Language in the body:
Multimodality in grammar and discourse

by

Jennifer Hinnell

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Linguistics
University of Alberta

Examining Committee

Dr. Sally Rice, Supervisor
Dr. Elena Nicoladis, Supervisory Committee
Dr. Juhani Järvikivi, Supervisory Committee
Dr. Xiaoting Li, Examiner
Dr. Mark Turner, External Examiner

ABSTRACT

In this dissertation, I investigate the role of the body as a critical part of linguistic meaning-making, taking a cognitive and usage-based approach to language. These approaches prioritize the investigation of the linguistic conventions of everyday interactive contexts, namely, of spontaneous conversation, since they posit the importance of spoken language, rather than speaker's intuition or written text, as primary data. As Enfield (Enfield 2017: 3) puts it, conversation "is where language lives and breathes." Placing face-to-face conversation at the centre of linguistic study requires a consideration of the multiple modalities involved in language in interaction. In addition to linguistic features in the utterance, these include movements of the body such as manual gestures, head movements, shoulder shrugs, postural shifts, eye-gaze and brow movements, known collectively as co-speech behaviour.

To examine the contribution of the body to linguistic meaning, I investigate language use in interaction across three broadly construed linguistic domains: ASPECT, CONTRAST, and DISCOURSE NAVIGATION. I use the Red Hen archive, an international multimedia database of broadcast media featuring over 400,000 hours of video and 5-billion words of time-aligned transcripts of largely North American English, to observe linguistic and co-speech behaviour in hundreds of spontaneous conversations by a wide range of speakers over these three linguistic domains. I search for a range of linguistic expressions (as text strings in Red Hen) that characterize each domain and describe the embodied structures that accompany them. Using a combination of established (Bressem 2013) and novel annotation methodologies, I examine the manual gestures and movements of the head, shoulders, face, and eyes, and apply quantitative and statistical methods of data analysis.

In the first case study (Chapter 3), I explore the multimodal expression of event structure expressed through ASPECT-marking constructions. In the second set of studies (Chapter 4), I examine the behaviours associated with the marking of CONTRAST in speech. Finally, in the third set of case studies (Chapter 5), I investigate the co-speech behaviours aligned with linguistic expressions that help speakers signal the way they move through a conversation, i.e. expressions of DISCOURSE NAVIGATION which involve stance-taking at levels well beyond the simple sentence. The findings strongly suggest that the embodied representation of these domains is conventionalized and, furthermore, reveal how different articulators (e.g. gesture, head, shoulder, and torso movement) are recruited uniquely in conventionalized ways in each of these domains. For instance, stance is strongly associated with upper body movement as well as the use of manual gesture.

This dissertation marshals evidence for the coordinated and recurrent bodily enactment of grammatical and discourse-level expressions. Thus, its aim is to contribute to a more robust understanding of the role of the body in the specific domains addressed here and more broadly in natural discourse. This focus on face-to-face interaction as a starting point for linguistic description and language documentation has important implications for the study of languages that rely predominantly on verbal and visual signals (e.g. signed and oral Indigenous languages), in addition to contributing to developments in multimedia technologies, e.g. in virtual agents that rely on human-like language use and animated dialogue in films and video games.

PREFACE

This thesis is the original work of Jennifer Hinnell.

Chapter 3 of this dissertation has been published as:

Hinnell, Jennifer. 2018. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics* 29(4): 773-806.

Parts of Chapter 4 appeared in:

Hinnell, Jennifer. 2019. The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics* (35: 1-2): 55-92.

The poem on page (ii) is from William Diconsons's forward to Bulwer's *Chironomia*, 1644, accessed in November 2019 from:

https://quod.lib.umich.edu/e/eebo/A30105.0001.001/1:5?rgn=div1;view=fulltext

At the first sight we learne to read; and then
By Natures rules to perce and construe Men:
So commenting upon their Gesture, finde
In them the truest copie of the Minde.
The Tongue and Heart th"intention oft divide:
The Hand and Meaning ever are ally'de.

William Diconson, 1644

ACKNOWLEDGEMENTS

I've had two great privileges since moving to Edmonton to start my PhD. One has been to work under Dr. Sally Rice. Sally, for your generous investment of time, energy, and thought, and for sharing with me your vast knowledge and enthusiasm for language and how it is deployed, I am so grateful. My PhD has been the most enriching and satisfying endeavour of my working life, thanks to you. In my non-academic life, I had the privilege of meeting and marrying Keith Warner. Keith, you have brought me so much joy. I am forever yours and forever grateful. This is for you.

I'm grateful to my supervisory committee, Elena Nicoladis and Juhani Järvikivi, for your support, feedback, and prodding in different ways over the last number of years. I am grateful as well to Xiaoting Li and Mark Turner, who completed my examination committee as external examiners. Thanks to you all for your exacting questions and making me think about questions of language from six to 60,000 feet.

I feel very fortunate to have created an academic family over the course of this work. I am grateful for my research partnership with Irene Mittelberg and the Natural Media Lab at the RWTH Aachen University, Germany. Irene, it has been such a blessing to have a research partnership that features so much fun and friendship alongside theory-testing with motion capture. My studies were also greatly enriched by several visits to the European-University Viadrina, Germany, and by thought-provoking and encouraging exchanges with Cornelia Müller. Cornelia, many thanks to you and your team for the enthusiastic welcome to the field. Similarly, I thank Geert Brône and colleagues at KU Leuven, Belgium, for many meaningful discussions about quantifying the multimodal signal in interaction. I am indebted to friends and colleagues in the Berkeley Gesture & Multimodality Group, who listened to versions of each of these studies as they developed and gave thoughtful comments and much encouragement along the way. Thanks especially to Eve Sweetser, for your interest in my work and for creating such a warm and collegial environment in the BGMG.

Closer to home, I extend heartfelt thanks to Barbara Dancygier and Maite Taboada, my Vancouver mentors whom I had the great fortune to meet very early on in my graduate school journey. Without your care and guidance, and your engaged response to my ideas during my MA degree, I may not have pursued this PhD. I am so glad that we still find ways to be in touch and work together. Finally, while various maternity leaves interrupted occasions to meet in recent years, the early years of my PhD would not have been the same without the SPLAT reading group, with Sally Rice, Elena Nicoladis, Xiaoting Li, and fellow gesture graduate student Lisa Smithson. To other graduate students at various stages along the way, I am grateful to have shared the road with you, especially Kristina Geeraert, Claudia Heinrich, Dana Rahim, Audra Phillips, and Brian Rusk and Rachel McGraw for a very productive writing group in the end stages.

To my family, both the bigs and the littles: Thank you! To the adults, for never doubting that I needed to do this and would succeed. To all the kids, for always keeping life honest and full of love and shenanigans. And to my own littles in particular, Charlotte and Mackenzie, for embodying joy and for inspiring me daily. Finally, to two of the most excellent women, Sarah and Erika. You've had my back for a long time. Love to you always.

Lastly, I am grateful for financial support from the Social Sciences and Humanities Research Council of Canada and the Killam Trusts.

Thank you all!

Jennifer

TABLE OF CONTENTS

ABSTRACT ii
PREFACE iv
ACKNOWLEDGEMENTS vi
TABLE OF CONTENTS viii
LIST OF TABLES xii
LIST OF FIGURES xiv
DICTIONARY OF ABBREVIATIONS xvii

Chapter 1 | A multimodal view of language 1

1.1 Introduction 1
1.2 Cognitive linguistics 8
1.2.1 An overview 8
1.2.2 Embodied cognition 9
1.2.3 Constructions and construction grammars 15
1.2.4 Usage-based and corpus linguistic methods 18
1.2.5 From interaction to multimodality 19
1.2.6 Summary: Cognitive Linguistics 21
1.3 Multimodality 21
1.3.1 What is gesture? 22
1.3.2 Typologies of gesture 23
1.3.3 Recurrent gestures and gesture families 27
1.3.4 Multimodality: Beyond the hands 32
1.3.5 Multimodal constructions 37
1.3.6 Summary: Multimodality 39
1.4 Conclusion 40
1.4.1 Structure of the dissertation 40
1.4.2 Closing comments 43

Chapter 2 | Methods of multimodal corpus analysis 46

2.1 Introduction 46
2.2 Corpora 47
2.2.1 Red Hen 48
2.2.2 Corpus of Contemporary American English 51
2.3 Data collection 52
2.3.1 Target utterance searches 52
2.3.2 Search returns 54
2.3.3 Downloading clips from Red Hen 57
2.4 Annotations 57

2.4.1 Degree of enactment ... 58
2.4.2 Kinesic form annotations ... 61
Gesture annotations ... 61
Upper body annotations ... 72
Partitioning ... 73
Multimodal profiles ... 75
2.4.3 Linguistic form annotations ... 76
2.5 Citing multimodal examples ... 79
2.6 Summary ... 80

Chapter 3 | *Keep on talking* — and gesturing: The multimodal marking of ASPECT ... 81

