-structure on gesture behavior for specifications may not be particularly noticeable – we expect engagement and gestures into the central Interaction Space for both. However, turn-initial HTT may encourage more presentation gestures as the speaker offers the first contribution of their turn to the discourse.

7.4 Variation in the gestural mode

Given the function of a specification move to introduce new information about an ongoing topic, and the proposed function of HTT to mark a given contribution as particularly important for achieving discourse goals, I predicted (i) consistent or heightened engagement with the Interaction Space, and (ii) presentation and locating gestures into the central Interaction Space. In the present data set, these gestural expressions of specification are indeed recurrent. However, we also see *addressing* and *stopping* gestures. The form and function of each type of gestural expression observed recurrently is summarized in Table 7.4. I will discuss *presentation, locating* and *stopping* gestures separately and in detail. *Addressing* gestures and forms of *engagement* will be discussed as they occur in presented examples.

Type	Description	Section
Presentation	hand gestures that literally or metaphorically introduce an object into the Interaction Space as a discourse topic.	*7.4.1*
Containment	two-handed gesture with palms oriented toward each other, as if to hold a rectangular object	
Pinch	one-handed gesture with fingers bunched, as if to hold a small delicate object	
PUOH	one- or two- handed gesture with open palms facing upward, as if to hold a small object up for inspection	
Refferring	hand gestures that deictically that draw joint attention to discourse participants or topics as metaphoric objects in the Interaction Space	
Locating	deictic gestures toward regions other than addressee and lacking the physical affordances of presentation	*7.4.2*
Addressing	deictic gesture toward addressee and lacking the physical affordances of presentation; used to elicit a response	*7.4.3*
Stopping	hand gestures that request for some action to be stopped through either iconicity (*blocking*) or convention (*pausing*)	*7.4.4*
Blocking	one- or two- handed gesture with open palms facing outward as if to hold an object away from the body	
Pausing	one- or two- handed gestures with index finger upward, as if to indicate the numeral "1"	
Engagement	non-manual orientation toward the speaker-hearer line including turning gaze, head, and body toward addressee; used to signal willingness to contribute to the discourse	–
Heightened engagement	non-manual gestures decreasing the space between participants, such as leaning forward while already oriented toward addressee	

Table 7.4: Reccurrent gestures accompanying HTT

Of the 100 instances of HTT analyzed, only 9 do not align directly with a manual gesture. Of these 9, 4 exhibit manual gestures during the specification immediately after HTT. 80 of the manual gestures aligning with HTT were categorized as one of the four gesture classes of interactive gesture described above. The distribution of these gestures according to class and interaction type is given in Table 7.5. The remaining 10 gestures aligned with HTT were categorized as self-adaptors (3), representational (3), or uncategorizable (4).[10]

	Presentation	Locating	Addressing	Stopping	**Total**
Monologue	7/52 (13%)	3/52 (6%)	10/52 (19%)	18/52 (35%)	**38/52** (73%)
Band interaction	1/9 (11%)	1/9 (11%)	5/9 (56%)	-	**7/9** (78%)
Interview	7/39 (18%)	12/39 (31%)	10/39 (26%)	6/39 (15%)	**35/39** (90%)
Total	**15/100** (15%)	**16/100** (16%)	**25/100** (25%)	**24/100** (24%)	**80/100** (80%)

Table 7.5: Gestures accompanying HTT by discourse type

Initially, these findings are surprising – the majority of observed interactive gestures are *addressing* or *stopping* gestures (49/80), neither of which are predicted for specification moves. However, of these 49 'unexpected' gestures, 17 are immediately followed by either a presentation gesture (13 instances) or a locating gesture (4 instances), the two classes of interactive gesture that were initially predicted for specification moves. I suggest that this supports a multifunctional analysis of HTT. In the introduction, it was proposed that HTT is used as an attention marker to signal that a proceeding specification is especially important for achieving discourse goals. Under this analysis, HTT can be said to perform two simultaneous functions: (i) alerting the addressee that they should pay attention; and (ii) marking the segment they should pay attention to as a specification move. I argue that stopping and addressing gestures express the first of these functions, whereas presentation and locating gestures express the second. This will be further explored in Section 7.4.4 in which I analyse two extended gesture sequences.

There are also several noticeable differences in the gestures performed across interaction types. First, the overall rate of interactive gesture is highest in interviews (90% versus 73%

[10]Observed self-adaptors included buttoning a jacket and itching the chin. Representational gestures included a cancelling gesture affiliated with an adjacent semantic negation (Harrison 2010; Kendon 2004), and a pantomime of talking on the phone. Uncategorizable interactive gestures included slapping the desk and moving hands from a rest position to a clasping position.

and 78%). This is unsurprising given that this is the one interactive context with a symmetric Interaction Space, that is a space in which the two participants have equal access to Interaction Space management. The overall rate of interactive gesture is lowest during monologues (73%). This is again unsurprising given that Colbert is the only participant with full access to Interaction Space management in these contexts. There are fewer reasons to use interactive gesture to manage the interaction when Colbert has full control over the interaction. Both of these observations align with previous work demonstrating that interactive gestures occur most frequently in face-to-face interaction, where a physical Interaction Space is being actively maintained by participants (Bavelas et al. 1992, 2008). In other communicative settings, such as on the phone or during a tape-recording, interactive gestures still occur, but at a significantly lower rate. In these contexts, much like in Colbert's interview, there is only one participant actively maintaining a largely imagined Interaction Space.

Second, the majority of gestures performed during an interaction with the band lead, Jon Batiste, were addressing gestures, all performed by Colbert. I suggest that this is related to the general monologue format and Batiste's role on the show. During the monologue segments of the show, the primary Interaction Space is between Colbert and the audience. To activate the secondary Interaction Space between him and Batiste, he must mark the relatively unexpected shift in space. Additionally, Batiste's primary role on the show is that of band leader, *not* of an active discourse participant. To facilitate the shift in role from band leader to discourse participant, Colbert uses addressing gestures to, in a sense, invite Batiste's participation in the discourse.

Third, locating gestures occurred most frequently in interviews. This makes sense when one considers the physical affordances of the different Interaction Spaces. The Interaction Space in interviews is fully bound by the two active participants, and consists of a consistent space (the desk) on which to manage (and thus locate) Interaction Space contents. During monologues, especially standing monologues, there is a larger, more loosely defined space – Colbert walks around the stage and reorients himself frequently. This flexibility likely impacts the likelihood of a topic being located and referred to in a single region of space.

Finally, stopping gestures occurred most frequently in monologues. This finding is particularly interesting given the general function of stopping gestures – to stop some line of action (e.g. Bressem & Müller 2014, 2017; Bressem & Wegener 2021). Their frequency in monologues suggests Colbert's use of stopping gestures is a form of self-management. In Chapter 4, I argued that *blocking* gestures in self-management contexts enact a barrier between the speaker and unwanted objects (even when those objects were introduced by the speaker). This is also the case in blocking gestures used for addressee-management. However, addressee-directed blocking gestures have the additional physical affordance of 'stopping' the metaphoric forward motion of the addressee, evoking the metaphor Speaking is Forward motion (Wehling 2017). There is also some evidence that the number of stopping gestures in monologues is inflated by an idiosyncratic gestural construction in which Colbert performs a pausing gesture immediately followed by a precision grip presentation gesture.

We can better understand the observed gestural variation, especially results that seem

initially surprising, by considering other contextual factors. Section 7.3 reviewed the structural and functional variation of HTT observed in the present data set and how that variation may manifest in particular types of gestural variation. In particular, *stopping* gestures were predicted for *contradiction* and *qualification* relations. Where the HTT-marked specification occurred immediately after a topic-shift, dismissal (*away* and *stopping*) and initiation (*presentation*) gestures were predicted. These contextually mediated predictions are only partially supported by the observed gesture variation in the data set. This suggests that underlying discourse relations may not play a particularly important role in determining the form of gestural expressions of specification. This would mean that gestural markers of specification are underspecified for discourse relation, much like the lexical marker HTT. Table 7.6 summarizes gestural variation according to discourse structural context. Despite inconsistencies with predictions, and the relatively small number of gestures occurring with each discourse relation, there are several interesting trends worth noting.

First, the *background* relation had the lowest rate of accompanying interactive gesture (9/14; 64%). If background information is considered already part of the common ground, as it is in (140), then there may be less motivation to mark it with an interactive gesture. This is reminiscent of Enfield et al.'s (2007) work on variation in the articulation of points. They showed that when points provided background information, information that was assumed to be in the common ground of both participants, the points were smaller and less central, and argued that under-articulation served a kind of face-saving function – to draw too much attention to information your interlocutor already knew would be rude. Of the 9 interactive gestures that do align with HTT and a background relation, the majority (5/9) are locating gestures. This further suggests that the information is, in a sense, already present in the discourse and common ground.

Second, elaborations occur most frequently with stopping gestures (7/16; 44%). This is surprising given the function of elaborations to simply add additional information to an ongoing discourse topic. However, there are two interesting things about these six examples that may account for their gestural exceptionality. First, 6 of the 7 stopping gestures with elaborations are *pausing* gestures in which a speaker holds their index finger pointing upward as if to say "wait a second". These are distinct from *blocking* gestures which request that some action be permanently stopped. The one *blocking* gesture occurred with a "but HTT" sequence, which means that the blocking gesture may be expressing the cancellative function of "but" rather than the elaboration relation. Second, all 6 of these examples are accompanied by an additional preceding discourse marker. There are two "and HTT" and one "so HTT" examples. The stopping gestures in these three examples are immediately followed by a presentation gesture, the type of gesture we expect to see with elaboration relations. The remaining three examples occur with "but" and are *not* followed by a presentation gesture.

	Presentation	Locating	Addressing	Stopping	**Total**
Background	-	5/14 (36%)	3/14 (21%)	1/14 (6%)	**9/14** (64%)
Circumstance	1/16 (6%)	2/16 (13%)	6/16 (38%)	6/16 (38%)	**15/16** (94%)
Contradiction	3/10 (30%)	2/10 (20%)	1/10 (10%)	3/10 (30%)	**9/10** (90%)
Elaboration	4/16 (25%)	1/16 (6%)	3/16 (19%)	7/16 (44%)	**15/16** (94%)
Qualification	-	-	1/3 (33%)	1/3 (33%)	**2/3** (67%)
Return Shift	5/30 (17%)	5/30 (17%)	9/30 (30%)	3/30 (10%)	**22/30** (73%)
Full Shift	1/5 (20%)	1/5 (20%)	-	2/5 (40%)	**4/5** (80%)

Table 7.6: Gestures accompanying HTT by functional context

We also predicted turn-position to influence gestural expression. Most notably, presentation gestures are expected in turn-initial, regardless of discourse move or relation, as participants mark their role as speaker and make their first contribution to the Interaction Space. This is partially supported by the data, as summarized in Table 7.7. During interviews, presentation gestures are used more frequently in turn-initial position than in turn-medial position. However, *no* presentation gestures occur in 'initial'-position in monologues. As noted turn-'initial' in monologues is qualitatively different in that Colbert is not taking the turn from a co-present interlocutor, but rather shifting between imagined dialogues and video clips. There are no other particularly noteworthy turn-structural patterns.

		Presentation	Locating	Addressing	Stopping	**Total**
Monologue	*initial*	-	2/14 (14%)	4/14 (29%)	3/14 (21%)	**9/14** (64%)
	medial	7/38 (18%)	1/38 (3%)	6/38 (16%)	15/38 (39%)	**30/38** (79%)
Interview	*initial*	6/24 (25%)	9/24 (38%)	4/24 (17%)	3/24 (13%)	**22/24** (92%)
	medial	1/15 (7%)	3/15 (20%)	6/15 (40%)	3/15 (20%)	**13/15** (87%)

Table 7.7: Gestures accompanying HTT by turn position

These summaries of observed gesture variation provide helpful initial insights into the extent of variation in gestures accompanying HTT. However, to actually understand how *specification* is being expressed in the gestural mode, we have to consider not only the gesture aligned with HTT itself, but also the gestures occurring during the specification. In the remainder of this section I will demonstrate how the four classes of gestures are used, often in concert with one another, to profile different discourse structural and social aspects of individual specification moves.

7.4.1 Presentation

In this section, I discuss three cases of presentation gestures aligned with HTT or the immediately proceeding specification, one for each variant of presentation described in Table 7.4 and depicted in Figure 7.1. All three variants introduce a metaphoric object into the Interaction Space, and so all are expected to occur during specification moves. However, the implied physical properties of the metaphoric object introduced are different for each type. We expect these formal differences to be indicative of functional differences.

Figure 7.1: Types of presentation gestures as observed with HTT

Containment gestures imply a medium-sized flat-sided object with clearly defined edges, such as a box. The containment gesture is the only type of presentation gesture that is obligatorily two-handed, meaning that a *region* of space is demarcated rather than a single position. This allows further manipulation of the "contained" region such that metaphoric *contents* can be individually located, organized, and removed. This affordance is particularly conducive to argumentation in which an argument (the bounded region) is made through a series of examples and particular points (the metaphoric contents). It is also the only type of presentation gesture that refers specifically to the object's edge, emphasizing a metaphoric boundary between the topic presented and anything else. This affordance can be exploited to express contrast by dividing the space into separate bounded regions (e.g. Calbris 2003, 2008; Hinnell 2019). This is demonstrated in example (161) below.

Pinching gestures imply a very small singular object. Frequently the gesturer leans in while making this gesture as if to get closer to the small object for inspection. Kendon (2004: Ch. 12, Streeck 2008) argues that variants of the pinching gesture (which Kendon terms 'precision grip' gestures) are used to emphasize precision in, for example, the argument being made or the information being requested. As such, pinching gestures are conducive to expressing the particular importance of a single point, especially amidst other potential arguments. We will see a particularly clear example of this (ex. 157) in which the two speakers, Colbert and American politician Elizabeth Warren, use pinching gestures while contesting an appropriate and honest answer to a very specific question.

Palm-up open-hand gestures (PUOH) have been noted for their pervasiveness across cultures and contexts (Cooperrider et al. 2018; Müller 2004). Perhaps related to their apparent polysemy, PUOH gestures are also the most underspecified presentation gestures as to properties of the metaphoric object presented. Potentially, PUOH gestures imply a smallish object that can be held in the palm, though two-handed variants of the gesture complicate this. Unlike the other two presentation gestures, no metaphoric information about shape or delicateness is conveyed.

Given these characteristics, we can say that (i) containment gestures are particularly well-suited for presenting complex arguments and establishing distinctions between concepts, (ii) pinching gestures are particularly well-suited for presenting specific points for careful consideration, and (iii) PUOH gestures are well-suited for presenting ideas, but not for conveying specific information about the nature or intent of the contribution.

These three types of presentation gesture occur with HTT 16 times in the data set. An additional 15 presentation gestures occur immediately after HTT, during the specification. The distribution of presentation gesture type and timing is given in Table 7.8.

	Containment	Pinch	PUOH	**Total**
HTT	3	3	9	**15**
Specification	5	8	3	**16**
Total	**8**	**11**	**12**	**31**

Table 7.8: presentation gestures accompanying HTT and the proceeding specification

The variation in timing of the three types of presentation gestures is indicative of functional differences. PUOH gestures largely align with HTT, whereas pinching gestures largely align with the specification. This aligns well with the differences in affordances described above – PUOH gestures function well to present an idea and draw attention to a general contribution, whereas pinching gestures draw attention to specific information within a contribution. Containment gestures do not seem to strongly favor one alignment over the other. Interestingly, two of the containment gestures aligned with the specification, but not HTT, immediately follow a PUOH gesture aligned with HTT. This again supports the underspecification of PUOH gestures. This transition from more schematic to more semiotically enriched gestures is demonstrated several times in the following examples.

7.4.1.1 Containment

The first example I'll discuss demonstrates both of the aforementioned affordances of containment gestures: (i) the ability to "contain" complex arguments, and (ii) the ability to demarcate space into separate, meaningful, and mutually exclusive spaces. In this clip, Colbert is discussing Republicans' attempts to deligitimize the impeachment trials against then-president Donald Trump. It begins with Colbert displaying a recent tweet from Trump which Colbert reads out in an exaggerated impersonation of Trump (lines 1-3). When he is finished reading the tweet, the camera returns to Colbert standing on stage. Colbert acts as if to collect his thoughts in order to respond calmly, looking down and clasping his hands together as he says "okay" (line 3, CG1). He then begins his rebuttal of Trumps tweet, marking the contradiction-specification with HTT and a two handed blocking gesture (line 3, CG2). Colbert begins by specifying the truth, "there's a non-partisan stenographer in the room", bringing his hands down and to the right into a containment gesture (line 4, CG3). In the same space, Colbert then performs a representational gesture depicting the stenographer's typing (line 5, CG4). After this, he explicitly contradicts Trump's claim that there are partisan versions of the trial transcript (lines 5-6). Using a series of containment gestures, Colbert positions the "Democratic" version at his lower right (CG5) and the "Republican" version to his lower left (CG6). He repeats this sequence for his analogy to partisan physics (lines 6-7, CG7 & CG8, not depicted).

(153) TRANSCRIPT 15: STEPHEN COLBERT
[UID:c7cce1dc-ff9e-11e9-9f3f-089e01ba0335,728]

```
1  SC   "Republicans, dot dot, dot dot dot dot dot, should give their
2       own transcripts of the interviews to contrast with Schiff's
3       manipulated propaganda." *Okay (.)* *Here's- here's the
                                  *CG1- - -* *CG2- - - - - - - >
4       thing,* *there's a non-partisan stenographer in the room*
        <- - -* *CG3- - - - - - - - - - - - - - - - - - - - - *
5       *typing every word.* *There's not a Democratic* *or
        *CG4- - - - - - - -* *CG5- - - - - - - - - - -* *CG6->
6       Republican version of the transcript,* *any more than there
        <- - - - - - - - - - - - - - - - - - * *CG7- - - - - - - ->
7       is a Democratic* *or Republican laws of physics.*
        <- - - - - - - * *CG8- - - - - - - - - - - - - -*
CG1:  two hands, fingers intertwined, held to stomach
CG2:  two hands, flat palms out, chest level
CG3, CG5, CG7:  two hands, containment, held center-right
CG4:  two hands, typing pantomime with lateral movements
CG6, CG8:  two hands, containment, held center-left
```

Figure 7.2: Gesture sequence corresponding to TRANSCRIPT 15: COLBERT, ex. 153

The blocking gesture performed during HTT itself profiles the contradiction relation – Colbert signals a refusal to accept Trump's claims by metaphorically blocking them and requesting their removal from the Interaction Space. This is represented formally in (154). Given that this example is from a monologue, no turn-transition occurs, and the requested Interaction Space state can be considered automatically granted.

(154) **Management actions by Colbert, ex. 153, Line 2**

$$IS1_{t} = \begin{cases} \text{Participants: } SC_S, \text{ Audience} \\ \text{Content : Trump's claims} \\ \text{Management :} \\ \quad \textsc{Request}(\text{CG2}): IS1_{t+1} \end{cases} \qquad IS1_{t+1} = \begin{cases} \text{Participants: } SC_S, \text{ Audience} \\ \text{Content : } \sout{\text{Trump's claims}} \\ \text{Management :} \\ \quad \textsc{Remove}(\text{G} \rightarrow \\ \quad\quad \text{Trump's claims}) \end{cases}$$

With a cleared Interaction Space, Colbert can begin to build his rebuttal. The first containment gesture (CG3) then introduces Colbert's main point to the Interaction Space – there is a non-partisan transcript written by a non-partisan stenographer. Information is added about this transcript by gesturing in the same space, as he does when pantomiming "typing every word".[11] Colbert then performs a second containment gesture (CG5), introducing the concept of a "Democratic" transcript. Importantly, this presentational gesture is performed in the same region of the Interaction Space as the first containment gesture. I argue that this COMBINES the two topics, expressing some shared attribute or purpose. This management action is the inverse of the SEPARATE action that was introduced in Chapter 3 and serves to express some contrast in attribute or purpose. These actions are represented in (155).

(155) **Management actions by Colbert, Lines 4-5, ex. 153**

$$IS1_{1} = \begin{cases} \text{Participants: } SC_{\mathbf{S}}, \text{ audience} \\ \text{Content : } +\text{Transcript}_{\text{Real}} \\ \quad +\text{Transcript}_{\text{Dem}} \\ \text{Management :} \\ \quad \textsc{Present}(\text{CG3}\rightarrow\text{Transcript}_{\text{Real}}) \\ \quad \textsc{Present}(\text{CG5}\rightarrow\text{Transcript}_{\text{Dem}}) \\ \quad \textsc{Combine}(\text{CG3}+\text{CG5}\rightarrow\text{Transcript}_{\text{Real}}+\text{Transcript}_{\text{Dem}}) \end{cases}$$

It may at first seem odd to equate (via gesture) the real non-partisan transcript to the "Democratic" transcript Colbert is arguing against. However, his point is that what Trump *believes* is a Democratic transcript is actually the real non-partisan transcript. Representing these two interpretations in the same region of the Interaction Space makes Colbert's contrast with a "Republican" transcript all the more effective. As he says "or Republican version of the transcript", Colbert moves his containment gesture in an arc across his body, presenting the "Republican" transcript into a region of the Interaction Space to his left. In doing so, he contrasts the imagined "Republican" transcript not only with the imagined "Democratic" transcript, but also with the real transcript metaphorically occupying the same space. This is represented in (156).

[11] This gesture is primarily a representational gesture, conveying meaning about the event of typing. As such, I do not consider it as enacting Interaction Space management. Despite this, the gesture still conveys discourse structural meaning through its location in a meaningful region of the Interaction Space.

(156) **Management actions by Colbert, Line 5-6, ex. 153**

$$IS1_2 = \begin{cases} \text{Participants: SC}_{\mathbf{S}}\text{, audience} \\ \text{Content : Transcript}_{\text{Real+Dem}} \\ \quad +\text{Transcript}_{\text{Rep}} \\ \text{Management :} \\ \quad \textsc{Present}(\text{CG6}\rightarrow\text{Transcript}_{\text{Rep}}) \\ \quad \textsc{Separate}(\text{CG5+CG6}\rightarrow\text{Transcript}_{\text{Real+Dem}}+\text{Transcript}_{\text{Rep}}) \end{cases}$$

This process of Combine and Separate actions is repeated as Colbert draws an analogy to claiming there to be "Democratic or Republican laws of physics" (line 7). By the end of this rebuttal, Colbert has two bounded regions of space, one metaphorically containing the true transcript, an imagined Democratic transcript, and an imagined Democratic physics, the other metaphorically containing the imagined Republican transcript and an imagined Republican physics. With this strategy, Colbert effectively contrasts "Democratic" and "Republican" while implying that a "Democratic" transcript or physics would be closer to reality than the "Republican" versions would be.

7.4.1.2 Pinching

The second example demonstrates the alternating use of containment and pinching gestures during specification. This is a particularly long clip with many interesting gesture sequences. For the sake of clarity, I focus only on gestures performed by Warren. Only Colbert's gestures at turn-transition points are annotated.

In this clip, Colbert is asking American politician Elizabeth Warren about her proposals for fixing the American healthcare system. He is especially concerned about Warren being honest as to how her plan would be funded. When she initially tries to say that "costs" will go up for wealthy Americans in order to pay for the plan (lines 1-2), Colbert interrupts to clarify that *costs* means *taxes* (line 3). After quickly confirming this, Warren continues, claiming that costs for "hard-working middle-class families" will go down (lines 4-5), performing a loose containment gesture in the central Interaction Space as she does so (WG1). As she begins her next sentence, Warren returns her left hand to rest and forms a pinching hand shape with her right hand. Colbert then interrupts again to clarify the definition of "cost", performing a representational tracing gesture as he asks whether the taxes of "hard-working middle-class families" will go up (CG1). Warren does not confirm this time, and instead begins a specification, acknowledging Colbert's viewpoint with "well", signalling disagreement with "but", and drawing attention to her forthcoming argument with HTT. As she uses this string of lexical discourse markers, Warren maintains the pinching hand shape and leans in sharply toward Colbert (WG2). Colbert than interrupts again, this time fully taking the turn to offer her his opinion on discussions concerning taxes and social welfare. When Warren finally regains the turn (line 20), She acknowledges and agrees with Colbert's main points, performing containment gestures as she does so. After this acknowledgement, Warren performs a pinching gesture as she begins to explain, again, how her plan will help,

not hurt, the budgets of middle- and lower-class people. She performs another containment gesture as she shifts from specifying *how* she knows this to specifying *what* she knows.

(157) Transcript 16: Elizabeth Warren & Colbert
[UID:6af48f4c-ecba-11e9-bb9b-089e01ba0335,1954]

```
1  EW  So here's how we're gonna do this, uh, costs are going to go
2      up for the wealthiest Americans, for big corporations-
3  SC  Taxes are what you mean by cost?
4  EW  Yeah.  *And hard-working middle-class families are going to
               *WG1- - - - - - - - - - - - - - - - - - - - - - - ->
5      see their costs go down.* *And-
       <- - - - - - - - - - - -* *WG2->
6  SC  *But will their taxes go up?*
       *CG1- - - - - - - - - - - - *
7  EW  Well- but here's the thing, [I spent]
       <- - - - - - - - - - - - - - - - - ->
8  SC                              *[No but here's the thing,] I've
                                   *CG2- - - - - - - - - - - - - ->
9      listened to these answers a few times before, and I- I just
       <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
10     want to, I just want to make a parallel suggestion for you
       <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
11     about how you might defend the taxes that* perhaps you're not
       <- - - - - - - - - - - - - - - - - - - - *
12     mentioning in your sentence - is that, isn't Medicare for All
13     *like public school?*
       *CG3- - - - - - - - *
14 EW *See, y'know, I actually-*
       *WG3- - - - - - - - - - -*
15 SC there might be taxes for it, but you certainly save a lot of
16     money on sending your kids to school, and do you want to live
17     in a world where kids aren't educated?  Do you want to live
18     in a world where your- your fellow citizens are *dying even
                                                       *CG4- - ->
19     i it costs a little bit of money?*
       <- - - - - - - - - - - - - - - - - *
20 EW [*(.)]  So, I accept your point and I believe in your point,
       *WG4- - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
21     healthcare is a basic human right,* *we fight for basic human
       <- - - - - - - - - - - - - - - - - -* *WG5- - - - - - - - - - >
```

```
22      rights and that's Medicare for All,* *everyone gets covered*
        <- - - - - - - - - - - - - - - - - - * *WG6- - - - - - - - -*
23      *(.)  But here's how I look at it, I've spent a big chunk of
        *WG7- - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
24      me career* studying why families go broke, *and a big reason
        <- - - - *                                  *WG8- - - - - ->
25      families go broke is healthcare*
        <- - - - - - - - - - - - - - - - *
WG1, WG4:  two hands, loose containment, central Interaction Space
WG2, WG7:  right hand, pinch into central Interaction Space
WG3:  right hand, index-finger point toward Colbert's gesture
WG5, WG8:  two hands, containment gesture, central Interaction Space
WG6:  both hands, open palms out, arch outward from center
CG1:  right hand, index-finger point traces upward curve, low-central
      Interaction Space
CG2:  right hand, pinch into central Interaction Space, beats
      throughout
CG3:  right hand, loose PUOH, elbow on table
CG4:  left hand, PUOH, arm extended leftward on desk
```

Figure 7.3: Gesture sequence corresponding to Transcript 16: Elizabeth Warren & Colbert, ex. 157

Warren performs four types of interactive gestures in this sequence – four containment gestures (WG1, WG4, WG5, WG8), two pinching gestures (WG2, WG7), one locating gesture (WG3), and one removal gesture (WG6). Each of these gestures performs a distinct action in relation to the Interaction Space and brings about a distinct Interaction Space state. Most relevant to the present discussion are her transitions between containment gestures and pinching gestures. Let's consider the first instance of this as she performs gestures WG1 & WG2. As Warren makes a general claim about costs going down for middle-class Americans, she performs a loose containment gesture (WG1). When she transitions from this general claim to trying to address Colbert's specific concern, raising taxes, she performs a pinching gesture (WG2), as if to pick out the most relevant information from her general claims. This is represented in (158).

(158) **Management actions by Warren, Lines 4-5, ex. 157**

$$IS1_1 = \begin{cases} \text{Participants: } \text{EW}_{\mathbf{S}}, \text{SC}_{\text{A}} \\ \text{Content : +Costs} \\ \quad \text{+Taxes} \\ \text{Management :} \\ \quad \textsc{Present}(\text{WG1}\rightarrow\text{Costs}) \\ \quad \textsc{Present}(\text{WG2}\rightarrow\text{Taxes}) \\ \quad \textsc{Combine}(\text{WG1+WG2}\rightarrow\text{Costs+Taxes}) \end{cases}$$

As in the previous example, Warren's two presentation gestures into the same region of the Interaction Space constitute a Combine action. However, because the presentation gestures are distinct in this case (containment vs. pinching), we do not get an equivalency between *costs* and *taxes*. Instead we understand *taxes* as metaphorically occupying a subregion of *costs*. This is a very effective management strategy as the pinching gesture, via its precision and the smallness of the implied metaphoric object, conveys that the taxes Colbert is concerned with is a relatively small part of the broader issue of cost that Warren is concerned with – there are other costs that need to be considered. Of course, Colbert refutes this, interrupting Warren and taking the turn before she can finish her specification. He achieves this by mirroring her verbally and gesturally, performing a pinching gesture and leaning in as he takes the floor with HTT (line 8, CG2).

Warren attempts to regain the turn shortly after Colbert's interruption (line 14). At this point, Colbert has just performed a PUOH gesture, introducing an analogy between Warren's healthcare proposal and the public school system. Colbert's Present move is represented in the first Interaction Space state in (159). With this analogy now in the space as a metaphoric object, Warren attempts to take the turn by referring to this analogy through a Refer action, pointing to Colbert's open hand. This is represented in the second Interaction Space state in (159).

(159) **Management actions by Colbert & Warren, Lines 12-14, ex. 157**

$$IS1_2 = \left\{ \begin{array}{l} \text{Participants: } SC_S, EW_A \\ \text{Content : +Public school} \\ \text{Management :} \\ \quad \textsc{Present}(CG3\rightarrow \\ \quad\quad \text{Public school}) \end{array} \right. \qquad IS1_3 = \left\{ \begin{array}{l} \text{Participants: } EW_{+S}, SC_A \\ \text{Content : Public school} \\ \text{Management :} \\ \quad \textsc{Engage}(WG3) \\ \quad \textsc{Refer}(WG3\rightarrow \\ \quad\quad \text{Public school}) \end{array} \right.$$

Warren seems to intend her gesture to achieve both a REFER action, specifying what topic she wishes to contribute to, and an ENGAGE action, allowing her to take the turn. However, Colbert has not performed a DISENGAGE action signalling a willingness to cede the floor. Because of this, Warren's relatively subtle ENGAGE action is insufficient to achieve a turn-transition. Instead, Colbert continues his turn (lines 15-19).

When Warren does eventually succeed in taking the floor (line 20), she performs another extended containment-pinch sequence. She uses a larger and looser containment gesture when acknowledging Colbert's contributions (lines 20-21, WG4), as if to physically gather the points he has presented. The physical affordances of this gesture put its status as a proper containment gesture in question – the orientation of her two hands could not actually support an object. Note that she performs this gesture as Colbert maintains a PUOH presentation gesture (CG4). I suggest that the unique physical affordances of WG4 indicate a kind of *collaborative* containment as Warren attempts to sufficiently acknowledge and distill Colbert's contributions, which he is still physically presenting, before moving on. Warren then tenses her hands and moves them closer together (WG5), presenting her proposal, Medicare for All (M4A), as an example of the fight for human rights Colbert has been suggesting. This is represented in the first Interaction Space state in (160).

Warren then rotates her palms outward, and moves them in an arching sweep away from the speaker-hearer line, as if to clear a vertical surface (WG6). I suggest that this gesture performing both a semantic and discourse management function. Semantically, the sweeping motion co-expresses the exhaustivity expressed in "everyone" (line 22).[12] The gesture also serves to remove metaphoric objects from the Interaction Space, as discussed extensively in Chapter 5 on *anyway*. This is represented in the second Interaction Space state in (160).

[12]This interpretation is supported by Kendon's analysis of sweeping gestures as, in a sense, negating all possible alternatives (Kendon 2004, ch. 13; see also Harrison 2010).

(160) **Management actions by Warren, Lines 20-22, ex. 157**

$$IS1_4 = \begin{cases} \text{Participants: } EW_S, SC_A \\ \text{Content : +Human rights} \\ \quad \text{+M4A} \\ \text{Management :} \\ \quad \textsc{Present}(\text{WG4}\rightarrow \\ \quad\quad \text{Human rights}) \\ \quad \textsc{Present}(\text{WG5}\rightarrow\text{M4A}) \\ \quad \textsc{Combine}(\text{WG4+WG5}\rightarrow \\ \quad\quad \text{Human rights+M4A}) \end{cases} \qquad IS1_5 = \begin{cases} \text{Participants: } EW_S, SC_A \\ \text{Content :} \\ \quad \sout{\text{Human rights+M4A}} \\ \text{Management :} \\ \quad \textsc{Remove}(\text{WG6}\rightarrow \\ \quad\quad \text{Human rights+M4A}) \end{cases}$$

Having cleared the Interaction Space, Warren can reintroduce the topic she was pursuing before Colbert's derailment – explaining that the primary cost faced by middle-class and working-class Americans is health care, rather than taxes. To signal this return, she performs another pinching gesture, this time aligned with "here's how I look at it" (line 23). Finally, she transitions from this pinching gesture to another containment gesture as she begins a new line of inquiry related to the costs families incur for healthcare services.

This discourse excerpt has demonstrated the complex ways in which a gesturer can exploit different presentation gestures to profile different aspects of a specification move. In all cases, the pinching gesture is used to draw attention to a particular contested point to which information is being added. Containment gesture, on the other hand, situate an entire topic or set of topics that can be further explored and managed.

7.4.1.3 Palm-up Open-hand

The final clip in this section features a PUOH-containment sequence performed by American attorney and politician Kamala Harris. In this clip, Colbert and Harris are discussing the impeachment trials for then-president Donald Trump. Colbert has asked Harris who she thinks should be called to the witness stand, given her experience as a prosecutor. She answers that she thinks all of the men closest to the president should be called to the witness stand (lines 1-2). Her and Colbert than collaboratively form a list of these men, with their turns overlapping. When Harris says "Pence" (then-vice president Mike Pence), Colbert stops to express shock ("ooh", lines 4 and 6). Harris, after reinforcing her suggestion (lines 5 & 7), dismisses any reason for deeming this unacceptable, saying "enough of that" (line 7). At this dismissal, Harris performs a gesture (HG1) typically associated with negation, shaking her down-turned open hand as if to clear a flat surface (e.g. Bressem & Müller 2014, 2017; Gawne 2021; Harrison 2010; Kendon 2004). She then goes on to explain why she thinks such an action is appropriate, using HTT to mark the initial specification (line 8), and performing a PUOH gesture into the central Interaction Space as she does so (HG2). She then uses a series of containment gestures, spatially separating what needs to happen (HG3 & HG5) from what the listed men are trying to do (HG4). At the end of this sequence, as

Harris is summarizing her point (line 11-12), Colbert points toward her containment gesture (CG1) without saying anything contentful. He then moves to end the interview (lines 13-14).