3.1 Introduction ... 81
3.1.1 Aspect ... 83
3.1.2 Iconicity in gesture ... 88
3.1.3 Studies of aspect in gesture ... 90
3.2 Methods ... 93
3.2.1 Target utterances ... 93
3.2.2 Data collection ... 94
3.2.3 Data annotation ... 95
3.3 Findings99
3.3.1 Semantic overview by auxiliary ... 99
Open aspect auxiliaries ... 99
Phase aspect auxiliaries ... 104
3.3.2 Quantitative findings ... 108
Open aspect auxiliaries ... 108
Phase aspect auxiliaries ... 110
3.4 A statistical comparison across auxiliaries ... 113
3.5 Discussion ... 115
3.6 Conclusion ... 119

Chapter 4 | *On the one hand, on the other hand:* CONTRAST in the body ... 122

4.1 Introduction ... 122
4.1.1 Contrast background ... 125
4.1.2 Expressions and domains of contrast ... 128
4.1.3 Embodied contrast ... 129
4.1.4 Conventionalization in the embodied signal ... 134
4.1.5 Summary ... 135
4.2 Methods ... 136
4.2.1 Target utterances ... 137
4.2.2 Data collection in Red Hen ... 139
4.2.3 Data annotation ... 143
Kinesic form annotations ... 143

Linguistic form annotations ... 146
4.2.4 Corpus comparison for on (the) one hand/on the other hand ... 147
4.2.5 Summary: Methods ... 148
4.3 Real world CONTRAST ... 149
4.3.1 On (the) one hand/on the other hand ... 150
4.3.2 Other phrasal and lexical expressions of CONTRAST ... 158
4.3.3 Quantitative case studies ... 165
On (the) one hand/on the other hand ... 165
Better than/worse than ... 177
4.3.4 Summary: Real world CONTRAST ... 180
4.4 Possible world CONTRAST ... 181
4.4.1 Should I or shouldn't I? ... 182
4.4.2 Quantitative study: Should I/shouldn't I ... 186
4.4.3 Summary: Possible world CONTRAST ... 189
4.5 Conclusion ... 190

Chapter 5 | *That being said*: Discourse constructions are multimodal ... 193
5.1 Introduction ... 193
5.1.1 Background ... 196
5.1.2 Stance and constructional specificity ... 198
5.1.3 Summary ... 203
5.2 Methods and data overview ... 204
5.2.1 Target utterances ... 205
5.2.2 Data annotation ... 213
5.3 Excursion markers ... 215
5.3.1 Excursion microstudies ... 218
Which is true ... 220
Which is fine ... 224
Which, by the way ... 228
How shall/should I put it? ... 230
Although, you know ... 233
5.3.2 Summary: Excursions ... 236
5.4 Return markers ... 238
5.4.1 Return microstudies ... 242
Anyhow ... 244
At any rate ... 247
So anyway(s) and *but anyway(s)* ... 249
However, I ... 259
Nevertheless ... 264
That being said ... 267
5.4.2 Summary and discussion: Returns ... 270
5.5 Excursions and returns as stance-marking, multimodal constructions ... 274
5.5.1 The nature of stance: Embodied, stacked, and idiomatic ... 274

5.5.2 Concluding remarks 279

Chapter 6 | Discussion and conclusion 282

6.1 Introduction 282
6.2 Implications for a constructionalist view of language 287
6.2.1 Convention and variation 287
6.2.2 Schematicity, iconicity, and grammaticalization in the body signal 290
6.3 Applications in linguistics and beyond 294
6.3.1 Language documentation and language pedagogy 295
6.3.2 Embodied conversational agents 297
6.4 Final observations 302

REFERENCES 303

APPENDIX A: Video clips and corpus metadata 316

A1. Instructions for viewing videos clips 316
A2. Metadata for Red Hen examples 316

APPENDIX B: Red Hen program selections and settings 317

B1. Red Hen network and series selections 317
B2. Sample Red Hen search interface 318

LIST OF TABLES

Table 2.1. Manual gesture annotation parameters by domain 63
Table 2.2. Upper body annotation by domain 73
Table 2.3. Linguistic annotations by domain 78
Table 3.1. Characteristics of types of aspect and their target auxiliaries 88
Table 3.2. Gesture annotation schema 96
Table 3.3. Predicted relationships between gesture variables and aspect type 99
Table 3.4. Summary of annotation results 108
Table 3.5. [CONTINUE to VVI] – Summary table (n=37) 109
Table 3.6. [KEEP VVG] – Summary table (n=29) 110
Table 3.7. [START VVG] – Summary table (n=25) 111
Table 3.8. [STOP VVG] – Summary table (n=27) 112
Table 3.9. [QUIT VVG] – Summary table (n=29) 112
Table 3.10. Interactions by aspect type 113
Table 3.11. Relative frequency distribution of movement type by aspect 114
Table 3.12. Movement type by mean action phases and gesture timing 115
Table 4.1. Distribution by construction type 142
Table 4.2. Variables and levels of kinesic movement annotation 145
Table 4.3. Annotation schema for target utterance context 146
Table 4.4. Frequency of [O1H] and [OOH] in Red Hen and $COCA_{sp}$ 148
Table 4.5. Degree of enactment for [O1H] and [OOH] 167
Table 4.6. Palm orientation for [O1H] in [O1H-exp] constructions 169
Table 4.7. Palm orientation for [OOH] in [exp-OOH] constructions 170
Table 4.8. Frequency of upper body movement for [O1H] and [OOH] 172
Table 4.9. Body partitioning for [O1H-OOH] by construction type 173
Table 4.10. Palm orientation, [BT-WT] 179
Table 4.11. Upper body movement for [BT-WT] 179
Table 4.12. Body partitioning for *should I/shouldn't I* 187
Table 4.13. Upper body movement for *should I/shouldn't I* 187
Table 5.1. Top 20 collocates for [which is ADJ] in $COCA_{sp}$ 206
Table 5.2. Frequency of excursion expressions 207
Table 5.3. Return utterance case studies 213
Table 5.4. Kinesic annotation variables 214
Table 5.5. Degree of enactment for excursion expressions in Red Hen 217
Table 5.6. Degree of enactment by excursion marker 219
Table 5.7. Distribution of articulator forms: *which is true* 221
Table 5.8. Distribution of articulator forms: *which is fine* 225
Table 5.9. Distribution of articulator forms: *which, by the way* 228
Table 5.10. Distribution of articulator forms: *how shall I put it* 231
Table 5.11. Distribution of articulator forms: *although, you know* 234
Table 5.12. Dominant gesture form by excursion expression 238
Table 5.13. Dominant head movement by excursion expression 238
Table 5.14. Degree of enactment by return marker 243
Table 5.15. Distribution of articulator forms: *anyhow* 247
Table 5.16. Distribution of articulator forms: *at any rate* 249

Table 5.17. Distribution of articulator forms: *so anyway(s)*..251
Table 5.18. Distribution of articulator forms: *but anyway(s)*..255
Table 5.19. Pronoun frequency in [however, PRO] in $COCA_{sp}$....................................260
Table 5.20. Five most common verb collocates for [however , PRO]..........................261
Table 5.21. Collocating stanced expressions with *however, I*...261
Table 5.22. Distribution of articulator forms: *however, I*..263
Table 5.23. Distribution of articulator forms: *nevertheless*..265
Table 5.24. Distribution of articulator forms: *that being said*...269
Table 5.25. Dominant gesture form by return expression 272
Table 5.26. Dominant head movement by return expression..273