(161) Transcript 17: Kamala Harris
[UID:97a9a7ce-0cfa-11ea-8374-089e01ba0770,2909]

```
1   KH   Going forward, again, all the president's men
2        bring 'em all [forward.  Pompeo, ah- ah-] Pence
3   SC                 [Pompeo, Bolton]
4        Ooh
5   KH   Yes
6   SC   Ooh
7   KH   Yes, yeah, yeah, yes.  *Enough of that.  Y'know,* *I mean,
                                *HG1- - - - - - - - - - -* *HG2- ->
8        because here's the thing*, I mean, *you can't hide behind*
         <- - - - - - - - - - - -*         *HG3- - - - - - - - - *
9        *y'know the president on this, but you- you can't prance around
         *HG4- - - - - - - - - - - - - - - - - - - - - - - - - - - - - >
10       talking about being vice president with all of the benefits that
         <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - ->
11       come with that* *and not take the responsibility that comes with
         <- - - - - - -* *HG5- - - - - - - - - - - - - - - - - - - - ->
12       that position*
         <- - - - - -*
13 SC    *(.)  Ah- I-* well Senator, thank you so much for being here,
         *CG1- - - - *
14       lovely to see you again
HG1:  right hand, open palm down, small lateral shakes, central
      Interaction Space
HG2:  right hand, open palm up, central Interaction Space
HG3, HG5:  two hands, loose containment, movement into central
      Interaction Space then down and right
HG4:  two hands, open palm up, repeated outward movement, peripheral
      Interaction Space
CG1:  right hand, index finger point toward Harris' gesture
```

Figure 7.4: Gesture sequence corresponding to Transcript 17: Kamala Harris, ex. 161

The first gesture depicted for Harris in this excerpt in which she shakes a down-turned open hand back and forth as if to clear crumbs from a surface,[13] performs two functions: it (i) co-expresses the negative assessment in "enough of that" and the accompanying head shake she performs, and (ii) clears the Interaction Space for her to then present her arguments. The subsequent PUOH gesture introduces the set of general arguments she is about to make, beginning a new line of inquiry to address the established question of why it is appropriate to call these men as witnesses. Harris then introduces her main argument into the Interaction Space with a containment gesture (HG3), sweeping both hands down and rightward to the desk's edge. She contrasts this argument, that these men need to take responsibility for their actions, with what they are trying to do by shifting her body and gestures away from the center of the Interaction Space (HG4). She repeats cyclic upward PUOH gestures in the right periphery of the Interaction Space as she describes the men's actions (lines 9-11) – "prance around", "talking about being vice president", "all the benefits that come with that".[14] Harris then reinforces this contrast by returning to the central Interaction Space and repeating a containment gesture as she summarizes her point (lines 11-12, HG5). This sequence of management actions is represented in (162).

[13]See Bressem & Müller (2014, 2017) and Teßendorf (2014) for discussion of so-called "brushing aside" gestures.

[14]See Müller (2004) and Ladewig (2011, 2014a) for discussion of cyclic PUOH gestures.

(162) **Management actions by Harris, Line 7-12, ex. 161**

$$IS1_1 = \begin{cases} \text{Participants: } \text{KH}_{\mathbf{S}}, \text{SC}_{\text{A}} \\ \text{Content : +Taking responsibility} \\ \quad \text{+Acting with impunity} \\ \text{Management :} \\ \quad \textsc{Present}(\text{HG3}\rightarrow\text{Taking responsibility}) \\ \quad \textsc{Present}(\text{HG4}\rightarrow\text{Acting with impunity}) \\ \quad \textsc{Separate}(\text{HG3+HG4}\rightarrow\text{Taking responsibility+Acting with impunity}) \\ \quad \textsc{Separate}(\text{HG4+HG5}\rightarrow\text{Acting with impunity+Taking responsibility}) \end{cases}$$

The PRESENT action performed by HG5 can be further divided into a series of PRESENT and COMBINE actions as she repeats the PUOH gesture in the same space. The SEPARATION actions she performs in this sequence are especially effective in their relative positioning in the Interaction Space. Her argument is placed centrally on the desk between her and Colbert, and it is this argument that is most important to achieving the discourse goal. She places what amounts to their argument, that they should be able to act with impunity, on the periphery of the Interaction Space, signaling that their viewpoint is should not be considered legitimate when addressing discourse goals. Colbert then, in a sense, confirms his agreement by LOCATING Harris' argument, pointing toward her final containment gesture (CG1). Pointing gestures such as this have been argued to confirm uptake in common ground management (Holler 2009), which seems to be the role of Colbert's gesture here.

This sequence has shown how presentation gestures can not only introduce particular specifications to the discourse, but can also mark those specifications as being relevant or not to achieving discourse goals by locating them in the center or peripheral Interaction Space.

7.4.2 Locating

Locating gestures are deictic gestures that refer to a discourse topic. This can be done *metonymically* by pointing to an object related to the topic, or *metaphorically* by pointing to a dedicated region of the Interaction Space.[15] We have already seen two examples of locating gestures in this chapter in which one interlocutor points toward the presentation gesture of another interlocutor to either indicate agreement or their intent to add to the discourse topic presented. In this section I discuss one additional example in which Colbert transitions from a presentation gesture to locating gestures within a single turn.

In (163), Colbert is talking to the show's bandleader, Jon Batiste, about his family's vacation home in South Carolina. As he introduces the vacation home, Colbert performs a two handed containment gesture into the central Interaction Space (lines 1-3, CG1). He then drops his left hand to a resting position on the desk, while holding his right hand out to specify a particular feature of the vacation home, a fig tree (lines 3-4, CG2). He then goes

[15]See McNeill et al. (1993) for early seminal work on 'abstract deixis' and the use of deictic gestures in topic organization.

on to describe a salient feature of the fig tree, noting that the figs all ripen at once, marking this specification as particular important with HTT (lines 3-5, CG2). As he specifies the particular time of year when this happens (line 5-6), he brings his right hand down to the desk, with his index finger metaphorically pointing to the "second week of July" (CG3).

(163) TRANSCRIPT 18: STEPHEN COLBERT
[UID:d6941018-a793-11e9-902d-089e01ba0335,1273]

```
1   SC   Um, but, uh, so we never debate on where to go when we have
2        time off, we go see our families in South Carolina.  And on-
3        *on the land down there, we got a little house down there*
         *CG1- - - - - - - - - - - - - - - - - - - - - - - - - - - -*
4        *and we have a fig tree, and here's the thing is that that
         *CG2- - - - - - - - - - - - - - - - - - - - - - - - - - - ->
5        fig tree really gives up its fruit.  The figs are really
         <- - - - - - - - - - - - - - - - - - - - - - - - - - - - - >
6        ripe* *the second week of July*
         <- -* *CG3 - - - - - - - - - *
CG1:  both hands, loose containment held into central Interaction
      Space
CG2:  right hand, loose palm down, held in central Interaction Space
CG3:  right hand, index finger point, moved down to touch central desk
```

Figure 7.5: Gesture sequence corresponding to TRANSCRIPT 18: STEPHEN COLBERT, ex. 163

As Colbert adds increasingly specific information, he deictically refers to increasingly specific regions of space. His initial loose containment gesture (CG1) introduces the topic of his vacation home in its entirety into the Interaction Space. Because we can easily accommodate the presence of a fig tree on the property, a locating gesture to a particular region of this space suffices when introducing the referent. Because we know that the existence of a fig tree entails ripe figs at a particular time of year, another locating gesture, toward an even more particular region of the 'vacation home' space is sufficient for introducing this

information.[16] This sequence is represented in (164). In this representation, the embedded parentheses indicate subsections of a meaningful region of the Interaction Space. For example, "REFER(CG2→Vacation home(fig tree))" represents a locating gesture that indicates a subregion of the introduced "vacation home" topic space, identified as the "fig tree".

(164) **Management actions by Colbert, Lines 2-6, ex. 163**

$$IS1_1 = \begin{cases} \text{Participants: } \text{SC}_{\mathbf{S}}, \text{JB}_{\text{A}} \\ \text{Content : +Vacation home} \\ \text{Management :} \\ \quad \textsc{Present}(\text{CG1}\rightarrow\text{Vacation home}) \\ \quad \textsc{Refer}(\text{CG2}\rightarrow\text{Vacation home(Fig tree)}) \\ \quad \textsc{Refer}(\text{CG3}\rightarrow\text{Vacation home(Fig tree(Fig harvest))}) \end{cases}$$

This representation of the Interaction Space highlights the functionality of Colbert's gestures progressing as they do. In order for either locating gesture (CG2, CG3) to effectively manage the space, it is necessary for it to follow a more general presentation gesture, and for it to indicate a space that overlaps with the more general topic space. Imagine if instead Colbert had moved his right hand further to the right, outside of the space indicated in CG1, as he described the fig tree. It would be very confusing. The REFER action would in a sense *separate* the fig tree from the vacation home space, which doesn't make much sense at all.

This sequence demonstrates the ways in which speaker's can use locating gestures to refer to inferable aspects of an introduced topic. Colbert successfully assigns a region of the Interaction Space as representing his vacation home. In doing so, he can then use locating gestures to elaborate on particular features of the vacation home. The position of each gesture, and the order that they occur in, is necessary for interpretation.

7.4.3 Addressing

In this section I discuss two instances of addressing gestures aligned with HTT. Though addressing gestures were not explicitly predicted, they are not unexpected. Prototypical uses of addressing gestures function to either refer explicitly to the addressee and their contributions or invite the addressee to take some action (Bavelas et al. 1992; Bavelas 1994). In these uses, the gesturer is drawing joint attention to the addressee – they either have already done something that should be acknowledged or are being asked to do something now. With HTT, addressing gestures seem to function differently, instead of drawing attention to the addressee, they seem to request the attention of the addressee.

[16] This final deictic gesture (CG3) may also convey representational information. The downward trajectory of the point could depict ripe fruit dropping, as evoked by Colbert's preceding description of the fig tree when it "gives up its fruit".

7.4.3.1 Role reinforcement

The first example consists of an *addressing-presenting* sequence in which Colbert addresses both the audience and his band leader, Jon Batiste, as he repeats HTT He then transitions into a pinching presentation gesture to specify the exciting information. In this clip, which occurs just after a commercial break, Colbert is talking about what is happening on the show that night. He mentions one of his guests, American politician Nancy Pelosi (lines 1-2), and then briefly pauses, looking down at the script in his hands (CG1). Colbert then continues, specifying why it is particularly exciting to have her on the show that evening, and prefacing the specification with HTT (lines 2-6). During the first HTT, Colbert lifts the script and points it toward the audience, sweeping from left to right, ensuring the entire audience is addressed. During the second HTT, he sweeps the script back rightward, addressing his band leader Jon Batiste while performing a series of small vertical beats. As he begins to explain the circumstances of Pelosi's visit (lines 3-4), Colbert places the script back down on the desk and performs a pinching gesture into the central Interaction Space between him and Batiste, leaning his head and body in as he does so.

(165) Transcript 19: Stephen Colbert
[UID:6a02fce6-159f-11ea-8a36-089e01ba0335,1037]

```
1  SC   You know who I got coming up in just a minute here?  Nancy
2       Pelosi is gonna be right over here *(.)* *Here's the thing,
                                          *CG1* *CG2- - - - - - ->
3       here's the thing* *this is what we call in the bizz an
        <- - - - - - - -* *CG3- - - - - - - - - - - - - - - - ->
4       exclusive.  Today was the vote to formalize the impeachment,*
        <- - - - - - - - - - - - - - - - - - - - - - - - - - - - -*
5       open up the hearings.  She's been nowhere, she's done
6       nothing.  This is it.  Straight from there, right to the seat
CG1:  both hands hold script on desk
CG2:  right hand, lifts script and points in sweeping motion to
      Batiste, then the audience then back to Batiste
CG3:  right hand, precision grip, held in central interaction space,
      beats performed throughout
```

Figure 7.6: Gesture sequence corresponding to TRANSCRIPT 19: STEPHEN COLBERT, ex. 165

As stated, there are two typical uses of addressing gestures – citing an addressee's contribution, and requesting a contribution (e.g. Bavelas et al. 1992; Bavelas 1994). When used to cite an addressee's contribution, the gesture performs a REFER action, referring to a topic metonymically by referring deictically to it's source. When used to request a contribution, the addressing gesture still performs a REFER action but composes with a DISENGAGE action to perform a composite REQUEST action. This composite action requests a subsequent Interaction Space state in which the addressee engages as speaker.

I argue that in specification contexts, addressing gestures perform a REFER action, as usual, by deictically referring to the addressee. However, in specification contexts, the addressing gesture composes with an ENGAGE action instead of a DISENGAGE, this results in a reinforcement of interlocutor roles, rather than a switch in interlocutor roles. This process is represented in (166). The ENGAGE action is performed by Colbert's reorientation toward Batiste and acts upon his role in the discourse. In turn-transitions, the gesturer's role at the time of performance is that of addressee, resulting in the role as speaker being established. In this case, Colbert's role at the time of performance is already that of speaker, and so his role as speaker is reinforced rather than established. The ENGAGE and REFER compose to request that the other interlocutor cede the floor by performing a DISENGAGE action. In this context, the action would act upon JB whose role is already that of addressee, resulting, again, in a reinforcement rather than establishment of roles.

(166) **Management actions by Colbert, Lines 2-3, ex. 165**

$$IS2_1 = \begin{cases} \text{Participants: } \text{SC}_{+\text{S}}, \text{JB}_{+\text{A}} \\ \text{Content : Pelosi} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{CP2}{\rightarrow}\text{SC}_{\text{S}}) \\ \quad \textsc{Refer}(\text{CG2}{\rightarrow}\text{JB}_{\text{A}}) \\ \quad \textsc{Request}(\text{CP2+CG2:IS1}_2) \end{cases} \qquad IS2_2 = \begin{cases} \text{Participants: } \text{SC}_{\text{S}}, \text{JB}_{+\text{A}} \\ \text{Content : Pelosi} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{BG}{\rightarrow}\text{JB}_{\text{A}}) \end{cases}$$

With HTT marking an utterance as particularly important, and the addressing gesture requesting a reinforcement of interlocutor roles, Colbert then moves on to the specification itself. Colbert's posture shift (leaning further into the Interaction Space, CP3) performs

another reinforcing ENGAGE action. The pinching gesture introduces the topic of the interview being an "exclusive" via a PRESENT action, as we have seen many times before. This is represented in (167).

(167) **Management actions by Colbert, Lines 3-4, ex. 165**

$$IS2_3 = \begin{cases} \text{Participants: } SC_{+S}, JB_A \\ \text{Content : Pelosi} \\ \quad +\text{Exclusive} \\ \text{Management :} \\ \quad \text{ENGAGE}(CP3 \rightarrow SC_S) \\ \quad \text{PRESENT}(CG3 \rightarrow \text{Exclusive}) \end{cases}$$

This example has demonstrated how addressing gestures request attention from an addressee by reinforcing participant roles. Though posture shifts were not the focus of previous examples in this chapter, it is likely the case that role reinforcement occurs elsewhere. This is particularly apparent in the recurrent association between pinching gestures and leaning forward into the Interaction Space.

7.4.3.2 Interaction Space negotiation

The second addressing sequence I discuss demonstrates the use of addressing gestures during emotionally charged interactions. This discourse excerpt was previously presented as an example of the circumstance relation during interviews (ex. 141).

In this clip, Indian-American actor Mindy Kahling and Colbert are in the process of recovering from a derailment after an automated phone call from a reward program interrupted the interview. Kahling is annoyed by the interruption, and expresses confusion as to why someone as rich as Colbert would have a rewards program (lines 3-5). She points to herself as she reiterates this, saying she doesn't have a rewards program (lines 4-5, KG1). Kahling then disengages with the primary Interaction Space to make an appeal to the audience. Colbert takes this disengagement as an opportunity to try to explain himself and apologize, reaching toward Kahling and leaning far over the desk as he does so (line 6, CG1). Kahling rejects this attempt, saying no, and mirroring his HTT marker, signalling that the audience should be listening to her, not him. Once Colbert gives up trying to explain himself, Kahling continues her appeal to the audience, specifying the circumstances under which to interpret the situation – Colbert has the number one show on television, and so can assumed to be rich, and so has no reason to have a rewards program (lines 8-9). This appeal is meant to, in a way, legitimize her outrage – her interview was interrupted by something that shouldn't exist in the first place.

(168) TRANSCRIPT 20: STEPHEN COLBERT & MINDY KAHLING
[UID:fc6004ce-bb38-11e9-afb5-089e01ba0770,2163]

```
1  SC   This is how you get rich, you go for the rewards.  Exactly.
2       I also steal rolls from au bon pain.
3  MK   I don't even- y'know, I don't even know.  I have so many more
```

```
4        questions about your financial situation.  *I don't use a
                                                    *KG1- - - - ->
5        rewards program.* *He has-
         <- - - - - - - * *KG2- ->
6   SC   *Here's the thing.  [Hold on.  Here's the thing.  hold on]
         *CG1- - - - - - - - - - - - - - - - - - - - - - - - - - -≫
7   MK                    [No.  He has-* *no, no.* *Here's the thing*]
         <- - - - - - - - - - - - - -* *KG3- -* *KG4- - - - - - -*
8        *he has the- he has the number one show on t.v.  and he's
         *KG5- - - - - - - - - - - - - - - - - - - - - - - - - - ->
9        like* *"ah, I gotta make sure* that if I travel, that me and
         <- -* *KG6- - - - - - - - - -*
10       my wife and my three children are being hosted in an
11       economical way"
KG1:  left hand, index finger point toward self
KG2:  right hand, index finger up, held toward audience
KG3:  right hand, index finger up, held toward audience, moved in
     lateral arch back and forth
KG4:  right hand, index finger point toward Colbert, small beats
KG5:  two hands, index fingers up, held toward audience
KG6:  two hand pantomime of using smart phone
CG1:  two hand far reach across desk toward Kahling
```

Figure 7.7: Gesture sequence corresponding to Transcript 20: Stephen Colbert & Mindy Kahling, ex. 168

Both Colbert and Kahling perform addressing gestures aligned with HTT during this sequence. However, they perform different functions determined by (i) the participant roles

they fill at the time of performance, and (ii) the broader social context. First, Colbert performs his addressing gesture, reaching dramatically over the desk (CG1), while he does not have the floor. In fact, his gesture and the accompanying HTT utterance interrupt Kahling mid-sentence. This sequence is represented in (169). This representation is typical of a contested turn-transition. Kahling is filling the role of speaker at the time of performance (MK_S).[17] Colbert's ENGAGE action acts upon his role as addressee (ENGAGE(CP2→SC_A)) and switches his role upon performance to speaker (SC_{+S}). This ENGAGE actions composes with his addressing gesture to achieve a REQUEST action, requesting that Kahling bring about $IS1_2$ in which she disengages from her role as speaker and cedes her turn to Colbert. At this point, Kahling can either enact this request and allow Colbert to explain himself, or deny the request and continue her turn. She opts for the second.

(169) **Management actions by Colbert, Line 6, ex. 168**

$$IS1_1 = \begin{cases} \text{Participants: } SC_{+S}, MK_S \\ \text{Content : Rewards program} \\ \quad \text{Interruption} \\ \text{Management :} \\ \quad \textsc{Engage}(CP2 \rightarrow SC_A) \\ \quad \textsc{Refer}(CG1 \rightarrow MK_S) \\ \quad \textsc{Request}(CP2{+}CG1{:}IS1_2) \end{cases} \qquad IS1_2 = \begin{cases} \text{Participants: } SC_S, MK_{\text{-}S} \\ \text{Content : Rewards program} \\ \quad \text{Interruption} \\ \text{Management :} \\ \quad \textsc{Disengage}(KG \rightarrow MK_S) \end{cases}$$

In response to Colbert's request for a turn-transition, Kahling directly denies the request, repeating "no" three times (line 7). This direct denial is co-expressed in the verbal mode as she shakes her head and moves her upheld finger in large lateral sweeps (KG3). After this direct denial, she repeats Colbert's use of HTT to reinforce her role as speaker and as the contributor of particularly important information. This is the point at which she performs her addressing gesture, maintaining her index-finger point and moving it downward to point to Colbert, moving her hand in small vertical beats as she does so (KG4). However, as she performs this addressing gesture, she remains oriented toward the audience, engaged in this secondary Interaction Space and, importantly, *disengaged* from the primary space between her and Colbert. This means that her addressing gesture performs a REFER action that cannot compose with an ENGAGE action to reinforce her role as speaker, which seems to be her intention. This is represented in (170). Note that the Interaction Space in which the management is enacted is necessarily the Interaction Space she is engaged in at the time of the gesture's performance (IS2).

[17] Kahling is actually engaged in the secondary Interaction Space between her and the audience at this point. However, her role as speaker in that space is known by Colbert, and interruption of that role still violates the *one speaker at a time* principle (Sacks et al. 1978), and thus still constitutes an overlapping turn.

(170) **Management actions by Kahling, Line 7, ex. 168**

$$IS2_1 = \begin{cases} \text{Participants: } \text{MK}_\text{S}, \text{Audience}_\text{A} \\ \text{Content : Rewards program} \\ \quad \text{Interruption} \\ \text{Management :} \\ \quad \textsc{Refer}(\text{KG4}\rightarrow\text{SC}_\text{A}) \end{cases}$$

This form of role reinforcement is, in a way, passive. The addressing gesture acknowledges Colbert's attempt at turn-transition, while Kahling's refusal to engage in the relevant Interaction Space is what gesturally expresses her denial of Colbert's request. I consider this an interesting demonstration of the *optionality of expression* principle of Multimodal Discourse Management which states that "the use of each expressive strategy is always optional" (Chapter 1). Kahling could reorient and perform an ENGAGE action in the Interaction Space between her and Colbert to reinforce her role as speaker. However, she has already used very clear verbal strategies to do this (saying "no" and repeating HTT). It is thus not necessary to also employ a gestural strategy of reinforcement.

After this interactive gesture, Kahling then performs a series of representational gestures that contribute information about the circumstance against which the audience should judge Colbert's participation in the rewards program (KG5 & KG6). I argue that this discourse structural context highlights the efficacy of Kahling's choice in expressive strategies described above – her primary discourse goal is to convince the audience of the ridiculousness of the interruption, and it is thus in her interest to remain engaged in the Interaction Space between her and the audience rather than spend unnecessary time attending to Colbert.

The gestural context described above can also help us to better understand the form that Colbert's addressing gesture takes. This form is particularly extreme, as he leans as far as he can over the desk and reaches into Kahling's personal space. In the gestural context, the extremity of form seems almost necessary in his attempt to perform an ENGAGE action. Kahling is not maintaining the Interaction Space between her and Colbert at the time, and thus Colbert is not even an active participant in the discourse. To achieve his intent, he thus must not only engage in the Interaction Space, he also must make Kahling re-engage in the Interaction Space. This requires him to in some way insert himself in an interaction he is not actually a participant in.

7.4.4 Stopping

Stopping gestures are manual gestures that request some action to be stopped. This is associated with a DISENGAGE action upon a particular discourse topic. I identify two stopping gesture variants in this data set: *blocking* and *pausing*. *Blocking* gestures, performed with open palms facing outward, enact a boundary between the speaker and an unwanted metaphoric object. *Pausing* gestures, performed with the index-finger pointed upward, seem to indicate the numeral one, as if to ask "wait one second". These two variants are depicted in Figure 7.8. The transcripts corresponding to these examples are given in (171).

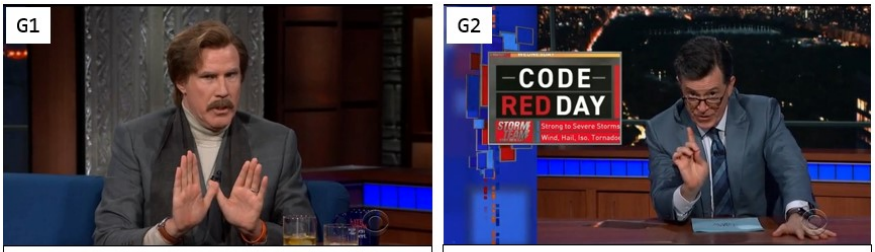


Figure 7.8: Types of stopping gestures as observed with HTT

These stopping gestures counter our predictions for gestural expressions of specification. In specification moves, we expect consistent engagement with the Interaction Space and topics therein as speakers actively contribute to discourse goals. Stopping gestures are associated with the exact opposite: a *refusal* to contribute to current discourse goals. I argue that the occurrence of *stopping* gestures with specification moves profile aspects of the discourse structure other than the specification itself. Often, this is supported by the co-occurrence of additional lexical discourse markers, as in (171). Again, these two transcripts correspond to the screenshots in Figure 7.8 above demonstrating stopping gesture variants.

(171) a. Transcript 21: Will Ferrell
[UID:d129e820-ba6f-11e9-a245-089e01ba0335,2886]

```
1  SC   Did you ever and would you ever sleep with a subject to
2       get an interview?
3  WF   Well, (.)  *here's the thing*, it is completely unethical
                    *G1- - - - - - - *
4       it is abuse of power.  But yes, I did it, about f-, about
5       fifteen times.
```

b. Transcript 22: Stephen Colbert
[UID:afbec190-9d5c-11e9-9e5f-089e01ba0770,1190]

```
1  SC   There's controversy involving WICS, Springfield Illinois'
2       news leader.  They were told by their corporate owners
3       that anytime there was any bad weather, everything from a
4       tornado to a light drizzle, they had to announce it to
5       their viewers with the graphic "Code Red".  The theory
6       behind the "Code Red" is that it's exciting, it's grabbing
7       and more people would keep watching if you scare them until
8       their pants are at code brown.  *Here's the thing though*,
                                        *G2- - - - - - - - - - -*
9       it's not all that informative.  So viewers complained to
10      the station with thousands of comments on social media,
```

```
11      letters to the editor, frequent calls to local talk-radio
11      shows.
G1:  two hand blocking gesture toward Colbert/camera
G2:  right hand pausing gesture toward audience
```

In both of these excerpts, there is a another lexical discourse marker adjacent to HTT that serves a discourse function other than marking specification or emphasis. As previously described, the "well" in (171a) indicates an acknowledgement of contrasting viewpoints (e.g. Aijmer & Simon-Vandenbergen 2003; Le Lan 2007). In this particular case, the American actor Will Ferrell (playing as his character Ron Burgundy) is acknowledging that people are going to judge the answer he is about to give. The blocking gesture Ferrell performs acts upon these perceived judgements rather than the qualifying specification itself. Similarly, the *pausing* gesture in (171b) may be profiling the oppositional properties expressed by the adjacent "though", rather than the specification. In this case, Colbert is rejecting the legitimacy of the reasons for using the phrase "Code Red" and then elaborating on the fallout of misusing the phrase.

Table 7.9 gives the distribution of stopping gesture variants by discourse structural context.[18] *Pausing* gestures have a noticeably more restricted distribution, and occur only with the most 'typical' flavors of specification – those which contribute directly to discourse goals. *Blocking* gestures, on the other hand, are more evenly distributed across functional contexts. Given these findings, I hold that both *blocking* and *pausing* gestures mark disengagement from a discourse topic, as previously suggested. However, I argue that the manner of t with the topic is different in two important ways.

First, I argue that blocking gestures contribute to a REQUEST for topic *removal*, whereas pausing gestures do not. This is supported by the accompanying postural shifts we see for each variant. Blocking gestures are often accompanied by some kind of non-manual DISENGAGE action, such as looking or leaning away (e.g. Figures 7.8 & 7.9). As discussed in Chapter 4, the non-manual and manual disengagement compose to REQUEST the removal of a discourse topic. Pausing gestures, on the other hand, are frequently accompanied by postural shifts that *increase* engagement with the Interaction Space, such as that in Figure 7.8. Without an accompanying DISENGAGE action, a REQUEST action is not composed.

Second, the topic-disengagement in pausing gestures is *temporary*. In *background, circumstance, contradiction* and *elaboration* use contexts, the ongoing discourse is, in a sense, *paused* while the specification is made, as if to say "we can't move on without this point being made". The topic-disengagement, or *requested* topic-disengagement, with blocking gestures is *permanent*. In topic-shift contexts, this is related to the dismissal of a topic (see Chapter 5). In the remaining contexts, this related to the dismissal of an erroneous assessment (as in the *circumstance* relation), a false statement (as in the *contradiction* relation), or a judgement (as in the *qualification* relation). This leaves only the use of a blocking gesture with an elaboration relation exceptional.

[18]Note that one *pausing* gesture occurred in a repeated HTT and thus is not counted in this table.

	Blocking	Pausing	**Total**
Background	-	1/14 (7%)	**1/14** (7%)
Circumstance	3/16 (19%)	3/16 (19%)	**6/16** (38%)
Contradiction	1/10 (10%)	2/10 (20%)	**3/10** (30%)
Elaboration	1/16 (6%)	6/16 (38%)	**7/16** (44%)
Qualification	1/3 (33%)	-	**1/3** (33%)
Return Shift	3/30 (10%)	-	**3/30** (10%)
Full Shift	2/5 (40%)	-	**2/5** (40%)

Table 7.9: Stopping gestures accompanying HTT by functional context

In the remainder of this section, I discuss two clips that demonstrate the use of stopping gestures during a specification move. Through these two close analyses I show that stopping gestures can provide complimentary information about the discourse structural and social context of the specification that is not expressed directly in the verbal mode.

7.4.4.1 Blocking a question

The following example demonstrates the use of stopping gestures as a kind of hedge when answering difficult or unwanted questions. In this clip, Colbert asks American comedian and political podcaster Jon Lovett whether or not we can trust political polls. At the time, polls consistently showed Donald Trump losing the 2020 election to several potential candidates. However, trust in polls had plummeted after the 2016 election in which Trump won the presidency despite nearly every poll predicting otherwise. Lovett is very hesitant to answer, looking away from Colbert and performing a blocking gesture before he even begins speaking. After some hesitation, he says "yes", moving his hands outward laterally (line 4, LG2). He then quickly contradicts himself, saying "but don't" and pushing his hands forward toward

the audience (line 4, LG3).[19] Colbert then jumps in to also interact with the audience. A series of overlapping turns occur as Lovett tries to regain the turn and clarify himself. First, he uses a *pausing* gesture aligned with "wait, that was confusing" (lines 6 & 8, LG4). Then he performs an exaggerated pinching gesture, leaning far forward toward the audience as he requests a chance to explain himself (line 10, LG5). He fully recovers his turn as he says HTT and repeats the blocking gesture he performed prior to the derailment (line 11, LG6). As he begins to actually explain the circumstances against which to interpret his initial answer, he turns his head and gaze back to Colbert and performs a containment gesture into the central Interaction Space (lines 12, LG8). In the end, Lovett seems to indicate that Colbert is asking the wrong question given that polls don't determine elections, voting determines elections (lines 17-18)

(172) Transcript 23: Jon Lovett
[UID:ea5070e6-ef15-11e9-8194-089e01ba0335,3541]

```
1   SC   As- as somebody who's again who's- who's been there
2        intimately and seen the internal polling and stuff on
3        campaigns, *can we trust polls* *at all?
                    *CG1- - - - - - - -* *CG2- - >
4   JL   Ah- *(sigh)* *yes.* *But [don't]*
             *LG1- -* *LG2-* *LG3- - - - *
5   SC                              *[You] heard him*
                                    *CG3- - - - - -*
6   JL   *Wait
         *LG4- >
7   SC   *Five people
         *CG4- - - - ≫
8   JL   that was confusing*
         <- - - - - - - - - -*
9   SC   [Five people can beat Trump]
10  JL   *[I'll explain it.*]
         *LG5- - - - - - - *
11       *Here's the thing* uh *I get why we're all looking at
         *LG6- - - - - - -*    *LG7- - - - - - - - - - - - - - >
12       polls.* *Our brains are profoundly broken*, uh we are- we
         <- - -* *LG8- - - - - - - - - - - - - - - *
13       are drowning in political coverage, we want this nightmare to
14       end, and we want someone to give us certainty.  Just tell me,
15       please tell me that it is going to be over soon.  And so
```

[19] Lovett's third blocking gestures (LG3) also conveys representational information, as he is literally asking Colbert and the audience to not trust political polls.

```
16      ya click up- ya click up your little polls and the say it's
17      gonna be over soon.  But (.)  the only way it's gonna be over
18      is if we do our part to make it over
LG1, LG6:  two hand blocking gesture toward audience
LG2, LG7:  two hand blocking gesture toward audience, arcing sweep
     outward from center
LG3:  two hand blocking gesture toward audience, arcing sweep outward
     from center with push forward
LG4:  both hands, index finger up, palm out, held toward Colbert then
     audience
LG5:  right hand, pinch, moved out and down
LG8:  two hand containment gesture held toward Colbert
CG1:  right hand, palm up open hand, held toward Lovett
CG2:  right hand, fist held down on desk toward Lovett, palm side
CG3:  right hand, fist rotated to palm up orientation
CG4:  right hand, palm out open hand, fingers spread representing "5"
```

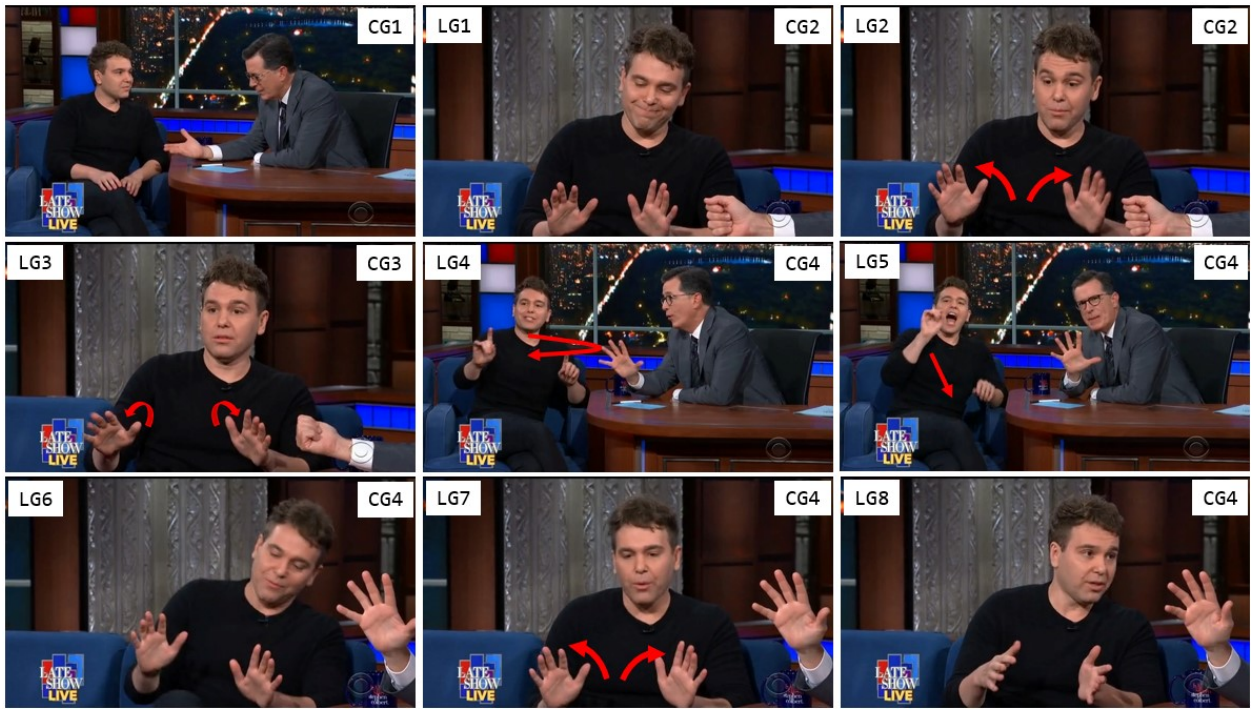


Figure 7.9: Gesture sequence corresponding to TRANSCRIPT 23: JON LOVETT, ex. 172

Colbert begins this sequence with a typical presentation gesture, leaning toward Lovett to ask the question and offering the topic as a metaphoric object on his open hand (CG1).