LIST OF FIGURES

Figure 2.1 Screenshot of Red Hen search interface 53
Figure 2.2. Screenshot of Red Hen search return page..54
Figure 2.3. Data point excluded due to chyron...56
Figure 2.4. Example of pre-existing gesture handshape 59
Figure 2.5. Model of gesture space (McNeill 1992: 89) 64
Figure 2.6. Gesture movement annotation: Axis and direction (Bressem 2013: 1090) 65
Figure 2.7. Gesture movement annotation: Movement type (Bressem 2013: 1088).....65
Figure 2.8. Spiral movement trace as captured by MoCap...67
Figure 2.9. Arc movement trace as captured by MoCap 67
Figure 2.10. One-handed and two-handed PUOH gesture forms..................................69
Figure 2.11. Recurrent gesture forms 70
Figure 2.12. *on the one hand/on the other hand*, PUOH gesture 79
Figure 3.1. *continued to grow,* segmented gesture on vertical axis 97
Figure 3.2. *continue to grow*, singular continuous gesture on horizontal axis..................101
Figure 3.3. *continues to grow,* cyclic gesture 101
Figure 3.4. *keep coming back*, iterated cyclic gesture..103
Figure 3.5. *keep getting paid,* iterated downward stroke...103
Figure 3.6. *keeps getting higher*, spiral movement on vertical axis 104
Figure 3.7. *start making blessings,* single phase cyclic gesture 105
Figure 3.8. *I'm gonna quit smoking*, bimanual, horizontal, outward gesture stroke........107
Figure 3.9. *stopped talking,* bimanual, horizontal, outward gesture stroke 107
Figure 3.10. *quit talking*, iterated cyclic gesture 117
Figure 4.1. Bulwer's rhetorical gesture for *antitheses* (Bulwer 1644)..............................126
Figure 4.2. Lateral gestures marking CONTRAST: *insanely defensive of him* vs. *vilifies him* .127
Figure 4.3. Line drawing of *on the one hand/on the other hand*..131
Figure 4.4. Distribution of construction types in Red Hen and $\mathrm{COCA}_{\mathrm{sp}}$ 148
Figure 4.5. *on the one hand/on the other hand*, PUOH gesture 150
Figure 4.6. *on the one hand/on the other hand*, PUOH gesture 151
Figure 4.7. *on one hand*, container gesture...153
Figure 4.8. *on the one hand,* container gesture 154
Figure 4.9. *on the one hand/on the other hand,* bilateral point sequence 155
Figure 4.10. *on one hand,* bimanual palm-lateral gesture 156
Figure 4.11. *on the other hand,* PUOH gesture 157
Figure 4.12. Bipartite expression, container and PUOH gestures................................158
Figure 4.13. *something I do* vs. *people do it way worse,* bilateral gestures..............................159
Figure 4.14. *defense* vs. *offense,* antonymic contrast with bilateral gestures 160
Figure 4.15. antonymic noun phrases with *versus,* container gesture.............................161
Figure 4.16. *a good news bad news situation,* bilateral indexical gestures 161
Figure 4.17. *manicure or pedicure,* PUOH gesture...162
Figure 4.18. *insanely defensive of him* vs. *vilifies him*, bilateral gesture.................................162
Figure 4.19. *something that is* vs. *something that isn't,* PLTC gesture...................................163
Figure 4.20. *better than/worse than*, PUOH gestures 164
Figure 4.21. *better than/worse than*, head tilt 164
Figure 4.22. Symmetry and laterality by construction type...168

Figure 4.23. *on one hand,* palm-down gesture 171
Figure 4.24. *on the other hand,* palm-down gesture 171
Figure 4.25. Body partitioning profile by construction type 173
Figure 4.26. Body partitioning profile for *better than/worse than* 180
Figure 4.27. Contrastive enactments in irrealis marking 185
Figure 4.28. *should I/shouldn't I,* subtle head tilt 188
Figure 4.29. *should I/shouldn't I,* pronounced head tilt with lateral gestures 188
Figure 4.30. *should I/shouldn't I,* pronounced head tilt 189
Figure 4.31. *if…/if…*, head tilt 190
Figure 5.1. *what can I say?,* PUOH gesture 202
Figure 5.2. *what can I do?,* representative enactments 203
Figure 5.3. Relative frequency of selected return expressions in COCA_{sp} 212
Figure 5.4. Genre preference for return markers in COCA 213
Figure 5.5. Body partitioning for excursion markers 219
Figure 5.6. *which is true*, PUOH and head nod 222
Figure 5.7. *which is true*, palm lateral gestures to each side 223
Figure 5.8. *which is true*, head turn 223
Figure 5.9. *which is fine,* three examples with holding away gesture 226
Figure 5.10. *which is fine,* palm-down gesture 226
Figure 5.11. *which is fine,* PUOH gesture 227
Figure 5.12. *which, by the way,* holding away gesture 229
Figure 5.13. *which, by the way,* point gesture and head movement 229
Figure 5.14. *how should I put it,* self-touch gesture and head turn 231
Figure 5.15. *how shall I put it,* head movement only 232
Figure 5.16. *although, you know,* PUOH and head tilt 234
Figure 5.17. *although, you know,* equivocation gesture 235
Figure 5.18. Distribution by articulator for all excursion expressions 237
Figure 5.19. *anyhow*, head tilt and mouth shrug 240
Figure 5.20. *at any rate*, holding away gesture, brow raise, ... 241
Figure 5.21. *anyway*, shoulder shrug 242
Figure 5.22. Body partitioning profile for return markers 244
Figure 5.23. Distribution by articulator: *anyhow* 246
Figure 5.24. Distribution by articulator: *at any rate* 248
Figure 5.25. Distribution by articulator: *so anyway(s)* 250
Figure 5.26. *so anyway(s),* PUOH 251
Figure 5.27. *so anyway(s)*, PLTC and head tilt 252
Figure 5.28. Distribution by articulator: *but anyway(s)* 254
Figure 5.29. *but anyway,* palm-down open-hand 255
Figure 5.30. *but anyway,* one-handed holding away gesture (palm-vertical) 256
Figure 5.31. *but anyway,* bimanual vertical palm-away and palm-down gestures 257
Figure 5.32. *but anyway*, outward-moving PDOH 257
Figure 5.33. Distribution by articulator: *however, I* 262
Figure 5.34. *however, I,* pointing gesture 263
Figure 5.35. *however, I,* pursed hand 263
Figure 5.36. Distribution by articulator: *nevertheless* 265
Figure 5.37. *nevertheless,* range of deictic gesture forms 266
Figure 5.38. Distribution by articulator: *that being said* 269
Figure 5.39. *that being said*, holding away gesture with brow raise and head turn 270

Figure 5.40. Distribution by articulator for all return expressions with *any*..................271
Figure 5.41. Distribution by articulator for [however, I], [nevertheless], and [TBS]...272
Figure 5.42. Degree of enactment for non-restrictive relative clauses 278
Figure 6.1. Screenshot of virtual reality training scenario for doctor-patient interaction in 'breaking bad news' disclosures 298
Figure 6.2. Cassandra in dialogue with Solas, *Dragon Age: Inquisition* 301
Figure 6.3. Leliana and Varric in dialogue, *Dragon Age: Inquisition* 301

DICTIONARY OF ABBREVIATIONS

CONSTRUCTIONS

Chapter 3: ASPECT

[CONTINUE to VVI]	*continue to VERB*
[KEEP VVG]	*keep VERB-ing*
[START VVG]	*start VERB-ing*
[STOP VVG]	*stop VERB-ing*
[QUIT VVG]	*quit VERB-ing*

Chapter 4: CONTRAST

[O1H-OOH]	*on (the) one hand/on the other hand*
[BT-WT]	*better than/worse than*
[SI~SI]	*should I/shouldn't I*

Chapter 5: DISCOURSE

Excursions

[WIT]	*which is true*
[WIF]	*which is fine*
[WBTW]	*which, by the way*
[HSIPI]	*how shall/should I put it*
[AYK]	*although, you know*

Returns

[anyhow]	*anyhow*
[AAR]	*at any rate*
[so anyway(s)]	*so anyway(s)*
[but anyway(s)]	*but anyway(s)*
[however, I]	*however, I*
[nevertheless]	*nevertheless*
[TBS]	*that being said*

Chapter 1 | A multimodal view of language

1.1 Introduction

Human beings express themselves through their bodies from the earliest ages. Think of a child at two years of age. She has likely started to say short utterances, but is hampered by a lack of vocabulary, a lack of syntax, and a lack of ability to accurately produce many of the sounds of her first language. What she can do is recruit additional modalities of communication to express herself. A head tilt, alongside rising intonation in the utterance of a phonetic sequence resembling the sounds for "daddy", suffices to communicate her desire to know where her daddy is. While playing outside, a child who is looking for a dog she had just seen might hold up her hands around waist height with palms facing upwards, raise her shoulders, and say /go/ as a proximal sound for the word *dog*. Given the context, the palm-up and shrug gesture could convey something akin to a more complex utterance such as *Where's the dog? (I don't know where it went)*. Gesture use is not merely the domain of children who have yet to acquire a full range of oral language skills. As an adult, think of a time when you have been able to see, but not hear, a friend with whom you wish to communicate. Perhaps it was across a crowded party room or an airport lounge? Gestures such as pointing first to the empty glass in your friend's hand and then to the location of the bar likely helped you accomplish the communicative goal of your bringing your friend a refill.

Beyond the use of pointing to achieve a specific goal, you likely will also have observed the regularity with which people's bodies move as they speak. These movements tend to include gestures formed by the hands, as well as head movements such as tilts, nods, and shakes, eyebrow and gaze shifts, and shoulder movements. The meaning of these movements may not be as easily decipherable as the pointing series that helped you to order a refill in the scenario described above. Without the help of a verbal utterance, you may not be able to discern the precise content of a conversation simply by observing accompanying body movements. One thing is clear: regardless of which language they are speaking or whether they can be heard or not, when people talk, they also communicate with their bodies. In this dissertation, I explore the ways in which communicative body movements contribute to linguistic meaning and consider how the interplay between speech and the body needs to be integrated into our understanding of language as a whole.