He then cedes his turn, signalled through the completion of the verbal question and the retraction of his presentation gesture to a rest position (CG2). This sequence is represented in (173). Lovett is then expected to take the turn and address the presented topic.

(173) **Management actions by Colbert, Lines 2-3, ex. 172**

$$IS1_1 = \begin{cases} \text{Participants: } SC_{\text{-S}}, JL_{\text{A}} \\ \text{Content : +Trust in polls} \\ \text{Management :} \\ \quad \textsc{Present}(\text{CG1}\rightarrow \\ \quad\quad \text{Trust in polls}) \\ \quad \textsc{Disengage}(\text{CG2}\rightarrow SC_{\text{S}}) \\ \quad \textsc{Request}(\text{CG1+CG2:IS1}_2) \end{cases} \quad IS1_2 = \begin{cases} \text{Participants: } JL_{+\text{S}}, SC_{\text{A}} \\ \text{Content : Trust in polls} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{L}\rightarrow JL_{\text{A}}) \\ \quad \textsc{Refer}(\text{LG}\rightarrow \\ \quad\quad \text{Trust in polls}) \end{cases}$$

In the verbal mode, Lovett acts cooperatively, at least initially. In the gestural mode, however, he does not cooperate. He *disengages* from the Interaction Space between him and Colbert by averting his gaze, and performs a *blocking* gesture as if to refuse the question altogether. Instead of enacting the Interaction Space state Colbert requested, he enacts an almost inverse state, represented in (174).

(174) **Management actions by Lovett, Line 4, ex. 172**

$$IS1_2 = \begin{cases} \text{Participants: } JL_{\text{S}}, SC_{\text{A}} \\ \text{Content : +Trust in polls} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{LG1}\rightarrow \\ \quad\quad \text{Trust in polls}) \\ \quad \textsc{Disengage}(\text{LP2}\rightarrow JL_{\text{S}}) \\ \quad \textsc{Request}(\text{LG1+LP2:IS1}_2) \end{cases} \quad IS1_3 = \begin{cases} \text{Participants: } SC_{+\text{S}}, JL_{\text{A}} \\ \text{Content : } \sout{\text{Trust in polls}} \\ \text{Management :} \\ \quad \textsc{Remove}(\text{CG}\rightarrow \\ \quad\quad \text{Trust in polls}) \end{cases}$$

Again, the requested Interaction Space state fails. Colbert, in good humor, ignores Lovett's hesitation and acts as if his question has been answered. Only after the derailment and several other hedging statements does Lovett finally get to achieve his desired Interaction Space state, rejecting the question of poll trustworthiness and insisting that people should vote, not worry about polls. The blocking gesture is repeated throughout the derailment and Lovett's hedging statements as a reiteration of his request to dismiss the topic. He only transitions out of a blocking gesture to a presentation gesture (LG7 to LG8) after he has secured his turn and thus secured the dismissal of Colbert's desired topic.

7.4.4.2 Pausing within a specification

The final example demonstrates the use of pausing gestures to indicate a break in the intended specification. The gesturer, American politician Nancy Pelosi, uses two pausing gestures, in concert with shifting body orientation, to indicate the start and end of a brief digression.

In this clip, Colbert and American politician Nancy Pelosi are discussing then-president Donald Trump's controversial phone call with the president of Ukraine. The White House had recently released a rough transcript of the phone call, and Pelosi is explaining why that happened. The excerpt begins with Pelosi giving her negative assessment of the phone call (line 5), performing a loose two-handed PUOH gesture, as if to support a circular object (PG1). At this point the audience cheers, and Pelosi turns partially toward them (PP2). She then moves on to her explanation as to the release of the transcript, prefacing her specification with HTT. At "here's the thing", Pelosi performs an addressing gesture, lightly tapping Colbert's hand, as if to make sure he is paying attention (line 7, PG2). She then holds up a pausing gesture toward the audience, as if to ask them to wait a moment (PG3). Pelosi then turns toward the audience for a brief background digression, reminding them that Republicans are claiming that Trump's actions weren't wrong. She performs a two-handed PUOH gesture while introducing the digression (line 8-9, PG4), and then a full body shrug while speaking as a Republican, conveying the innocence they are claiming (lines 8-9, PG5). She then returns to the main Interaction Space between her and Colbert, performing another pausing gesture as she introduces her main point – it was because of a whistle blower that the transcript was released (lines 9-10, PG6). She then elaborates on this point, performing another abbreviated addressing gesture (PG7) and a pinching gesture (PG8) as she reiterates that the transcript would not have been released without the whistle blower (line 11).

(175) Transcript 24: Nancy Pelosi
[UID:6a02fce6-159f-11ea-8a36-089e01ba0335,1991]

```
1   NP   When I went out to uh formalize the inquiry, the inquiry has
2        been going on for a very long time, but a week after that
3        September seventeenth was the morning I have- the president
4        called about how perfect the phone call was and I thought-
5        *my view was that it was perfectly wrong, and- but-*
         *PG1- - - - - - - - - - - - - - - - - - - - - - - - -*
6   --   (cheers)
7   NP   But- *but here's the thing* *to remember* *they make a big
                 *PG2- - - - - - - - -* *PG3- - - -* *PG4 - - - - - ->
8        thing of saying* *"well if it was so wrong, why would the
         <- - - - - - - * *PG5- - - - - - - - - - - - - - - - - ->
9        president put out the notes from the meeting"* *it was a
         <- - - - - - - - - - - - - - - - - - - - - - -* *PG6- - ->
10       whistle blower.* *We would never have known about this*
         <- - - - - - - * *PG7- - - - - - - - - - - - - - - - *
11       *absent the whistle blower coming forward*
         *PG8- - - - - - - - - - - - - - - - - - - - - - -*
PG1:  both hands, palms up in cupping shape, resting on desk
PG2:  left hand reaches, palm down, taps Colbert's right hand
```

```
PG3:  left hand, index finger up, held toward audience
PG4:  two hands, loose palm up, central gesture space
PG5:  two hands, palm up, outward lateral movement, shoulder shrug
PG6:  right hand, index finger up, held toward Colbert
PG7:  right hand, loose palm down held in central Interaction Space
PG8:  right hand, pinching, central Interaction Space
```

Figure 7.10: Gesture sequence corresponding to TRANSCRIPT 24: NANCY PELOSI, ex. 175

The addressing gesture Pelosi performs at HTT functions in the same as the addressing gestures discussed in Section 7.4.3. By performing an ENGAGE, turning her head toward Colbert, and a REFER action upon Colbert, she reinforces the current participant roles. This is represented in (176).

(176) **Management actions by Pelosi, Line 7, ex. 175**

$$IS1_1 = \begin{cases} \text{Participants: } \text{NP}_{+\text{S}}, \text{SC}_{+\text{A}} \\ \text{Content : Transcript release} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{NP3}\rightarrow\text{NP}_{\text{S}}) \\ \quad \textsc{Refer}(\text{NG2}\rightarrow\text{SC}_{\text{A}}) \\ \quad \textsc{Request}(\text{NP3+NG2:IS1}_2) \end{cases} \qquad IS1_2 = \begin{cases} \text{Participants: } \text{NP}_{\text{S}}, \text{SC}_{+\text{A}} \\ \text{Content : Transcript release} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{CG}\rightarrow\text{SC}_{\text{A}}) \end{cases}$$

What is really interesting about this clip is that it is at this point, just after Pelosi has reinforced her role as speaker and requested that Colbert reinforce his role as addressee, that she disengages from the primary Interaction Space. She does this through several gestural strategies: (i) she breaks the speaker-hearer line by averting her gaze (PP4), (ii) she performs a pausing gesture to indicate a request to stop some aspect of the discourse (PG3), and (iii) she reorients her body completely to engage with the secondary Interaction Space between her and the audience (PP5). While in this space she presents claims by the Republicans that are incompatible with her own.[20] Her disengagement with the primary Interaction Space (IS1) and subsequent engagement with the secondary Interaction Space (IS2) is represented in (177).

(177) **Management actions by Pelosi, Lines 7-9, ex. 175**

$$IS1_1 = \left\{ \begin{array}{l} \text{Participants: NP}_{\text{-S}}\text{, SC}_{\text{A}} \\ \text{Content : Transcript release} \\ \text{Management :} \\ \quad \textsc{Disengage}(\text{PG3}\rightarrow \\ \quad\quad \text{Transcript release}) \\ \quad \textsc{Disengage}(\text{PP4}\rightarrow\text{NP}_{\text{S}}) \\ \quad \textsc{Disengage}(\text{PP5}\rightarrow\text{NP}_{\text{S}}) \end{array} \right. \qquad IS2_1 = \left\{ \begin{array}{l} \text{Participants: NP}_{\text{+S}}\text{, Audience} \\ \text{Content : Transcript release} \\ \quad \text{+Republican claims} \\ \text{Management :} \\ \quad \textsc{Engage}(\text{PP5}\rightarrow\text{NP}_{\text{S}}) \\ \quad \textsc{Present}(\text{PG4}\rightarrow) \\ \quad\quad \text{Republican claims}) \end{array} \right.$$

Pelosi then repeats the reorientation process in reverse, performing another pausing gesture (PG6) and an abbreviated addressing gesture (PG7) as she turns back to Colbert. It is only at this point that she continues her specification regarding why the transcript was released.

In Chapter 3, we saw that reorientation across Interaction Spaces can help parse discourse segments into digressive and non-digressive utterances. This is another such example. HTT does preface an important partial answer to the question 'why did Trump release the notes', but the partial answer is not actually in the immediately proceeding utterance. This disjunction between the discourse marker and the actual point Pelosi wishes to make is expressed gesturally rather than verbally. First, Pelosi disengages from the primary Interaction Space between her and Colbert during the intrusive segment that does *not* provide the important specifying information. Second, the beginning and end of the intrusive segment is marked with a one-second gesture (PG3 & PG6). Third, the actual content of the partial answer, that we have the notes because a whistle blower came forward, is marked with a pinching gesture held into the primary central Interaction Space.

[20]This shift may also function to spatially separate Pelosi's claims from the reported Republican claims. We saw a very similar example to this in Section 7.4.1 where Harris physically separated her argument from the hypothetical argument of the other side, but maintained engagement with the primary Interaction Space throughout.

7.5 Conclusion

In this chapter I have explored the gestural expression of specification moves. Presentation and locating gestures express specification directly by introducing and referring to metaphoric objects in the Interaction Space. Presentation gestures in particular can provide additional information as to the complexity and specificity of the information being provided. Addressing gestures perform a very similar function to HTT, drawing attention to the speaker and their contribution. Stopping gestures relate to some parallel discourse structural or social factor rather than the specification move itself, and so can provide information that is not obviously expressed in the verbal mode.

Chapter 8

Concluding remarks

Over the previous seven chapters of this dissertation, I have worked to develop and implement a predictive model for interactive meaning in gesture. The proposed model is centered around *the Interaction Space* – the physical space in which discourse participants co-construct a cooperative discourse. By combining action schematic approaches to gesture meaning (Cienki 2013; Mittelberg 2018, 2019; Müller 2017) with a metaphoric understanding of discourse structure (e.g. Lakoff & Johnson 1980; Reddy 1979), I proposed an inventory of management actions that participants perform relative to the Interaction Space via interactive gestures. PRESENT actions introduce a topic into the Interaction Space as a metaphoric object; REFER actions direct joint attention to a discourse participant or an existing topic in the Interaction Space; REMOVE actions remove a topic from the Interaction Space; and STOP actions request that some subsequent management action be taken. ENGAGE and DISENGAGE actions signal participation, or lack thereof, in the pursuit of immediate discourse goals.

Each management action was motivated by the physical affordances and metaphoric construal of the gestures that enact it. PRESENT actions are achieved through gestures that exhibit the physical affordances of object presentation, such as when an up-turned open hand is extended as if to offer an object for inspection (Cooperrider et al. 2018; Müller 2004). REFER actions are achieved through deictic gestures that do not exhibit the physical affordances of PRESENTATION or REMOVAL and are used to direct joint attention to particular regions of shared space (e.g. McNeill et al. 1993; McNeill 2003). REMOVAL actions are achieved through gestures that exhibit the physical affordances of removing objects from a gesturer's immediate bodily space, such as when down-turned open hands are moved quickly and forcefully outward, as if to clear a surface of objects (e.g. Bressem & Müller 2014, 2017; Harrison 2010). STOP actions are achieved through gestures that exhibit the physical affordances of stopping an incoming object, such as when a flat, open hand is held with the palm facing away from the body, as if to prevent an object from entering immediate bodily space (e.g. Bressem & Wegener 2021; Wehling 2017). Finally, ENGAGE and DISENGAGE actions are performed by non-manual movements that exhibit the physical affordances of enabling or disabling physical interaction with nearby objects, such as when a gesturer leans toward their addressee, as if to be able to better perceive them and objects near them (e.g. Kendon

2010).

Each management action also places restrictions on subsequent actions that can be performed in the Interaction Space. For example, once a topic has been REMOVED from the Interaction Space, it can not longer be REFERRED to in subsequent actions, unless it is reintroduced to the space via a PRESENT action. These restrictions are also motivated by the physical affordances of the space and the gestures used to manage it. For example, removing an object from shared space physically prevents participants from further interaction with it. This motivates the restriction described above. It is these physically-motivated restrictions that make the model predictive – the forms that interactive gestures take relative to the Interaction Space are partially determined by preceding gestures and the impacts they have on the contents of the Interaction Space.

I set two primary goals for the development of this model. First, I sought to demonstrate the systematicity and communicative capacity of interactive gesture by developing a model of interactive meaning around the semiotic affordances of gesture itself. Second, I hoped to show the value of integrating such a model with existing theories of discourse structure, especially question-based models (e.g. van Kuppevelt 1995; Roberts 1996, 2012). I argued that keeping both goals in mind is important if we are to achieve a truly multimodal linguistics.

In the remainder of this final chapter, I summarize the findings of the three case studies conducted (8.1), discuss the broad theoretical implications of the findings (8.2), and suggest directions for future related work (8.3).

8.1 Summary of findings

After developing the Interaction Space model in Chapter 4, I tested the model's predictions in three case studies, each targeting a different kind of discourse move. The first considered interactive gestures accompanying the lexical discourse marker *anyway* in order to explore the multimodal expression of *topic-shifting*. The second considered interactive gestures accompanying *by the way* in order to explore the multimodal expression of *digression*. The third considered interactive gestures accompanying *here's the thing* in order to explore the multimodal expression of *specification*. These three discourse moves, *topic-shifting, digression,* and *specification*, were chosen for their distinct discourse-structural properties, and the distinct impacts they have on the co-construction of a cooperative discourse. *Specification* moves were considered the default move and maximally cooperative because they offer an answer to an open discourse question. *Topic-shifts* were considered maximally disruptive because they abandon an open discourse question in favor of a new one. *Digressions* were considered minimally disruptive because though they do not immediately offer an answer to an open discourse question, they also do not abandon an open discourse question.

Given the affordances and constraints of the proposed Interaction Space model, I made predictions as to what types of interactive gesture should appear with each discourse move. REMOVAL and PRESENTATION gestures should be used during topic-shifts to signal the removal of the abandoned question and the subsequent introduction of a new one. STOPPING

and REFERRING gestures should be used in digressions to signal a pause in the pursuit of discourse goals and the addition of supplementary information to existing discourse topics. REFERRING gestures should be used during specifications to signal the addition of important information to existing topics. One of these four gesture types appeared in three-quarters of the data (231/306). As summarized in Table 8.1, these are the patterns of gestural expression that occur in the data.

	PRESENTATION	REFERRING	REMOVAL	STOPPING	Total
Anyway	**29/106 (27%)**	15/106 (14%)	**25/106 (24%)**	11/106 (10%)	80/106 (75%)
By the way	16/100 (16%)	**23/100 (23%)**	2/100 (2%)	**28/100 (28%)**	69/100 (69%)
Here's the thing	15/100 (15%)	**41/100 (41%)**	2/100 (2%)	24/100 (24%)	82/100 (82%)
Total	60/306 (20%)	79/306 (26%)	29/306 (9%)	63/306 (21%)	231/306 (75%)

Table 8.1: Gestural expression across discourse moves

However, the most exciting thing about these findings is not that they broadly support the predictions made. The most exciting thing is that despite this broad support, there is also significant variation. To understand why this variation is exciting, we must remind ourselves of the nature of our data sets and of the behavior of the lexical discourse markers that shaped them.