Much of linguistic enquiry over the last half century has been dedicated to the properties of the mind that make language possible. These properties include, for example, rules of word formation and grammar that could explain how language is processed. This line of enquiry frequently takes as its starting point written language and focuses on decontextualized words (and sometimes phrases), in addition to the syntactic structures that are said to account for how these words and phrases are learned and joined together in the first place. Formalist approaches to language have idealized speakers and their language use to such an extent that language structure has been studied as an isolated entity, impervious to influence by 'outside' factors, such as pragmatic intent, other speakers' behaviours, and conversational conventions such as turn-taking and repair, that are evident in everyday conversation. Historically, the field has had less to say about the language of seemingly mundane, everyday exchanges in face-to-face interactions between speakers. This lack of focus on dialogue has been called a blind spot for linguistics (Enfield 2019). It ignores conversation, which, as Enfield insists, "is where language lives and breathes" (2017: 3). Enfield goes on to argue that "the inner workings of conversations [should] have their rightful place at the center of the language sciences" (ibid.).

Enfield is not alone in arguing that conversation should be at the centre of investigations of language. The field of conversation analysis (CA) has placed speech and interaction at the focus point of language investigation by prioritizing the intonation unit as a unit of analysis and describing patterns in oral phenomena such as turn-taking and repair. Along with CA, interactional linguistics is also characterized by a focus on genuinely interactive, sociocognitively grounded accounts of meaning, which examine the coordination of meaning between different interlocutors (Feyaerts et al. 2017: 136); however, until recently, analyses in these fields have remained largely qualitative and have been restricted to spoken utterances that are examined in text form (via transcriptions). The effect is that these fields have overlooked the role of the body in linguistic meaning-making. Instead, as soon as one centralizes face-to-face interaction as the object of linguistic study, one must consider the various modalities involved in language, i.e. the body in addition to elements in the speech stream. Recognizing this, there has been a recent 'multimodal turn' in some fields, including cognitive linguistics (CL), that has put full-bodied, face-to-face interaction under the microscope in linguistic analyses.

The language-body relationship is complex. Manual gestures, shoulder shrugs, facial expressions and gaze movements, and postural or torso shifts, all add meaning to a spoken utterance. Indeed, in contexts such as the gestured "conversation" in the pub that I described earlier, these embodied behaviours can convey meaning in the absence of a linguistic utterance altogether. Conversely, gesture and other so-called and easily dismissed "paralinguistic" phenomena such as gaze, facial expression, and shoulder and torso movements are not obligatory for the production and comprehension of speech. Their optionality may be a contributing factor to their rejection historically as objects of linguistic study; however, more recent studies within the fields of cognitive science, psychology, and interactive and usage-based approaches in linguistics have shown that the workings of the human mind are intricately bound to the workings of the human body (Glenberg & Kaschak 2002; Gibbs 2005; Bergen & Wheeler 2010; Müller, Cienki, et al. 2013; Müller et al. 2014). The study of language in many research fields has moved beyond the study of text or speech in isolation to the examination of full-bodied multimodal interaction. There is new focus on how internalized patterns of embodied behaviour "might drive meaningful expression, and visibly manifest themselves in communicative bodily movements integrated with spoken discourse" (Mittelberg 2013b: 750). The research program presented in this dissertation answers, in part, the call by Enfield and others for the close observation and description of interactional behaviour in conversation to explore the multimodal nature of language in face-to-face interaction (Enfield 2017; Feyaerts et al. 2017; Zima & Bergs 2017a).

This dissertation takes a cognitive, constructionalist approach to language, which is guided by the premise that all linguistic structure is meaningful. That is, if we consider language to be a situationally grounded, embodied, and interactional medium, then linguistic structures must be "related to and motivated by human conceptual knowledge, bodily experience, and the communicative functions of discourse" (Gibbs 2005: 11). If this is the case, then central areas of inquiry must include the form and nature of co-speech behaviours as linguistic structures in their own right and the degree of conventionalization, or more cognitively speaking, entrenchment, that these behaviours manifest. The driving research questions I seek to address in this dissertation include: Are there co-speech behaviours that regularly align with specific linguistic and/or conceptual structures? If so, what form do these behaviours take? And lastly, to what degree are these forms conventionalized? I seek to provide empirical evidence to inform our understanding of the nature of multimodal constructions in grammar and discourse.

To investigate the contribution of the body to linguistic meaning, I examine language use in interaction across three broadly construed conceptual domains: ASPECT, CONTRAST, and DISCOURSE NAVIGATION. The range of case studies presented in this dissertation examine linguistic expressions that characterize each domain and describe the embodied structures that accompany them. In the first case study (Chapter 3), I examine the multimodal expression of event structure in a set of ASPECT-marking constructions. In the second set of studies (Chapter 4), I capture the behaviours associated with the marking of CONTRAST in speech. Finally, in the third set of case studies (Chapter 5), I investigate the co-speech behaviours aligned with linguistic expressions of speaker stance; expressions in the domain of DISCOURSE NAVIGATION are especially relevant here as they involve stance-taking at levels well beyond the simple sentence. As described in each chapter, the findings suggest that the embodied representation of these domains is conventionalized and, furthermore, reveal how different articulators (e.g. the hands, head, shoulder, eyebrow, and torso movement) are recruited uniquely in conventionalized ways in each of these domains.

Within the three conceptual domains of ASPECT, CONTRAST, and DISCOURSE NAVIGATION, I explore reflexes in the body alongside the verbalization of specific linguistic expressions. In North American English, speakers have available to them a very heterogeneous set of linguistic constructions with which they can express each domain. These range from fixed expressions (i.e. individual words and short phrases) to more schematic grammatical structures. For example, ASPECT can be expressed morphologically (as in the English *-ing* suffix to mark progressive aspect), periphrastically through lexically rooted auxiliary constructions (*keep saying*), through prepositions (*on and on)*, adverbials (*again*), and many other devices. CONTRAST can be marked through logical operators (*and*, *but,* and *or*), fixed phrases (*by contrast, in comparison*), and through a variety of phrasal constructions that vary from more to less fixed in nature (e.g. the fixed bipartite pairing *on (the) one hand/on the other hand* and less fixed phrases within a similar bipartite setup such as *in one way/it is also*). Finally, the realm of stance – an umbrella term that captures the expression of attitude and evaluation – is expressed through a wide range of expressions, from adverbials (*seriously, frankly)*, to verb phrases (*I think, I feel*), discourse markers (*like, well, now)*, and impersonal clauses (*it seems that),* to name only a few. More specifically, the set of discourse navigation markers I examine in Chapter 5 includes both highly stanced fixed expressions that serve as parenthetical asides (*which is fine, which is true),* and expressions that return a speaker to the main flow of discourse. This latter function is frequently marked through concessive devices such as *however, nevertheless,* and *anyways*.

Within cognitive and usage-based approaches to language, linguists have been divested of the notion of a separate lexicon and grammar in favour of a continuum of sub-lexical and multiword units (Bybee 2006). The range of expressions I have selected exemplifies a continuum of highly lexical to highly grammatical constructions that are recruited to express ASPECT, CONTRAST, and DISCOURSE NAVIGATION. The case studies presented throughout this dissertation shed light on whether there is an analogous, schematized organization of the *bodily* structures used to express these domains, i.e. whether there is a range of gestural and/or postural structures that convey (more or less) the same conceptual material as the linguistic expressions, analogous to the continuum of resources available in speech. Furthermore, the case studies investigate the differential role of the modes of expression available in the body (e.g. manual gesture vs. head movement vs. posture shifts). By examining the hands and behaviour of other articulators within the same study, as I do for the studies of CONTRAST and DISCOURSE NAVIGATION, I address whether the body enacts these domains differentially such that some articulator movements (e.g. head tilts, nods) or clusters (e.g. upper body movements more generally) conventionally express certain conceptual structures. In particular, I explore whether the upper body is recruited more frequently to express highly subjective material – i.e. stance-marking, while manual gesture is recruited to express more objective, imageable characteristics of objects and events.

The field of gesture studies has tended to focus almost exclusively on referential gestures, including iconic, metaphoric, and deictic gestures, all of which are characterized by a relatively straightforward form-to-meaning mapping in terms of form features, such as hand shape (McNeill 2005). Instead, I focus on more abstract domains in order to explore conventionalized mappings in co-speech forms. The corpus-driven and quantitative nature of the studies presented in this dissertation also represent a departure from most multimodality studies to date. While co-speech gesture research has been largely based on case studies of a small number of items, often involving conversation elicited in a lab setting, I use advances in digital data, particularly the increasing availability of video-based language corpora, to examine hundreds of spontaneous conversations by a wide range of speakers. By taking a corpus-based approach, I can control for linguistic variables. I also collect a large enough data set for each domain to conduct quantitative and/or statistical analyses and to compare utterances within domains. The aim of this dissertation is to contribute to a more robust understanding of the role of the body in the specific domains addressed here and more broadly in natural discourse.

In the remainder of this introductory chapter, I outline some basic tenets from the general philosophical framework of cognitive linguistics, including its take on embodied cognition, its adoption of corpus-linguistic methods to support its commitment to usage-based description, and its theoretical operationalization in grammatical approaches called construction grammars (§1.2). In §1.3, I provide an overview of the field of multimodality, including an introduction to gesture studies, conventionalized gesture forms, and the integration of co-speech behaviour into a theory of language, for example in a multimodal construction grammar.

1.2 Cognitive linguistics

1.2.1 An overview

Cognitive linguistics (CL) reflects a commitment to investigating and accounting for all aspects of human language in a way that is informed by and aligns with what we know about cognition and the human brain. Lakoff termed this the *cognitive commitment* (Lakoff 1990: 40). It is this principle that is the hallmark of what many refer to as second generation (or embodied) cognitive science. Since the 1980s, empirical studies of the syntax, semantics, and pragmatics of languages have provided evidence that language develops as a result of our cognitive capacities for perception, bodily motion, and action. Early studies of metaphor, body part projections, and spatial relations, revealed the crucial impact of our bodies and our interactions with forces in the world on meaning-making generally, and linguistic meaning-making more specifically (Lakoff & Johnson 1980, 1999; Talmy 2000a, 2000b). Thus, for today's cognitive linguists, language is seen as "being shaped by all aspects of our bodily being in the world –

from perception to movement to feeling" (Johnson 2017: 33). This view of language is part of the larger theoretical notion of embodiment, which refers to the ways our bodies and our interactions with the world shape our minds, our actions, and our personal and cultural identities (Gibbs 2005: 450).

In the following sections, I introduce the principal theoretical underpinnings of cognitive linguistics as they relate to the major contributions and themes of this dissertation. These include the theory of embodiment, the importance of usage in how we learn and use language, and, increasingly, the role of face-to-face interaction as the primary locus of language use. The view of language put forward here is one of a multimodal "constructicon" (Goldberg 2006; Steen & Turner 2013). This blend of *lexicon* and *construction* implies that language comprises a structured inventory of conventionalized forms (Langacker 1991b: 15), rather than a series of phrase structure rules as in formalist approaches. This review therefore also examines literature that explores the construction as the primary unit of linguistic analysis, discusses multimodality as a general characteristic of language, and introduces the notion of *multimodal constructions* as instantiations of the inherently multimodal yet highly conventionalized nature of language.

1.2.2 Embodied cognition

Recent advances in cognitive science have placed the human body at the centre of the exploration of cognition. In the mid-1990s, the embodied simulation hypothesis emerged. It posited that the understanding of language emerges from simulations in our minds of the physical experience that the language describes. Barsalou (2010: 618) described simulation as "the re-enactment of perceptual, motor, and introspective states acquired during experience with the world, body, and mind." The central role ascribed to the simulation of real-world experience represented a significant shift from the long-held view of cognition, language, and meaning-making as autonomous, modular processes within the brain. For "[v]iewing the brain simply as an information-processing or computational device, as the center of cognition, ignores the centrality of animate form in human thought" (Gibbs 2005: 9). Given the wide acceptance of the term *embodied cognition* in cognitive linguistic circles, in this dissertation I use the term in a general sense to capture the central role of the body (of both speaker and hearer) in our experience of the physical world. For, as Gibbs states:

> People's subjective, felt experiences of their bodies in action provide part of the fundamental grounding for language and thought. Cognition is what occurs when the body engages the physical, cultural world and must be studied in terms of the dynamical interactions between people and the environment. Human language and thought emerge from recurring patterns of embodied activity that constrain ongoing intelligent behaviour. We must not assume cognition to be purely internal, symbolic, computational, and disembodied, but seek out the gross and detailed ways that language and thought are inextricably shaped by embodied action.
>
> (Gibbs 2005: 9)

In an embodied cognition view of meaning, meaning "isn't something that is distilled away from our bodily experiences but is instead tightly bound by them" (Bergen 2012: 12). Original studies in embodiment were based on investigating metaphor in language and thought. Where previously metaphor had been construed as arising from literal similarities between objects, in one embodied approach to metaphor known as Conceptual Metaphor Theory (CMT), Lakoff and Johnson (1980, 1999) developed a theory of metaphor in which metaphors were understood to be *mappings* between two semantic and experiential domains known as *frames* (Fillmore 1982). In these mappings, aspects of the *source* domain are transferred to aspects of the *target* domain. The source domain is generally characterized by concrete forces that can be directly perceived or experienced. The target domain is the semantic domain containing the topic of the metaphor and is generally abstract in nature. For example, in the phrase *spend time*, TIME (the target domain) is compared to MONEY (the source domain). One of the main interests in empirical studies of metaphor has been the construal of TIME in terms of SPACE. The metaphoric construal of time as space is exemplified in (1). Here, the literal meaning of *at* is locative, but it has been extended to the domain of time (Croft & Cruse 2004: 194). Likewise, in (2), the preposition *in*, also normally a locative expression, provides for a metaphorical extension in which "a state (danger) is conceived as a container that one can be inside of or outside of" (ibid.: 196).

(1) *I'll see you at 2'oclock.*
(2) *He is in danger.*

(Croft & Cruse 2004: 194-196)

Where (1) exemplifies the TIME IS SPACE metaphor, (2) reflects the STATES ARE CONTAINERS metaphor (Lakoff & Johnson 1980: 31). We also transfer our understanding of and interaction with objects onto our understanding of ideas, as shown in the fact that in English we can *grasp a concept* and *toy with an idea.* In other words, we transfer our knowledge of the OBJECT frame to the IDEA frame (Stickles 2016).

While CMT was originally based in studies of linguistic expressions, research in co-speech gesture has provided a new lens into conceptual metaphors. The research addresses questions such as whether metaphorical mappings are universal. In short, they aren't. For example, while in English UNDERSTANDING is conceived of as SEEING, as in *I see your point*, in certain Australian Indigenous languages, UNDERSTANDING is conceived of as HEARING (Evans & Wilkins 2000; Gaby & Sweetser 2017). Additionally, studies of temporal gestures have shown that mappings of time to space vary across cultures as well (Núñez & Sweetser 2006; Cooperrider & Núñez 2009; Casasanto & Jasmin 2012).

Because of their basis in embodied experience, the source domains in conceptual metaphors have been described as image-schematic. The term *image schema* denotes an intrinsically meaningful and embodied pattern of recurrent experiences (Johnson 1987). Image schemas are generally related to visual perception, sensorimotor routines, and interactions with the physical and social world. For example, the fact that we are subject to the force of gravity generates recurring experiences of up and down, which are realized in the image schema UP/DOWN. This image schema impacts our conceptualizing of physical objects and values as rising or falling and of being on top or below spatially as well as socially. Similarly, our regular experience of balance gives rise to the BALANCE schema, which we then apply to physical objects and extend to balancing inner states, mathematical equations, and notions of political fairness and justice (Johnson 2017). Common image schemas include SOURCE-PATH-GOAL, CENTER-PERIPHERY, CYCLE, and many others (Cienki 1997; Hampe & Grady 2005; Cienki 2013).

Further support for an embodied view of language and cognition comes from studies of language processing, as well as studies of how meaning is processed in other modes of expression (e.g. gesture, film, music, and dance). As discussed above, the embodied simulation hypothesis proposes that in processing language, language users construct a mental experience of what it would be like to perceive or interact with the objects and events that are referred to in the instance of language use (Barsalou 1999, 2010; Bergen 2012). A groundswell of experimental studies have shown evidence for the critical role of simulation in language processing. Glenberg and Kaschak (2002) first reported on the *action-sentence compatibility effect* (ACE) that showed that sentence processing leads people to perform mental simulation. The original ACE study showed that when a sentence implied meaning in one direction, participants found it difficult to judge sentence meaning for an utterance that required an experimental response in the opposite direction. For example, the sentence *close the drawer* implies action away from the body and was shown to interfere with real physical action inwards toward the body in the manipulation of a joy stick device for responding to a stimulus sentence. They concluded that their data were "inconsistent with theories of language comprehension in which meaning is represented as a set of relations among nodes. Instead, the data support an embodied theory of meaning that relates the meaning of sentences to human action" (Glenberg & Kaschak 2002: 558). In essence, language users simulate in the moment of language processing a "what this must be like" process, which makes use of their embodied experience in the world (Gibbs 2017: 458).

Many studies have subsequently used the ACE methodology to further examine the effects of mental simulations on language comprehension. Richardson et al. (2003) examined the activation of spatial representations through directionally associated words (e.g. *push* for horizontal motion, *bomb* for vertical motion) and found location-specific interference of language on visual processing. This was argued to show that language-induced mental simulation ties up location-specific parts of the vision system. Follow-up studies suggested that this holds for actual motion (*The mule climbed),* but not for metaphorical motion (*The prices climbed)* (Bergen et al. 2007); however, other studies have shown that even abstract levels of language processing are grounded in body movement.