The data gathered for the three case studies did not constitute collections of particular discourse moves, but rather collections of proxy expressions (the lexical discourse markers *anyway, by the way* and *here's the thing*) that were thought to indicate the discourse move's presence. The first half of each case study was dedicated to demonstrating the imperfect mapping between discourse marker on the one hand and discourse move on the other. For example, we saw in Chapter 5 that *anyway* could be used with very different *types* of topic-shifts that served very different discursive and social functions. All three lexical discourse markers were shown to be *multifunctional* and *underspecified*. For example, *here's the thing* simultaneously (i) expressed the presence of a specification move and (ii) emphasized the perceived importance of the contribution. This is a demonstration of multifunctionality – one linguistic form simultaneously performs two functions. *Here's the thing* also occurred variably with different discourse relations. For example, some of the marked specifications provided important *background* information about a discourse topic, while others *contradicted* particular aspects of previous contributions. This is a demonstration of underspecification – one linguistic form performs different functions in different contexts.

With this in mind, consider what it would mean if the predictions for gestural discourse markers *were* perfectly met in the present data sets, without variation. This would mean that the gestures were co-expressing the lexical discourse marker, rather than discourse-structural features directly. Though this might be an interesting finding for gesture studies, it would not be an interesting finding for multimodal linguistics more generally – the gestures would not be expressing anything other than what was already expressed in the verbal mode. And if gestures do not express anything other than what is already expressed in the verbal mode, what can the study of gesture offer linguistic theory?

That the gestures in this work *broadly* aligned with predictions while also exhibiting variation is an incredibly important finding for establishing the value of gestural data to linguistic theory. The broad alignment with predictions supports the position that interactive gestures convey discourse-structural information. The observed variation supports the position that interactive gestures convey discourse-structural information *independently from* accompanying verbal cues.

8.2 Theoretical implications

The capacity of gesture to independently contribute to discourse management has significant implications for the development of our theories of discourse structure. Very generally, it means that we can use gesture as a tool for identifying discourse-structural features and testing discourse structure analyses. I will briefly discuss two research areas in which I think the inclusion of interactive gesture analysis would prove particularly meaningful.

First, our theories of discourse moves and relations rely heavily on the overt expression of these moves and relations by things like lexical discourse markers. However, the theories we've built based on these suggest that much of discourse structure remains *implicit*, unexpressed in the verbal mode. This runs the risk of circularity – a given discourse relation is present because the theory predicts that it should be present (e.g. Kim et al. 2020; Lin et al. 2009). Interactive gestures may prove incredibly useful for diagnosing 'implicit' structure and avoiding this circularity. What is *verbally implicit* may be *gesturally explicit*.

Second, interactive gesture can be used in the discovery of polysemy networks in lexical discourse markers. Contrastive approaches to discourse markers suggest that you can learn about the multifunctionality of a lexical item by seeing how it is translated in different use-contexts into different languages (e.g. Aijmer & Simon-Vandenbergen 2003; Cuenca 2008; Fischer 2000; Takahara 1998). Considering the variation in gestural expression with a discourse marker may serve as a kind of within-language contrastive approach – variation in the gestural mode suggests multifunctionality in the verbal mode.

8.3 Future directions

Multimodal discourse analysis is currently an exciting and open field of study, and I believe will remain so for the foreseeable future. This means that there is an indefinite number of interesting and worthwhile questions to ask regarding the gesture-discourse structure interface. I briefly outline three ways in which the particular proposals and findings in the present work can be extended.

First, there are straightforward empirical extensions of this work that I believe are worth pursuing. Most obviously, case studies like those reported on in Chapters 5-7 can be conducted for any number of other lexical discourse markers, both within English and cross-linguistically. I chose to look at *anyway, by the way,* and *here's the thing* because they seemed to me maximally distinct in important ways, each relating to a distinct question-answer structure. This meant that my focus was on identifying interactive gestures that were *unique* to each case study. However, taking the opposite approach would be just as interesting. For example, one could look for *similarities* in the gestures used with *here's the thing*, *the thing is*, and *the fact is*, all of which are associated with emphasized specification. Conducting similar case studies on different types of data, especially naturalistic data, would be an important extension for testing the generalizeability of the proposals in this work.

Second, the questions asked in this work can be addressed using different methodologies. For example, I think that there is an immense potential for experimental work on the use of the Interaction Space. Because it is proposed as a predictive model, one should be able to develop experimental procedures to target and test particular predictions. One very clear prediction is that *referring* gestures should not be felicitous if the relevant discourse topic has been *removed* from the Interaction Space. Designing experimental production tasks to test this prediction should not prove particularly difficult. This would compliment existing experimental work on the use of gesture in establishing common ground (e.g. Holler 2009; Holler et al. 2011; Holler & Wilkin 2011).

Finally, there are theoretical extensions that can be made based on this work. Of central interest to me is the integration of the proposed Interaction Space model with the Question Under Discussion framework. Throughout this work I have discussed alignments between the two – each management action in the Interaction Space corresponded to a discourse move in the QUD structure. However, these alignments were not formalized. Formalizing this integration would be an important step toward a truly multimodal theory of discourse structure.

I look forward to participating in the multimodal discourse analysis community as it grows. I also look forward to seeing strengthened collaborations between gesture researchers and scholars of particular discourse structural frameworks, especially Question Under Discussion, but also Segmented Discourse Representational Theory and Rhetorical Structure Theory. There is much collaborative work to do if we are to improve our understanding of both interactive gesture and discourse structure. I hope this work can serve as an inspiration for such collaborations.

Bibliography

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and linguistics compass*, 9(11), 437–451.

Aijmer, K. (2002). *English discourse particles: Evidence from a corpus*. John Benjamins Publishing.

Aijmer, K. (2007). The interface between discourse and grammar: The fact is that. *Pragmatics & beyond*, 161, 31–48.

Aijmer, K. (2013). *Understanding pragmatic markers: A variational pragmatic approach*. Edinburgh: Edinburgh University Press.

Aijmer, K. & Simon-Vandenbergen, A.-M. (2003). The discourse particle well and its equivalents in Swedish and Dutch. *Linguistics*, 41(6), 1123–1161.

Alahverdzhieva, K., Lascarides, A., & Flickinger, D. (2017). Aligning speech and co-speech gesture in a constraint-based grammar. *Journal of Language Modelling*, 5.

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.

Ariel, M. (2010). *Defining pragmatics*. Cambridge University Press.

Asher, N. (1993). *Reference to abstract objects in discourse*. Dordrecht: Kluwer Academic Publishers.

Asher, N. & Lascarides, A. (2003). *Logics of Conversation*. Cambridge University Press.

Asr, F. T. & Demberg, V. (2012). Implicitness of discourse relations. In *Proceedings of COLING 2012* (pp. 2669–2684).

Azar, Z. & Özyürek, A. (2015). Discourse management: Reference tracking in speech and gesture in Turkish narratives. *Dutch Journal of Applied Linguistics*, 4(2), 222–240.

Battersby, S. A. (2011). *Moving Together: The organisation of non-verbal cues during multiparty conversation*. PhD thesis, Queen Mary University of London.

Bavelas, J., 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.

Bavelas, J. B. (1994). Gestures as part of speech: Methodological implications. *Research on language and social interaction*, 27(3), 201–221.

Bavelas, J. B. & Chovil, N. (2018). Some pragmatic functions of conversational facial gestures. *Gesture*, 17(1), 98–127.

Bavelas, J. B., Chovil, N., Coates, L., & Roe, L. (1995). Gestures specialized for dialogue. *Personality and social psychology bulletin*, 21(4), 394–405.

Bavelas, J. B., Chovil, N., Lawrie, D. A., & Wade, A. (1992). Interactive gestures. *Discourse processes*, 15(4), 469–489.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, 52(3), 566–580.

Beattie, G. (2004). *Visible thought: The new psychology of body language*. Routledge.

Beattie, G. & Shovelton, H. (1999). Mapping the range of information contained in the iconic hand gestures that accompany spontaneous speech. *Journal of language and social psychology*, 18(4), 438–462.

Beattie, G. W. (1981). The regulation of speaker turns in face-to-face conversation: Some implications for conversation in sound-only communication channels. *Semiotica*, 34, 55–70.

Bell, D. M. (1998). Cancellative discourse markers: a core/periphery approach. *Pragmatics*, 8(4), 515–541.

Bergen, B. & Chang, N. (2013). Embodied construction grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 168–190). Oxford University Press.

Bohle, U. (2014). Gesture and conversational units. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1560–1567). De Gruyter Mouton Berlin & Boston.

Bolden, G. B. (2009). Implementing incipient actions: The discourse marker 'so' in English conversation. *Journal of pragmatics*, 41(5), 974–998.

Bolinger, D. (1986). *Intonation and its parts: Melody in spoken English*. Stanford University Press.

Bressem, J. & Ladewig, S. H. (2011). Rethinking gesture phases: Articulatory features of gestural movement? *Semiotica*, 2011(184), 53–91.

Bressem, J. & Müller, C. (2014). The family of away gestures: Negation, refusal, and negative assessment. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1592–1604). De Gruyter Mouton Berlin & Boston.

Bressem, J. & Müller, C. (2017). The "negative-assessment-construction"–a multimodal pattern based on a recurrent gesture? *Linguistics Vanguard*, 3(s1).

Bressem, J., Stein, N., & Wegener, C. (2017). Multimodal language use in Savosavo: Refusing, excluding and negating with speech and gesture. *Pragmatics*, 27(2), 173–206.

Bressem, J. & Wegener, C. (2021). Handling talk: A cross-linguistic perspective on discursive functions of gestures in German and Savosavo. *Gesture*, 20(2), 219–253.

Brône, G., Oben, B., Jehoul, A., Vranjes, J., & Feyaerts, K. (2017). Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics*, 28(3), 449–483.

Brugman, C. & Lakoff, G. (1988). Cognitive topology and lexical networks. In S. Small, G. Cottrell, & M. Tananhaus (Eds.), *Lexical ambiguity resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence* (pp. 477–508). San Mateo, CA: Morgan Kaufmann.

Brunner, L. J. (1979). Smiles can be back channels. *Journal of personality and social psychology*, 37(5), 728–734.

Bublitz, W. (1988). *Supportive fellow-speakers and cooperative conversations*. Amsterdam: John Benjamins.

Bühler, K. (1982). The deictic field of language and deictic words. *Speech, place, and action*, (pp. 9–30).

Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmerer (Eds.), *Usage-based models of language* (pp. 65–85). Stanford:CSLI.

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 *Metaphor and gesture* (pp. 27–53). John Benjamins.

Calbris, G. (2013). Elements of meaning in gesture: the analogical links. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 1 (pp. 658–674). Berlin/Boston: De Gruyter Mouton.

Carston, R. (2002). *Thoughts and utterances: The pragmatics of explicit communication*. Oxford: Blackwell Publishers.

Casasanto, D. & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *Plos one*, 5(7), e11805.

Chambers, J. K. & Schilling, N., Eds. (2018). *The handbook of language variation and change*. John Wiley & Sons.

Charolles, M. (2020). Discourse topics and digressive markers. *Journal of Pragmatics*, 161, 57–77.

Chen, L. & Harper, M. P. (2009). Multimodal floor control shift detection. In *Proceedings of the 2009 international conference on Multimodal interfaces* (pp. 15–22).

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: The MIT press.

Chu, M., Meyer, A., Foulkes, L., & Kita, S. (2014). Individual differences in frequency and saliency of speech-accompanying gestures: the role of cognitive abilities and empathy. *Journal of Experimental Psychology*, 143(2), 694–709.

Church, R. B., Alibali, M. W., & Kelly, S. D., Eds. (2017). *Why Gesture?: How the hands function in speaking, thinking and communicating*. John Benjamins.

Church, R. B., Ayman-Nolley, S., & Mahootian, S. (2004). The role of gesture in bilingual education: Does gesture enhance learning? *International Journal of Bilingual Education and Bilingualism*, 7(4), 303–319.

Cienki, A. (2005). Image schemas and gesture. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics*, volume 29 (pp. 421–442). Berlin: de Gruyter Mouton.

Cienki, A. (2013). Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics. Published Under the Auspices of the Spanish Cognitive Linguistics Association*, 11(2), 417–432.

Cienki, A. (2017). Utterance Construction Grammar (UCxG) and the variable multimodality of constructions. *Linguistics Vanguard*, 3(s1).

Cienki, A. (2022). The study of gesture in cognitive linguistics: How it could inform and inspire other research in cognitive science. *Wiley Interdisciplinary Reviews: Cognitive Science*, (pp. e1623).

Clark, H. & Schaefer, E. F. (1989). Contributing to discourse. *Cognitive Science*, 13, 259–294.

Clark, H. H. & Marshall, C. R. (1981). Definite knowledge and mutual knowledge. In a. Joshi, B. Webber, & I. Sag (Eds.), *Elements of discourse understanding* (pp. 10–63). Cambridge University Press.

Clark, H. H. & Tree, J. E. F. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111.

Cohen, A. A. & Harrison, R. P. (1973). Intentionality in the use of hand illustrators in face-to-face communication situations. *Journal of personality and social psychology*, 28(2), 276–279.

Collier, D. & Levitsky, S. (1997). Democracy with adjectives: Conceptual innovation in comparative research. *World politics*, 49(3), 430–451.

Cooperrider, K. (2017). Foreground gesture, background gesture. *Gesture*, 16(2), 176–202.

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The palm-up puzzle: Meanings and origins of a widespread form in gesture and sign. *Frontiers in Communication*, 3, 23.

Croft, W. & Cruse, D. A. (2004). *Cognitive linguistics*. Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge University Press.

Cuenca, M.-J. (2008). Pragmatic markers in contrast: The case of well. *Journal of pragmatics*, 40(8), 1373–1391.

Dalrymple, M. (2001). *Lexical functional grammar*. Brill.

Dancygier, B. & Sweetser, E. (2005). *Mental spaces in grammar: Conditional constructions*, volume 108. Cambridge University Press.

Dancygier, B. & Sweetser, E. (2014). *Figurative language*. Cambridge University Press.

de Vignemont, F. & Iannetti, G. (2015). How many peripersonal spaces? *Neuropsychologia*, 70, 327–334.

Debras, C. (2017). The shrug: Forms and meanings of a compound enactment. *Gesture*, 16(1), 1–34.

Deppermann, A. (2013). Multimodal interaction from a conversation analytic perspective. *Journal of pragmatics*, 46(1), 1–7.

Dingemanse, M. (2012). Advances in the cross-linguistic study of ideophones. *Language and Linguistics compass*, 6(10), 654–672.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in cognitive sciences*, 19(10), 603–615.

Dittmann, A. T. & Llewellyn, L. G. (1968). Relationship between vocalizations and head nods as listener responses. *Journal of personality and social psychology*, 9(1), 79–84.

Dorgeloh, H. (2004). Conjunction in sentence and discourse: sentence-initial and and discourse structure. *Journal of Pragmatics*, 36(10), 1761–1779.

Drijvers, L. & Holler, J. (2022). The multimodal facilitation effect in human communication. *Psychonomic Bulletin & Review*, (pp. 1–10).

Dudis, P. G. (2004). Body partitioning and real-space blends. *Cognitive Linguistics*, 15(2), 223–238.

Duncan, S. (1974). On the structure of speaker–auditor interaction during speaking turns. *Language in society*, 3(2), 161–180.

Dynel, M. (2009). Where cooperation meets politeness: Revisiting politeness models in view of the Gricean framework. *Brno Studies in English*, 35(1), 23–43.

Efron, D. (1941). *Gesture, race, and culture*. Mouton, The Hague.

Ekman, P. & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Nonverbal communication, interaction, and gesture*, (pp. 57–106).

Enfield, N. J. (2009). *The anatomy of meaning: Speech, gesture, and composite utterances*. Cambridge University Press.

Enfield, N. J., Kita, S., & De Ruiter, J. P. (2007). Primary and secondary pragmatic functions of pointing gestures. *Journal of Pragmatics*, 39(10), 1722–1741.

Erteschik-Shir, N. (2007). *Information structure: The syntax-discourse interface*, volume 3. OUP Oxford.

Evans, N., Bergqvist, H., & San Roque, L. (2018). The grammar of engagement I: Framework and initial exemplification. *Language and Cognition*, 10(1), 110–140.

Falkum, I. L. (2015). The how and why of polysemy: A pragmatic account. *Lingua*, 157, 83–99.

Fauconnier, G. (1994). *Mental spaces: Aspects of meaning construction in natural language*. Cambridge University Press.

Fauconnier, G. & Turner, M. (2008). *The way we think: Conceptual blending and the mind's hidden complexities*. Basic Books.

Fauconnier, G. & Turner, M. B. (1998). Blending as a central process of grammar: Expanded version. In A. Goldberg (Ed.), *Conceptual Structure, Discourse, and Language* (pp. 113–130).

Ferrara, K. (1994). *Therapeutic ways with words*. Oxford University Press on Demand.

Ferrara, K. W. (1997). Form and function of the discourse marker anyway: implications for discourse analysis. *Linguistics*, 35, 343–378.

Ferrara, L. & Hodge, G. (2018). Language as description, indication, and depiction. *Frontiers in Psychology*, 9, 716.

Ferré, G. (2012). Functions of three open-palm hand gestures. *Journal Multimodal Communication*, 1(1), 5–20.

Fillmore, C. J. (1976). Frame semantics and the nature of language. In *Annals of the New York Academy of Sciences: Conference on the origin and development of language and speech*, volume 280 (pp. 20–32).: New York.

Fischer, K. (2000). *From cognitive semantics to lexical pragmatics: the functional polysemy of discourse particles*. Berlin: De Gruyter Mouton.