In addition to proposing that language processing is grounded in perception and cognition, studies have shown that speakers also express abstract conceptual and linguistic notions with their bodies. Such notions include ASPECT, the topic of the first set of case studies presented in Chapter 3. ASPECT captures "different ways of viewing the internal temporal constituency of a situation" (Comrie 1976: 3), for example, as bounded and complete (e.g. *she ate lunch*) or open (i.e. interruptible) and ongoing (e.g. *she was eating lunch (when…)*). For example, one experiment that examined gestures associated with progressive and perfect sentences found that "grammatical constructions such as aspect modulate how those simulations are performed" (Bergen & Wheeler 2010: 150). A number of other studies have drawn similar conclusions: linguistic meaning, including grammatical meaning, is dynamic and grounded in perceptual and cognitive experience (Matlock 2011; Matlock et al. 2012; Huette et al. 2014).

In addition to the studies exploring the gestural expression of aspect, eye-tracking studies have shown that eye movement aligns with the aspectual contour of an event expressed in speech. In a study in which participants listened to stories encoded as a series of ongoing actions in the past progressive form (e.g. *was speeding across the valley, was climbing a mountain*) or as a series of bounded (i.e. finished) events signaled by the simple past form (e.g. *sped across the valley, climbed a mountain*), eye movements differed according to the aspectual profile of the events. Eye movements that were aligned with stories that participants heard in the progressive aspect were both more dispersed (i.e. covered a wider area with more distinct points) and of a longer duration than eye movements in the simple past condition. The authors suggested that "the distribution and timing of eye movements mirrors the underlying conceptual structure of this linguistic difference" (Huette et al. 2014: 1).

In this section, I have provided a brief overview of embodied cognition. I have introduced key notions, including conceptual metaphors and image schemas, and reviewed experimental studies that have been taken as evidence that language processing is grounded in perception and cognition. I have also previewed studies of co-speech gesture that have demonstrated that speakers express underlying conceptual structure such as aspect in their co-speech behaviour.

If embodied cognition is one of the core underpinnings of cognitive linguistic approaches, how does this play out in descriptions and analyses of language? Beyond metaphors and image schemas, we need to consider units of linguistic analysis. In the next section, I introduce the notion of the construction as the basic unit of language and describe some of the basic tenets of construction grammars.

1.2.3 Constructions and construction grammars

Linguistic structures have long been analyzed as pairings of form and meaning. Known for his definition of the linguistic sign, de Saussure (1916) considered the central unit of language to be *signs*, arbitrary and conventional pairings of form (*signifiant)* and mental concept (*signifé*). More recently, in cognitive linguistic circles, the basic unit of linguistic structure is understood to be the *construction*, defined as a conventionalized pairing of form and meaning or form and function (Goldberg 2006: 1). Whereas the formalist 'items and rules'[1] approaches to grammar separate grammar and lexicon, in a constructionist view of language, grammar and the lexicon form a continuum of "assemblies of symbolic structures" (Langacker 2008: 15). Constructions are understood to exist at all levels of analysis, from morphemes and words to phrasal and discourse-level patterns. They also range in their degree of fixedness, for example, from lexical items and fixed idioms as in *kick the bucket*, which are inherently fixed (cf. (Geeraert 2016), to partially filled idiomatic expressions, such as the [the Xer the Yer] (*the more you think about it, the less you understand*) and [not X, let alone Y] (*I could barely get up*, *let alone make breakfast this morning)*. Such partially filled constructional forms are responsible, for example, for the recent additions to the North American English lexicon of the constructions [my X includes Y] (e.g. *My Canada*

[1] So named for the syntactic rules that combine words into sentences.

includes Quebec) and [X is the new Y], which is played upon in the title of the hit Netflix series *Orange is the New Black*. More general linguistic patterns also constitute constructions, such as the double object construction [V Object1 Object2] to encode the transfer of an object, as in *Mary baked John a cake,* and the caused motion construction, as in *Mary hit the ball out of the park*. The generalizability of constructional form allows for creative uses. For example, it is the caused motion construction that yields the novel sentence, *He sneezed his tooth right across town,* in Canadian children author Robert Munsch's (1998) story *Andrew's Loose Tooth.* (Munsch's sentence is itself an even more creative take on Goldberg's (1995) chestnut *He sneezed the napkin off the table*). Corpus linguists have played an important role in exploring constructions by focusing on collocations, or multiword chunks, which span both words and the grammatical constructions in which they occur. The effect has been to return conventionalized multi-word expressions, such as those explored throughout this dissertation, to the domain of linguistic analysis (Sinclair 1991; Stefanowitsch & Gries 2003; Gries & Stefanowitsch 2004; Stefanowitsch & Gries 2005).

The corpus-linguistic study of collocational patterns has led to the modeling of constructional networks, or schemas. Schemas are said to emerge from lexically particular constructions and become increasingly abstract and independent of those lexical expressions. For example, on hearing many tokens of the ditransitive construction, speakers abstract away from the specific utterances and generalize to the abstract construction. This network of constructions at all levels of structure that make up a language has been referred to as a *constructicon* (Goldberg 2003: 161; Steen & Turner 2013).

The family of approaches known as constructionist approaches include Berkeley Construction Grammar (Fillmore 1985; Fillmore et al. 1988), Sign-based construction grammar (Michaelis 2010; Boas & Sag 2012), Cognitive Grammar (CG) (Langacker 1987, 1991b, 2008), Cognitive Construction Grammar (CxG) (Goldberg 1995, 2006), Radical Construction Grammar (Croft 2001, 2012), and Embodied Construction Grammar (Bergen & Chang 2005, 2013)[2]. Although all of these approaches vary in ways that are non-trivial, the basic assumptions outlined in this section are shared (Hoffmann 2017: 312).[3] Construction grammars propose that grammar (i.e. structure) is meaningful. One of the significant shifts invoked by a constructionist approach is that grammar allows us to "construct and symbolize the more elaborate meanings of complex expressions [...] and is thus an essential aspect of the conceptual apparatus through which we *apprehend and engage* the world" (Langacker 2008: 3). To frame this in terms of embodiment, the grammatical patterns inherent to a language result from the embodied cognitive mechanisms that shape our conceptualizations (Johnson 2017: 28). Of course, embodiment is not the only factor in shaping language patterns, as cultural practices can also impact how patterns are conventionalized. Language structures can be iconic in vastly different ways, as mentioned in §1.2.2, above, with regard to the mapping of space and time. However, while patterns may vary according to culture and other factors, constructionalist approaches maintain that we learn our basic grammatical patterns through usage (Goldberg 2003), which leads me now to discuss the role of usage in learning and producing language and the importance of corpus methods to the study of usage.

---

[2] See Hoffmann (2017) for a full overview of each of these approaches and how they differ from and agree with each other.

[3] ECG differs from CG and CxG in that it is a computational implementation of CxG, whose goal is to represent an embodied simulation semantics model of language comprehension (Bergen 2012). As such it attempts to include the actual physical and social contexts of language use in its modeling.

1.2.4 Usage-based and corpus linguistic methods

To focus on two of the most widespread constructionist approaches, both Cognitive Grammar and Construction Grammar are usage-based models of language (Langacker 1987, 1991b; Kemmer & Barlow 2000; Bybee 2010). Usage-based models maintain that language is an inventory of conventional linguistic units that are abstracted from single usage events, such as the constructional schemas noted above. Structures arise from two cognitive processes over these usage events: schematization and categorization (Langacker 2008: 220). Language structures and units are thus seen as emergent:

> General cognitive capabilities of the human brain, which allow it to categorize and sort for identity, similarity, and difference, go to work on the language events a person encounters, categorizing and entering in memory these experiences. The result is a cognitive representation that can be called a grammar.
>
> (Bybee 2006: 711)

The usage-based perspective adopted by CL also entails that "generalizations (about language) are based on the analysis of authentic usage data provided by computerized corpora" (Gries et al. 2005: 635). Indeed, corpus-based methodologies have played a large supporting role in cognitive linguistic research. In the past decade, cognitive linguists have increasingly applied quantitative corpus analyses to major and minor patterns of language. These analyses have provided empirical evidence to support the exploration of major theoretical notions of particular interest to cognitive linguists, such as the claim that linguistic categories are radial categories evincing prototype effects. Usage-based studies have explored how to characterize patterns in language use with regard to prototype identification. Other important contributions of corpus studies include sense disambiguation and the relational structure of the network that constitutes language (Divjak & Gries 2006). Janda (2015) summarizes the outlook of cognitive linguistics, emphasizing the contribution of corpus methods to a cognitively oriented theory of language.

> Because cognitive linguistics is not in the business of prediction, it is also not looking for a set of concrete universals that would facilitate prediction, a goal that is probably neither desirable nor realistically achievable. In the big picture, cognitive linguistics' ultimate goal is to understand how human cognition motivates the phenomena of language, to be described in terms of statistical trends rather than absolute rules. One could say cognitive linguistics recognizes that human beings are not rule-guided algorithms, but individuals with a free will which they exercise in ways not entirely consistent and predictable.
>
> (Janda 2015: 133)

In this overview of cognitive linguistics, I have introduced the embodied nature of language, the theory of the construction as the primary building block of language, and the role of corpus linguistics in systematizing the analysis of actual language usage and exploring the principal tenets of a networked model of linguistic structure; however, while corpus linguistics systematized the study of actual language usage, as a field it remains dominantly text based. In the next section, I introduce recent developments in CL that have extended the commitment to usage to the analysis of language as it occurs in its most natural habitat, face-to-face interaction.