Flecha-García, M. L. (2010). Eyebrow raises in dialogue and their relation to discourse structure, utterance function and pitch accents in english. *Speech communication*, 52(6), 542–554.

Fraser, B. (1988). Types of English discourse markers. *Acta Linguistica Hungarica*, 38(1/4), 19–33.

Fraser, B. (1996). Pragmatic markers. *Pragmatics*, 6, 167–190.

Fraser, B. (1999). What are discourse markers? *Journal of pragmatics*, 31(7), 931–952.

Fraser, B. (2009). Topic orientation markers. *Journal of pragmatics*, 41(5), 892–898.

Fricke, E. (2014). Kinesthemes: Morphological complexity in co-speech gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1618–1629).

Fricke, E., Bressem, J., & Müller, C. (2014). Gesture families and gestural fields. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1630–1640). Berlin/Boston: De Gruyter Mouton.

Fried, M. & Östman, J.-O. (2004). Construction grammar: A thumbnail sketch. *Construction Grammar in a cross-language perspective*, (pp. 11–86).

Gaddy, M. L., van den Broek, P., & Sung, Y.-C. (2001). The influence of text cues on the allocation of attention during reading. In T. Sanders, J. Schilperoord, & W. Spooren (Eds.), *Text representation: Linguistic and psycholinguistic aspects* (pp. 89–110). Amsterdam: John Benjamins.

Gallie, W. B. (1955). Essentially contested concepts. In *Proceedings of the Aristotelian society*, volume 56 (pp. 167–198).: JSTOR.

Gallie, W. B. (1956). Art as an essentially contested concept. *The Philosophical Quarterly (1950-)*, 6(23), 97–114.

Gawne, L. (2021). 'Away' gestures associated with negative expressions in narrative discourse in Syuba (Kagate, Nepal) speakers. *Semiotica*, 239, 37–59.

Gibbs, R. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge: Cambridge University Press.

Gibbs, R. W. (2014). Conceptual metaphor in thought and social action. In M. Landau, M. Robinson, & B. Meier (Eds.), *The Power of metaphor: Examining its influence in Social Life* (pp. 17–40). American Psychological Association.

Gill, S. P., Kawamori, M., Katagiri, Y., & Shimojima, A. (2000). The role of body moves in dialogue. *International Journal of Language and Communication*, 12, 89–114.

Ginzburg, J. & Poesio, M. (2016). Grammar is a system that characterizes talk in interaction. *Frontiers in Psychology*, 7, 1938.

Givón, T. (1983). *Topic continuity in discourse*. Amsterdam: John Benjamins.

Givón, T. (1985). Iconicity, isomorphism and non-arbitrary coding in syntax. In J. Haiman (Ed.), *Iconicity in syntax* (pp. 187–219). Amsterdam: John Benjamins.

Givón, T. (1989). *Mind, code and context: Essays in pragmatics*. Hillsdale, NJ: Lawrence Erlbaum Associatess.

Glynn, S., Britton, B., & Tillman, M. (1982). Typographical cues in text: Management of the reader's attention. In D. Jonassen (Ed.), *The technology of text*, volume 2 (pp. 192–209). Englewoods Cliffs, NJ: Educational Technology Publications.

Goddard, C. (2012). Semantic primes, semantic molecules, semantic templates: Key concepts in the NSM approach to lexical typology. *Linguistics*, 50(3), 711–743.

Goffman, E. (1956). The nature of deference and demeanor. *American anthropologist*, 58(3), 473–502.

Goffman, E. (1963). *Behavior in public places*. New York: The Free Press.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.

Goldin-Meadow, S. (2017). Using our hands to change our minds. *Wiley Interdisciplinary Reviews: Cognitive Science*, 8(1-2), e1368.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Goodwin, C. (2000). Action and embodiment within situated human interaction. *Journal of pragmatics*, 32(10), 1489–1522.

Goodwin, M. & Goodwin, C. (1986). Gesture and coparticipation in the activity of searching for a word. *Semiotica*, 62(1), 51–75.

Grady, J. (1998). The "conduit metaphor" revisited: A reassessment of metaphors for communication. In J.-P. Koenig (Ed.), *Discourse and cognition: Bridging the gap* (pp. 205–218). Stanford, CA: CLSI Publications Stanford.

Grice, H. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*, volume 3 (pp. 41–58). New York: Seminar Press.

Grosz, B. J. & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational linguistics*, 12(3), 175–204.

Gullberg, M. & Holmqvist, K. (2006). What speakers do and what addressees look at: Visual attention to gestures in human interaction live and on video. *Pragmatics & Cognition*, 14(1), 53–82.

Gunnel, T. (2014). On the use of uh and um in American English. *Functions of Language*, 21(1), 6–29.

Hadi, A. et al. (2013). A critical appraisal of Grice's Cooperative Principle. *Open journal of modern linguistics*, 3(1), 69–72.

Hagemann, J. (2014). Proxemics and axial orientation. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1310–1323). Berlin: De Gruyter Mouton.

Hall, E. (1995). *Handbook for proxemic research*. Washington, DC: Society for the Anthropology of Visual Communication.

Halliday, M. (1985). *An Introduction to Functional Grammar*. London: Edward Arnold.

Harris, R. (1980). *The language-makers*. Ithaca, NY: Cornell University Press.

Harris, R. (1981). *The language myth*. London: Duckworth.

Harris, R. (1993). *The Linguistics Wars*. Oxford University Press on Demand.

Harris, R. (2002). *The language myth in Western culture*. Richmond, Surrey: Curzon Press.

Harrison, S. (2010). Evidence for node and scope of negation in coverbal gesture. *Gesture*, 10(1), 29–51.

Harrison, S. (2014). Head shakes: Variation in form, function, and cultural distribution of a head movement related to "no". In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction* (pp. 1496–1501). Berlin: De Gruyter Mouton.

Harrison, S. (2018). *The impulse to gesture: Where language, minds, and bodies intersect*. Cambridge University Press.

Hart, C. & Winter, B. (2022). Gesture and legitimation in the anti-immigration discourse of Nigel Farage. *Discourse & Society*, 33(1), 34–55.

Haspelmath, M. (2006). Against markedness (and what to replace it with). *Journal of linguistics*, 42(1), 25–70.

Haspelmath, M. & Michaelis, S. M. (2017). Analytic and synthetic. In I. Buchstaller & B. Siebenhaar (Eds.), *Language Variation-European Perspectives VI: Selected papers from the Eighth International Conference on Language Variation in Europe (ICLaVE 8), Leipzig 2015*: Amsterdam: John Benjamins.

Hayes, B. P. (1999). Phonetically driven phonology. *Functionalism and formalism in linguistics*, 1, 243–285.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29–29.

Heylen, D., Bevacqua, E., Pelachaud, C., Poggi, I., Gratch, J., & Schröder, M. (2011). Generating listening behaviour. In P. Petta, C. Pelachaud, & R. Cowie (Eds.), *Emotion-oriented systems* (pp. 321–347). Springer.

Hinnell, J. (2019). The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics*, 35, 55–92.

Hinnell, J. (2020). *Language in the body: multimodality in grammar and discourse*. PhD thesis, University of Alberta.

Hockett, C. F. (1978). In search of Jove's brow. *American speech*, 53(4), 243–313.

Hockett, C. F. & Hockett, C. D. (1960). The origin of speech. *Scientific American*, 203(3), 88–97.

Hoffmann, T. & Trousdale, G. (2013). *The Oxford handbook of construction grammar*. Oxford University Press.

Holler, J. (2009). Speakers' use of interactive gestures as markers of common ground. In *International Gesture Workshop* (pp. 11–22).: Springer.

Holler, J. & Bavelas, J. (2017). Multi-modal communication of common ground. In R. Breckinridge Church, M. Alibali, & S. Kelly (Eds.), *Why gesture? How the hands function in speaking, thinking, and communicating* (pp. 213–240).

Holler, J. & Beattie, G. (2003). Pragmatic aspects of representational gestures: Do speakers use them to clarify verbal ambiguity for the listener? *Gesture*, 3(2), 127–154.

Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic bulletin & review*, 25(5), 1900–1908.

Holler, J. & Levinson, S. C. (2019). Multimodal language processing in human communication. *Trends in Cognitive Sciences*, 23(8), 639–652.

Holler, J., Tutton, M., & Wilkin, K. (2011). Co-speech gestures in the process of meaning coordination. In *2nd GESPIN-Gesture & Speech in Interaction*: Bielefeld, Germany.

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.

Hornstein, N., Nunes, J., & Grohmann, K. K. (2005). *Understanding minimalism*. Cambridge University Press.

Hostetter, A. B. & Alibali, M. W. (2011). Cognitive skills and gesture–speech redundancy: Formulation difficulty or communicative strategy? *Gesture*, 11(1), 40–60.

Hovy, E. (1990). Parsimonious and profligate approaches to the question of discourse structure relations. In *Proceedings of the Fifth International Workshop on Text Generation* (pp. 59–65).

Hunley, S. B. & Lourenco, S. F. (2018). What is peripersonal space? an examination of unresolved empirical issues and emerging findings. *Wiley interdisciplinary reviews: Cognitive science*, 9(6), e1472.

Inbar, A. & Shor, L. (2019). Covert negation in Israeli Hebrew: Evidence from co-speech gestures. *Journal of Pragmatics*, 143, 85–95.

Jacobs, N. & Garnham, A. (2007). The role of conversational hand gestures in a narrative task. *Journal of Memory and Language*, 56(2), 291–303.

Jakobs, M. & Hüning, M. (2022). Scholars and their metaphors: on language making in linguistics. *International Journal of the Sociology of Language*, 2022(274), 29–50.

Jannedy, S. & Mendoza-Denton, N. (2005). Structuring information through gesture and intonation. *Interdisciplinary studies on information structure*, 3, 199–244.

Janzen, T. (2004). Space rotation, perspective shift, and verb morphology in asl. *Cognitive Linguistics*, 15(2), 149–174.

Janzen, T. (2012). Two ways of conceptualizing space: motivating the use of static and rotated vantage point space in asl discourse. In B. Dancygier & E. Sweetser (Eds.), *Viewpoint in language: A multimodal perspective* (pp. 156–174). Cambridge University Press.

Jasanoff, S. S. (1987). Contested boundaries in policy-relevant science. *Social studies of science*, 17(2), 195–230.

Jehoul, A., Brône, G., & Feyaerts, K. (2017). The shrug as marker of obviousness. *Linguistics Vanguard*, 3(s1).

Johnson, M. (1987). *The body in the mind: The bodily basis of meaning, imagination, and reason*. Chicago: University of Chicago Press.

Johnston, T. & Schembri, A. (2007). *Australian Sign Language (Auslan): An introduction to sign language linguistics*. Cambridge University Press.

Jokinen, K., Furukawa, H., Nishida, M., & Yamamoto, S. (2013). Gaze and turn-taking behavior in casual conversational interactions. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 3(2), 1–30.

Jokinen, K., Nishida, M., & Yamamoto, S. (2009). Eye-gaze experiments for conversation monitoring. In *Proceedings of the 3rd international universal communication symposium* (pp. 303–308).

Joo, J., Steen, F., & Turner, M. (2017). Red Hen Lab: Dataset and tools for multimodal human communication research. *KI-Künstliche Intelligenz*, 31(4), 357–361.

Jucker, A. H. & Ziv, Y. (1998). *Discourse markers: Descriptions and theory*, volume 57. John Benjamins Publishing.

Kamalski, J., Lentz, L., Sanders, T., & Zwaan, R. A. (2008). The forewarning effect of coherence markers in persuasive discourse: Evidence from persuasion and processing. *Discourse Processes*, 45(6), 545–579.

Kamp, H. & Reyle, U. (1993). *From discourse to logic*. Kluwer, Dordrecht.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22–63.

Kendon, A. (1972). Some relationships between body motion and speech. In A. Siegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). Elmsford, NY: Pergamon.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. *The relationship of verbal and nonverbal communication*, (25), 207–227.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters*, volume 7. CUP Archive.

Kendon, A. (1992). The negotiation of context in face-to-face interaction. *Rethinking Context:Language as in interactive phenomenon*.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of pragmatics*, 23(3), 247–279.

Kendon, A. (2000). Language and gesture: Unity or duality. In D. McNeill (Ed.), *Language and gesture* (pp. 47–63). Cambridge University Press.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kendon, A. (2010). Spacing and orientation in co-present interaction. In *Development of multimodal interfaces: Active listening and synchrony* (pp. 1–15). Springer.

Kendon, A. (2014). Semiotic diversity in utterance production and the concept of 'language'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1651).

Kendon, A. (2017). Pragmatic functions of gestures: Some observations on the history of their study and their nature. *Gesture*, 16(2), 157–175.

Kim, N., Feng, S., Gunasekara, C., & Lastras, L. (2020). Implicit discourse relation classification: We need to talk about evaluation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 5404–5414).: Association for Computational Linguistics.

Kipp, M., Neff, M., & Albrecht, I. (2007). An annotation scheme for conversational gestures: how to economically capture timing and form. *Language Resources and Evaluation*, 41(3-4), 325–339.

Kita, S. (2003). *Pointing: Where language, culture, and cognition meet*. Psychology 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.

Kok, K., Bergmann, K., Cienki, A., & Kopp, S. (2016). Mapping out the multifunctionality of speakers' gestures. *Gesture*, 15(1), 37–59.

Kok, K. & Cienki, A. (2016). Cognitive grammar and gesture: Points of convergence, advances and challenges. *Cognitive Linguistics*, 27(1), 67–100.

Krifka, M. (2008). Basic notions of information structure. *Acta Linguistica Hungarica*, 55(3-4), 243–276.

Ladewig, S. (2011). Putting the cyclic gesture on a cognitive basis. *CogniTextes. Revue de l'Association française de linguistique cognitive*, (Volume 6).

Ladewig, S. (2014a). The cyclic gesture. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1605–1618). Berlin/Boston: De Gruyter Mouton.

Ladewig, S. (2014b). Recurrent gestures. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1558–1575). Berlin/Boston: De Gruyter Mouton.

Lakoff, G. (2008). *Women, fire, and dangerous things: What categories reveal about the mind*. University of Chicago press.

Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.

Lakoff, R. (1973). The logic of politeness: Or, minding your p's and q's. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, volume 9 (pp. 292–305).: Chicago Linguistic Society.

Lambrecht, K. (1996). *Information structure and sentence form: Topic, focus, and the mental representations of discourse referents*, volume 71. Cambridge university press.

Landau, M. J., Meier, B. P., & Keefer, L. A. (2010). A metaphor-enriched social cognition. *Psychological bulletin*, 136(6), 1045.

Laparle, S. (2020). At the syntax-pragmatics interface: a quantitative study of aspect in Locative Inversion. In M. Asatryan, Y. Song, , & A. Whitmal (Eds.), *NELS 50: Proceedings of the Fiftieth Annual Meeting of the North East Linguistic Society*, volume 3 (pp. 153–162).: Amherst, MA: GLSA.

Laparle, S. (2021). Tracking discourse topics in co-speech gesture. In *International Conference on Human-Computer Interaction* (pp. 233–249).: Springer.

Laparle, S. (2022). The interaction space. In *International Conference on Human-Computer Interaction* (pp. 243–262).: Springer.

Lascarides, A. & Stone, M. (2009a). Discourse coherence and gesture interpretation. *Gesture*, 9(2), 147–180.

Lascarides, A. & Stone, M. (2009b). A formal semantic analysis of gesture. *Journal of Semantics*, 26(4), 393–449.

Le Lan, B. (2007). Orchestrating conversation: The multifunctionality of well and y'know in the joint construction of a verbal interaction. In A. Celle & R. Huart (Eds.), *Connectives as discourse landmarks* (pp. 103–116). John Benjamins.

Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, 1(2), 207–246.

Leonard, T. & Cummins, F. (2011). The temporal relation between beat gestures and speech. *Language and Cognitive Processes*, 26(10), 1457–1471.

Lewandowska-Tomaszczyk, B. (2007). Polysemy, prototypes, and radial categories. In H. Geeraerts, D Cuyckens (Ed.), *The Oxford handbook of cognitive linguistics* (pp. 139–169). Oxford: Oxford University Press.

Liddell, S. K. (2003). *Grammar, gesture, and meaning in American Sign Language*. Cambridge University Press.

Liddell, S. K. & Metzger, M. (1998). Gesture in sign language discourse. *Journal of pragmatics*, 30(6), 657–697.

Lin, Z., Kan, M.-Y., & Ng, H. T. (2009). Recognizing implicit discourse relations in the penn discourse treebank. In *Proceedings of the 2009 conference on empirical methods in natural language processing* (pp. 343–351).

Linell, P. (2005). *The written language bias in linguistics: Its nature, origins and transformations*. Routledge.

Local, J. (1992). Continuing and restarting. In P. Auer & A. di Luzio (Eds.), *The contextualization of language* (pp. 273–296). Amsterdam: John Benjamins.

Loehr, D. P. (2012). Temporal, structural, and pragmatic synchrony between intonation and gesture. *Laboratory phonology*, 3(1), 71–89.

Loock, R. (2007). Appositive relative clauses and their functions in discourse. *Journal of pragmatics*, 39(2), 336–362.

Macuch Silva, V., Holler, J., Ozyurek, A., & Roberts, S. G. (2020). Multimodality and the origin of a novel communication system in face-to-face interaction. *Royal Society open science*, 7(1), 182056.

Mann, W. C. & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text*, 8(3), 243–281.

Marrese, O. H., Raymond, C. W., Fox, B. A., Ford, C. E., & Pielke, M. (2021). The grammar of obviousness: The palm-up gesture in argument sequences. *Frontiers in Communication*, 6, 663067.

Matthiessen, C. & Halliday, M. (2009). Systemic functional grammar: A first step into the theory.

McClave, E. (2000). Linguistic functions of head movements in the context of speech. *Journal of pragmatics*, 32(7), 855–878.

McKee, R. L. & Wallingford, S. (2011). 'So, well, whatever': Discourse functions of palm-up in New Zealand Sign Language. *Sign Language & Linguistics*, 14(2), 213–247.

McNeill, D. (1985). So you think gestures are nonverbal? *Psychological review*, 92(3), 350–371.

McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.

McNeill, D. (2000). *Language and gesture*. Cambridge University Press Cambridge.

McNeill, D. (2003). Pointing and morality in Chicago. In S. Kita (Ed.), *Pointing: Where language, culture, and cognition meet* (pp. 293–306). Psychology Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

McNeill, D., Cassell, J., & Levy, E. T. (1993). Abstract deixis. *Semiotica*, 95(1-2), 5–20.

McNeill, D. & Levy, E. (1982). Conceptual representations in language activity and gesture. *Speech, place, and action*, (pp. 271–295).

McNeill, D., Quek, F., McCullough, K.-E., Duncan, S. D., Furuyama, N., Bryll, R., & Ansari, R. (2001). Catchments, prosody and discourse. *Gesture*, 1(1), 9–33.

Melinger, A. & Levelt, W. J. (2004). Gesture and the communicative intention of the speaker. *Gesture*, 4(2), 119–141.

Mittelberg, I. (2017). Multimodal existential constructions in German: Manual actions of giving as experiential substrate for grammatical and gestural patterns. *Linguistics Vanguard*, 3(s1).

Mittelberg, I. (2018). Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses. *Cognitive semiotics*, 11(1).

Mittelberg, I. (2019). Visuo-kinetic signs are inherently metonymic: How embodied metonymy motivates forms, functions, and schematic patterns in gesture. *Frontiers in Psychology*, 10, 254.

Mittwoch, A., Huddleston, R. D., & Collins, P. (2002). The clause: adjuncts. In R. Huddleston & G. Pullum (Eds.), *The Cambridge Grammar of the English Language* chapter 8. Cambridge University Press.

Mondada, L. (2007). Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse studies*, 9(2), 194–225.

Mondada, L. (2016). Challenges of multimodality: Language and the body in social interaction. *Journal of Sociolinguistics*, 20(3), 336–366.

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

Müller, C. (2004). Forms and uses of the palm up open hand: A case of a gesture family. *The semantics and pragmatics of everyday gestures*, 9, 233–256.

Müller, C. (2014). Ring-gestures across cultures and times: Dimensions of variation. In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1511–1522). Berlin: De Gruyter Mouton.

Müller, C. (2017). How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture*, 16(2), 277–304.

Müller, C. (2018). Gesture and sign: Cataclysmic break or dynamic relations? *Frontiers in Psychology*, 9, 1651.

Murray, J. D. (1997). Connectives and narrative text: The role of continuity. *Memory & Cognition*, 25(2), 227–236.

Murray, S. E. (2014). Varieties of update. *Semantics and Pragmatics*, 7, 1–53.

Neff, M., Toothman, N., Bowmani, R., Fox Tree, J. E., & Walker, M. A. (2011). Don't scratch! self-adaptors reflect emotional stability. In *International Workshop on Intelligent Virtual Agents* (pp. 398–411).: Springer.

Nevile, M. (2006). Making sequentiality salient: And-prefacing in the talk of airline pilots. *Discourse studies*, 8(2), 279–302.

Nevile, M. (2015). The embodied turn in research on language and social interaction. *Research on Language and Social Interaction*, 48(2), 121–151.

Newmeyer, F. (1992). Iconicity and generative grammar. *Language*, (pp. 756–796).

Norris, S. (2006). Multiparty interaction: a multimodal perspective on relevance. *Discourse studies*, 8(3), 401–421.

Oakley, T. (2017). Multimodal rhetoric: Fictive interaction strategies in political discourse. *Linguistics Vanguard*, 3(s1).

Ottati, V. C. & Renstrom, R. A. (2010). Metaphor and persuasive communication: A multifunctional approach. *Social and Personality Psychology Compass*, 4(9), 783–794.

Ozerov, P. (2022). Alternatives to QUD: Alternatives to questions. *Functions of Language*, 29(1), 86–115.

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.

Özyürek, A. & Kita, S. (1999). Expressing manner and path in English and Turkish: Differences in speech, gesture, and conceptualization. In *Twenty-first Annual Conference of the Cognitive Science Society* (pp. 507–512).: Erlbaum.

Park, I. (2010). Marking an impasse: The use of anyway as a sequence-closing device. *Journal of Pragmatics*, 42(12), 3283–3299.

Parrill, F. (2008). Form, meaning, and convention: A comparison of a metaphoric gesture with an emblem. *Metaphor and gesture*, (pp. 195–217).

Parrill, F. (2009). Dual viewpoint gestures. *Gesture*, 9(3), 271–289.

Parrill, F. (2010). Viewpoint in speech–gesture integration: Linguistic structure, discourse structure, and event structure. *Language and Cognitive Processes*, 25(5), 650–668.

Parrill, F. (2012). Interactions between discourse status and viewpoint in co-speech gesture. In B. Dancygier & E. Sweetser (Eds.), *Viewpoint in language: A multimodal perspective* (pp. 97–112). Cambridge University Press.

Parrill, F. & Sweetser, E. (2004). What we mean by meaning: Conceptual integration in gesture analysis and transcription. *Gesture*, 4(2), 197–219.

Peeters, D., Krahmer, E., & Maes, A. (2021). A conceptual framework for the study of demonstrative reference. *Psychonomic Bulletin & Review*, 28(2), 409–433.

Perniss, P. (2018). Why we should study multimodal language. *Frontiers in psychology*, 9, 1109.

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: evidence from spoken and signed languages. *Frontiers in psychology*, 1, 227.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency. In J. Bybee & P. Hopper (Eds.), *Frequency and the emergence of linguistic structure* (pp. 10–1075). John Benjamins.

Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of pragmatics*, 12(5-6), 601–638.

Pons Bordería, S. & Estellés Arguedas, M. (2009). Expressing digression linguistically: Do digressive markers exist? *Journal of Pragmatics*, 41(5), 921–936. Pragmatic Markers.

Prince, E. F. (1981). Towards a taxonomy of given-new information. *Radical pragmatics*, (pp. 223–255).

Rauscher, F. H., Krauss, R. M., & Chen, Y. (1996). Gesture, speech, and lexical access: The role of lexical movements in speech production. *Psychological science*, 7(4), 226–231.

Reagan, T. (2011). Ideological barriers to American Sign Language: Unpacking linguistic resistance. *Sign Language Studies*, 11(4), 606–636.

Reddy, M. (1979). The conduit metaphor. *Metaphor and thought*, 2, 285–324.

Redeker, G. (2006). Discourse markers as attentional cues at discourse transitions. In *Approaches to discourse particles* (pp. 339–358). Brill.

Riester, A. (2019). Constructing QUD trees. In *Questions in Discourse* (pp. 164–193). Brill.

Roberts, C. (1996). Information structure in discourse. In J. Yoon & A. Kathol (Eds.), *OSU Working Papers in Linguistics*, volume 49 (pp. 91–136). Ohio State University.

Roberts, C. (2012). Information structure: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics*, 5, 1–69.

Rohde, H., Johnson, A., Schneider, N., & Webber, B. (2018). Discourse coherence: Concurrent explicit and implicit relations. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 2257–2267).

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and Categorization* (pp. 27–48). Lawrence Erlbaum.

Rosch, E. (1983). Prototype classification and logical classification: The two systems. *New trends in conceptual representation: Challenges to Piaget's theory*, (pp. 73–86).

Rowbotham, S., Holler, J., Lloyd, D., & Wearden, A. (2012). How do we communicate about pain? a systematic analysis of the semantic contribution of co-speech gestures in pain-focused conversations. *Journal of nonverbal behavior*, 36(1), 1–21.

Ruhl, C. (1989). *On monosemy: A study in linguistic semantics*. State University of New York Press.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1978). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7–55). Elsevier.

Sanders, T. (2005). Coherence, causality and cognitive complexity in discourse. In *Proceedings/Actes SEM-05, First International Symposium on the exploration and modelling of meaning* (pp. 105–114).: University of Toulouse-le-Mirail Toulouse.

Sanders, T. & Spooren, W. (2009). The cognition of discourse coherence. In J. Renkema (Ed.), *Discourse, of course* (pp. 197–212). Amsterdam: John Benjamins.

Scha, R. & Polanyi, L. (1988). An augmented context free grammar for discourse. In *Coling Budapest 1988 Volume 2: International Conference on Computational Linguistics* (pp. 573–577).

Schegloff, E. (1984). On some gesture's relation to talk. In J. M. Atkinson & J. Heritage (Eds.), *Structures of social action: Studies in conversation analysis* (pp. 266–296). Cambridge University Press.

Schegloff, E. (1987). Analyzing single episodes of interaction: An exercise in conversation analysis. *Social psychology quarterly*, (pp. 101–114).

Schegloff, E. & Sacks, H. (1973). Opening up closings. *Semiotica*, 8(4), 289–327.

Schiffrin, D. (1986). Functions of and in discourse. *Journal of Pragmatics*, 10(1), 41–66.

Schiffrin, D. (1987). *Discourse markers*. Cambridge University Press.

Schlenker, P. (2020). Gestural grammar. *Natural Language & Linguistic Theory*, 38(3), 887–936.

Scholman, M. C., Rohde, H., & Demberg, V. (2017). "On the one hand" as a cue to anticipate upcoming discourse structure. *Journal of Memory and Language*, 97, 47–60.

Schoonjans, S. (2017). Multimodal construction grammar issues are construction grammar issues. *Linguistics Vanguard*, 3(s1).

Schourup, L. (1999). Discourse markers. *Lingua*, 107(3-4), 227–265.

Shattuck-Hufnagel, S. & Prieto, P. (2019). Dimensionalizing co-speech gestures. In *Proceedings of the International Congress of Phonetic Sciences*, volume 5 (pp. 1490–1494).

Simons, M., Tonhauser, J., Beaver, D., & Roberts, C. (2010). What projects and why. In N. Li & D. Lutz (Eds.), *Semantics and linguistic theory*, volume 20 (pp. 309–327).

Singer, M. A. & Goldin-Meadow, S. (2005). Children learn when their teacher's gestures and speech differ. *Psychological Science*, 16(2), 85–89.

Sorace, A. & Keller, F. (2005). Gradience in linguistic data. *Lingua*, 115(11), 1497–1524.

Sperber, D. & Wilson, D. (1998). The mapping between the mental and the public lexicon. In P. Carruthers & J. Boucher (Eds.), *Language and thought: Interdisciplinary themes* (pp. 184–200). Cambridge: Cambridge University Press.

Stalnaker, R. C. (1978). Assertion. In P. Cole (Ed.), *Syntax and Semantics 9: Pragmatics* (pp. 315–332). New York: Academic Press.

Stec, K. (2012). Meaningful shifts: A review of viewpoint markers in co-speech gesture and sign language. *Gesture*, 12(3), 327–360.

Steen, F. & Turner, M. B. (2013). : (pp. 1–20). Stanford, CA: CSLI Publications.

Stickles, E. (2016). *The interaction of syntax and metaphor in gesture: A corpus-experimental approach*. PhD thesis, University of California, Berkeley.

Stivers, T. (2008). Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on language and social interaction*, 41(1), 31–57.

Stokoe, W. (1960). *Sign Language Structure: An outline of the visual communications systems of American deaf*, volume 8. Linstok Press.

Stokoe, W. C., Casterline, D. C., & Croneberg, C. G. (1976). *A dictionary of American Sign Language on linguistic principles*. Linstok Press.

Streeck, J. (2008). Gesture in political communication: A case study of the democratic presidential candidates during the 2004 primary campaign. *Research on Language and Social Interaction*, 41(2), 154–186.

Streeck, J. (2009). *Gesturecraft: The manu-facture of meaning*, volume 2. John Benjamins Publishing.

Streeck, J., Goodwin, C., & LeBaron, C., Eds. (2011). *Embodied interaction: Language and body in the material world*. Cambridge University Press.

Streeck, J. & Hartge, U. (1992). Previews: Gestures at the transition place. *The contextualization of language*, (pp. 135–157).

Sweetser, E. (1990). *From etymology to pragmatics: Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Sweetser, E. (1992). English metaphors for language: Motivations, conventions, and creativity. *Poetics Today*, (pp. 705–724).

Sweetser, E. (2007). Looking at space to study mental spaces co-speech gesture as a crucial data source in cognitive linguistics. *Methods in cognitive linguistics*, 18, 201–224.

Sweetser, E. (2022). Gestural meaning is in the body(-space) as much as in the hands. In T. Janzen & B. Scheffer (Eds.), *Signed language and gesture research in cognitive linguistics* (pp. 357–366). Boston/Berlin: De Gruyter Mouton.

Sweetser, E. & Sizemore, M. (2008). Personal and interpersonal gesture spaces: Functional contrasts in language and gesture. *Language in the context of use: Discourse and cognitive approaches to language*, (pp. 25–51).

Sweetser, E. & Stec, K. (2016). Maintaining multiple viewpoints with gaze. In B. Dancygier, W.-l. Lu, & A. Verhagen (Eds.), *Viewpoint and the fabric of meaning: Form and use of viewpoint tools across languages and modalities* (pp. 237–258). Berlin: De Gruyter Mouton.

Taboada, M. (2006). Discourse markers as signals (or not) of rhetorical relations. *Journal of pragmatics*, 38(4), 567–592.

Taboada, M. (2009). Implicit and explicit coherence relations. In J. Renkema (Ed.), *Discourse, of course* (pp. 125–138). John Benjamins.

Taboada, M. & Mann, W. C. (2006). Rhetorical structure theory: Looking back and moving ahead. *Discourse studies*, 8(3), 423–459.

Tagliamonte, S. A. (2011). *Variationist sociolinguistics: Change, observation, interpretation*. John Wiley & Sons.

Takahara, P. O. (1998). Pragmatic functions of the english discourse marker anyway and its corresponding contrastive japanese discourse markers. *Pragmatics and beyond new series*, (pp. 327–352).

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive science*, 12(1), 49–100.

Tao, H. (2003). Turn initiators in spoken English: A corpus-based approach to interaction and grammar. In P. Leistyna & C. Meyer (Eds.), *Corpus Analysis: Language Structure and Language Use* (pp. 187–207). Brill.

Taub, S. F. (2001). *Language from the body: Iconicity and metaphor in American Sign Language*. Cambridge University Press.

Teßendorf, S. (2014). Pragmatic and metaphoric–combining functional with cognitive approaches in the analysis of the "brushing aside gesture". In C. Müller, A. Cienki, E. Fricke, S. Ladewig, D. McNeill, & S. Tessendorf (Eds.), *Body–language–communication: An international handbook on multimodality in human interaction*, volume 2 (pp. 1540–1557). Berlin: De Gruyter Mouton.

Thibodeau, P. H., Hendricks, R. K., & Boroditsky, L. (2017). How linguistic metaphor scaffolds reasoning. *Trends in cognitive sciences*, 21(11), 852–863.

Thompson, G. (2013). *Introducing functional grammar*. Routledge.

Traugott, E. (1985). On regularity in semantic change. *Journal of Literary Semantics*, 14(3), 155–173.

Traugott, E. (2020). The development of "digressive" discourse-topic shift markers in english. *Journal of Pragmatics*, 156, 121–135.

Traugott, E. (2022). *Discourse Structuring Markers in English: A historical constructionalist perspective on pragmatics*. John Benjamins.

Trudgill, P. (2001). *Sociolinguistic variation and change*. Edinburgh University Press.

Trujillo, J. P. & Holler, J. (2021). The kinematics of social action: Visual signals provide cues for what interlocutors do in conversation. *Brain Sciences*, 11(8), 996.

Tuggy, D. (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, 4(3), 273–290.

Turk, M. J. (2004). Using and in conversational interaction. *Research on Language and Social Interaction*, 37(2), 219–250.

Urgelles-Coll, M. (2010). *The syntax and semantics of discourse markers*. A&C Black.

van Kuppevelt, J. (1995). Discourse structure, topicality and questioning. *Journal of linguistics*, (pp. 109–147).

Violi, P. (2008). Beyond the body: Towards a full embodied semiosis. In T. Ziemke, J. Zlatev, & R. Frank (Eds.), *Body, language and mind*, volume 2 (pp. 53–76). Berlin: De Gruyter Mouton.

Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech communication*, 57, 209–232.

Wang, H. (2016). The (x) thing is: From a matrix clause to a discourse marker. *Poznan Studies in Contemporary Linguistics*, 52(3), 555–577.

Waxer, P. (1977). Nonverbal cues for anxiety: an examination of emotional leakage. *Journal of abnormal psychology*, 86(3), 306–314.

Wehling, E. (2017). Discourse management gestures. *Gesture*, 16(2), 245–276.

Wierzbicka, A. (1972). *Semantic primitives*. Frankfurt: Athenäum.

Wilcox, S. & Occhino, C. (2017). Signed languages. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics* (pp. 99–117). Cambridge University Press.

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). Elan: A professional framework for multimodality research. In *5th international conference on language resources and evaluation (LREC 2006)* (pp. 1556–1559).

Zima, E. (2017). On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard*, 3(s1).

Zima, E. & Bergs, A. (2017). Multimodality and construction grammar. *Linguistics Vanguard*, 3(s1).

Zwicky, A. M. (1985). Clitics and particles. *Language*, 61, 283–305.