1.2.5 From interaction to multimodality

While the commitment of cognitive linguistics to actual language usage has been prominent since the early days of the field, a deliberate focus on speech-in-interaction has emerged more recently. This 'interactional turn' has been characterized by "an extension of its more traditional focus on speaker-centered conceptualization and construal mechanisms to a genuinely interactive and therefore socio-cognitive account of meaning" (Feyaerts et al. 2017: 136). Extending the commitment to usage, this perspective has as its starting point the place of dialogue as the most basic form of language use, which I introduced in §1.1. In such a view of language, interlocutors are situated in a usage event, regardless of the medium of the actual instance of language use. That is:

> Even in situations where no other interlocutors are directly involved, as in writing a newspaper article, producing or interpreting language can always be seen from an interactional perspective, of which the inclusion of previous, current, and future assumed expectations, attitudes, emotions and/or knowledge of one or more other, (un)known interlocutors constitutes an essential part.
>
> (Feyaerts et al. 2017: 136).

Meaning-making in face-to-face interaction goes beyond the range of phenomena that has been the purview of conversation analysis, which as a field has explored the array of practices in talk-in-interaction such as turn-taking, speech acts, intonation units, and repair (Ochs et al. 1996; Sidnell & Stivers 2013). In addition to these dialogic features of conversation, interaction brings to the fore a range of modalities beyond speech, such as facial expressions, as well as hand, head, and postural movements. Including the body as an important part of face-to-face interaction between speakers and signers allows a focus on what Turner (2017: 96) calls "ecologically real" language and a view of meaning as co-created in dialogue between speakers. While manual gesture has long been a focus in the field of gesture studies, the conventionalized forms and functional associations of other articulators are also now being included in analyses. The past decade of research in CL has seen a marked increase in studies that have included multiple modalities, such as manual gesture, gaze, head, and shoulder movement. Through his focus on the inclusion of discourse in Cognitive Grammar, Langacker previewed the move to account for gesture and multimodality as part of the conventionalized units in language. Langacker (2001: 146) stated that "any aspect of a usage event, or even a sequence of usage events in a discourse, is capable of emerging as a linguistic unit, should it be a recurrent commonality". While Langacker was referring to discourse-level phenomena, which had also been excluded from models of language to that point, he was ultimately arguing that meaning is created in each interaction and that any channel, including channels beyond speech such as co-speech behaviour, can play a role in the conventionalized mechanisms of language. As I introduce in the next section (§1.3), CL is thus theoretically open to including the kinesic channels, which are often recruited to convey pragmatically rich material, as in the expression of stance. The term *stance* itself uses a bodily metaphor to refer to a positioning, or, even more literally, the ground or footing on which someone speaks.

1.2.6 Summary: Cognitive Linguistics

In this overview of cognitive linguistics, I have introduced the framework of a cognitive and embodied theory of language, including the necessity of investigating language in use, the value of using corpora to do that, and the turn towards interaction that promotes the inclusion of a fuller range of modalities in language description and analysis. In the next section, I introduce the body as a linguistic resource, which is grounded in the field of gesture studies and which has come to be known more recently as multimodal linguistics.

1.3 Multimodality

As described in the previous section, the current discussion within cognitive and usage-based linguistics is that the study of linguistic meaning should be based on language as it occurs in everyday interactive contexts. Such contexts are (or should be considered) multimodal contexts since they include the body (McNeill 1992; Kendon 2004; Zima 2017b). The notion of embodiment as a primary, if not *the* primary, resource for meaning-making and conceptualization has led to an expansion of linguistic research into multimodal communication (Sweetser 1998; Mittelberg 2006, 2010; Müller, Cienki, et al. 2013; Müller et al. 2014). As outlined in §1.2, cognitive linguistics is characterized by a "conceptual openness to *all levels* of usage features" (Zima & Bergs 2017a: 1; italics mine). As such, studies over the past few years have argued for the need to include kinesic (and prosodic) patterns in linguistic analysis (Schoonjans 2014; Debras 2017; Lanwer 2017; Zima 2017b). From a theoretical perspective, the question that arises, then, is whether the construction, as the basic unit of spoken language, should be considered to be inherently multimodal.

In this section, I review primary tenets and seminal research in the fields of gesture studies and multimodality. I begin with a working definition of gesture, including types of gesture. I move on to a brief survey of different approaches to the nature of gesture, including the various ways in which it can be considered iconic, and major views on how speech and gesture interact in language production and comprehension. I then turn toward recent major findings in studies in multimodal linguistics. I close this section by returning to the question of what constitutes the conventionalized form-meaning pairing known as a construction, from an interactional, multimodal perspective and consider the question: what does multimodality contribute to the study of language?

1.3.1 What is gesture?

As alluded to at the beginning of this chapter, there are a host of ways in which people move their bodies in communicative situations. In the field of gesture studies, the term *gesture* refers to meaningful bodily movements that are produced in the course of communication as part of communicative utterances. Such gestures are referred to as *co-speech gestures*. They have also been captured by the term *visible bodily action* (Kendon 2004). The key is that these movements manifest "deliberate expressiveness" (Kendon 2004: 15). Thus, the study of co-speech gesture excludes practices that are not related to a deliberate communicative act, such as sipping a cup of coffee or fiddling with one's hair. These actions are not integrated with speech in the same way as co-speech gestures.

Gestures can go by in a blur during the course of a conversation; however, when slowed down, the gestural movements of the hands can be segmented into units. Kendon referred to a movement excursion (or series of excursions) as a *gesture unit* (Kendon 1980: 111ff). This unit consists of a gesture phrase composed of five unique phases: rest, preparation, stroke, hold, and recovery (or retraction). The stroke phase is the apex of the gesture and is the most meaningful and effortful phase. When considering the relationship of speech to gesture, the stroke "correlates with the most relevant part of the verbal utterance" (Bressem 2013: 1102). The classification of gesture forms is based on the stroke phase.

With these basic definitions of gesture in mind, I now outline the ways in which gestures have been classified.

1.3.2 Typologies of gesture

In contemporary gesture studies, there are many different typologies that have been used for the classification and annotation of (mainly manual) gesture. While they vary in their specifics, what unites the typologies is their focus on defining gesture in terms of four parameters:

(i) how they relate to speech
(ii) the extent to which they have linguistic properties
(iii) the extent to which they are conventionalized, and
(iv) how they contrast in terms of their semiotic properties

(Kendon 2004: 106)

One of the most commonly cited typologies of gesture is 'Kendon's continuum' (McNeill 1992), which McNeill named in honour of Adam Kendon, one of the founders of contemporary gesture studies. McNeill (1992) suggested that there are four types of gesture:

gesticulation, pantomime, emblems, and signs in sign language. These gesture types differ from each other along the continua listed in (i) to (iv), above.[4] To begin with signs in sign language, these are gestures that function entirely independently of speech and are fixed in the same way that words in spoken language are fixed. Emblems and pantomimes are also gestures used in the absence of speech. Emblems are standardized gestures, such as the peace sign, which constitute a complete utterance by themselves, while pantomime, or depictive gestures, enact an event and can replace part of an utterance, such as: "And then I had to go like *this* [mimes wrenching a steering wheel to the side] to avoid hitting them." Finally, co-speech gestures, or gesticulations, complete the typology. These are gestures used as "unwitting accompaniments of speech" (McNeill 1992: 72). These gestures have long been described as idiosyncratic, ad hoc, and lacking in conventionality. Such spontaneous gestures have also been described as global and holistic as compared to the conventionalized signs of a sign language, which are compositional in structure and organization.

Gesture classification also relies heavily on the philosopher Charles S. Peirce's distinctions between the three ways that signs (linguistic or non-linguistic) can carry meaning. Firstly, in an indexical meaning relationship, gestures refer directly to a referent in space and time. Indexical signs in speech include deictic expressions such as demonstratives, verb tenses, adverbs of time and place, and social deixis. In gesture, indexicality usually centers on pointing gestures, which have been investigated across many cultures (McNeill 1992; Cooperrider & Núñez 2012; Fricke 2014; Cooperrider & Goldin-Meadow 2017). Gestures can also mean by

[4] While the term 'Kendon's continuum' originally gave the impression that there was one continuum on which the four types of gestures lie, McNeill later amended the notion to state that it is the four parameters that are continuua and that the gesture types can be contrasted with each other in terms of the continua in (i) to (iv) (McNeill 2000).

appearing to be similar to what they are referring to. By depicting at least some of the properties of its referent, such signs are iconic. Given the ability of the hands and arms to both form shapes and to move through space, the gestural modality excels at depicting human actions, motion through space, and spatial relationships (Cooperrider & Goldin-Meadow 2017: 120). This ability to be highly depictive has made iconic gesture forms a very robust area of research. Beyond depicting actions and spatial relationships, gestures also represent abstract ideas. Gestures that refer to abstract concepts are known as *metaphoric gestures,* to distinguish them from *iconic gestures*, in which the referent resembles the gesture form in some way.

According to Peirce's triad, the third way in which a sign or gesture can carry meaning is through conventionality. Conventional gestures include emblems, such as the peace sign noted above, or the 'thumbs-up' gesture known to North American English speakers. These are signs that have a standard form that is not compositional and cannot be altered. For example, one cannot replace the raised thumb with a raised finger in the thumbs-up gesture.

It is important to keep in mind that this account of ways in which gestures can mean is not intended to be exclusive. That is, a sign can simultaneously incorporate all three modes: indexicality, iconicity, and conventionality. It has been argued that all gestures are inherently indexical, since "they are always contextualized and indexical of the object they represent. They point to something other than themselves" (Mittelberg 2008: 124). Cooperrider and Goldin-Meadow describe how a beckoning gesture combines all three modes of meaning:

> The gesture is indexical in that it 'points' toward the intended recipient; it is iconic in that it represents the path between the gesturer and the recipient; and, finally, it is conventional to the extent that, in some cultures, people beckon with the palm up (e.g. in the US) and in other cultures, they beckon with the palm down (e.g. in Mexico).
>
> (Cooperrider & Goldin-Meadow 2017: 121)

Gestures have also been classified according to their function. Iconic and metaphoric gestures are known as *content gestures*, which relate to what the speaker is talking about. Content gestures have occupied the lions' share of studies in modern gesture studies. They differ from so-called *interactional* or *pragmatic gestures,* which operate at the level of the interaction (Bavelas et al. 1995; Kendon 2004). Since pragmatic gestures have been defined with regard to their function in the process of interaction, they have also been described as *speech-handling* (Streeck 2009) and most recently as *discourse management gestures* (Wehling 2017). In some ways, the term *speech-handling* is the most apt of these terms, given that pragmatic gestures manipulate objects in the discourse. They enact how a discourse object may be presented, held, withdrawn, forced on an interlocutor, wiped away, and so forth. When these gestures are made with regard to discourse objects, they display the speaker's stance towards the discourse object.

Pragmatic gestures also play a role in turn-taking, turn-holding, engaging the interlocutor, and carrying out speech acts. They have been shown to signal inclusion and cooperation or control-seeking in discourse. Gestures that indicate inclusion and cooperation include pointing gestures that reference the interlocutor (Bavelas et al. 1995). Metaphorical gestures such as the palm-up open-hand gesture can function as a pragmatic gesture through which speakers offer information to each other (Sweetser 1998; Kendon 2004; Müller 2004). Control-seeking gestures signal argumentative functions in discourse, e.g. head pushes and forward leans may signal offensive force towards a speaker or discourse content, while holding away gestures embody a defensive force (Wehling 2010, 2017).

Kendon describes the possible functions of pragmatic gestures in three groupings: *performative*, which show what type of speech act or move a speaker is engaging in; *modal*, which operate on a given unit of discourse and show how it is to be interpreted; or *parsing*, when they "contribute to the marking of various aspects of the structure of spoken discourse" (Kendon

2004: 225). These are not intended to be exclusive to a particular hand shape. As Kendon asserts, "any given gestural form, may, according to context, function now in one way, now in another" (ibid.). He gives as an example a palm-down gesture decisively moving laterally outwards. This form could in one context be understood as an act of rejection or denial, in which case it would be performative in its function. It could also be seen as an intensifier in a context in which negation is implied, in which case it would be considered modal. Finally, if it were used in the context of the end of a line of argument, at which point the speaker uses it to signal she will move on, it would be considered to be a parsing gesture.

In this section, I have provided a brief overview of types of gestures and how these are categorized, including the main considerations for categorizing their form and function. Of course, the typologies are never as cut and dried as the definitions here would suggest. I address the multifunctional nature of many gestures again in the next section (§1.3.3), in which I describe in greater detail pragmatic gestures, in particular those that have recurrent semantic themes and have therefore been ascribed to *gesture families*, or recurrent hand shapes and movements that subsume a common semantic or pragmatic core.

1.3.3 Recurrent gestures and gesture families

Studies of pragmatic gestures in their contexts-of-use have revealed groupings of recurring gestural forms known as gesture families, or "structural island[s] of interrelated gestures" (Bressem & Müller 2014: 1592). Gestures are said to be part of a family when they share a semantic and formational core over different speakers and different contexts-of-use. According to Kendon:

> When we refer to *families* of gestures, we refer to groupings of gestural expressions that have in common one or more kinesic or formational characteristics. […] Within each family, the different forms that may be

recognized in most cases are distinguished in terms of the different *movement patterns* that are employed. […] Each family not only shares in a distinct set of kinesic features but each is also distinct in its semantic themes. The forms within these families, distinguished as they are kinesically, also tend to differ semantically, although, within a given family, all forms share in a common semantic theme.

(Kendon 2004: 227; italics original)

More recent work has defined these recurrent pragmatic gestures as stable form-meaning pairing across a gesture and speech community (Ladewig 2014; Müller 2017). Müller describes families of gesture as follows:

> Their basic form concerns a shared Gestalt of selected features or parameters (for instance, hand-shape and orientation) that does not vary across contexts and that comes with a more or less conventionalized basic prototypical meaning. This meaningful form Gestalt is considered the kinesic core. The remaining formational features may be used to specify and alter the meaning of the kinesic core spontaneously and according to local affordances of the communicate situation. Recurrent gestures are hybrids of idiosyncratic and conventional elements.

(Müller 2017: 280)

Ladewig (2014) argues that pragmatic gestures have generally been investigated according to their conventional forms, rather than based on their pragmatic functions. She maintains that the form is tied to the pragmatic function and therefore argues for the use of the term *recurrent*, rather than pragmatic, to describe gesture families in order to better capture their conventional physical character in addition to the shared pragmatic function. In line with McNeill (1992), Ladewig argues for a distinction between singular and recurrent gestures. In her assessment, singular gestures constitute spontaneous creations and are part of the propositional content of an utterance (including the iconic and metaphoric gestures described above), whereas recurrent gestures fulfill a performative function and, crucially, form a repertoire of gestures shared within a culture. Ladewig also objects to the use of the term *pragmatic gesture* on the basis that the term implies that only a particular type of gesture can fulfill a pragmatic function. Rather, she notes, pragmatic gestures have variants that serve referential functions. She provides the example of a cyclic gesture that can depict either a scooping motion (a depictive gesture) or a thinking or word-finding process (a pragmatic gesture). These are variants of the same gesture form. She also notes that emblematic gestures (Kendon's emblems), although much more conventionalized than recurrent gestures, can also fulfil pragmatic functions. Müller captures the characteristics of recurrent gestures as follows:

> By merging conventional and idiosyncratic elements, recurrent gestures occupy a place between spontaneously created (singular) gestures and emblems as fully conventionalized gestural expressions on a continuum of increasing conventionalization (cf. Kendon's continuum).
>
> (Müller 2017: 276)

Studies of gesture families have focused on how form variants parallel differences in meaning and motivation of gesture. Well-known gesture families include the precision grip and the ring families of gesture (Morris et al. 1979; Kendon 2004), the palm-up and palm-down gesture families (Kendon 2004; Müller 2004), the cyclic gesture (Ladewig 2011, 2014), and the Away gesture family (Bressem & Müller 2014). Since many of these are treated in this dissertation, I review the most significant recurrent gesture families and gesture forms: the palm-up, palm-down, and Away families of gestures, as well as the cyclic gesture[5].

What these gesture families share is that they represent schematized versions of actions. The kinesic core of the palm-up open-hand (PUOH) is the open hand shape and upwards orientation. This form is said to relate the gestures of the family to their origin in actions of giving, showing, offering or receiving an object "by presenting it in the open hand"

[5] Naming, hyphenation, and capitalization conventions for gesture forms and gesture families vary among authors. For the purpose of this dissertation, I use lowercase for all gesture forms. I use capital letters only for abbreviations of common gesture forms, e.g. palm-up open-hand is abbreviated PUOH.

(Müller 2004: 254). These basic actions motivate the metaphorical use of the PUOH, in which a speaker-gesturer manipulates an idea or a discursive object. Varying the form of the kinesic core, e.g. by adding rotation or lateral movement extends the semantic meaning PUOH gesture to include continuation, a listing of ideas, and sequencing in discourse (ibid.).

The palm-down gesture family features gestures that are "used in contexts where something is being denied, negated, interrupted, or stopped, whether explicitly or by implication" and which "share the semantic theme of stopping or interrupting a line of action that is in progress" (Kendon 2004). The underlying semantic theme of the palm-down family has also been expressed as physical refusal (Calbris 2011: 248f). Calbris delineates between active and passive refusal, noting that this distinction "allows one to sort gestural signs into those derived from reflexes of rejection, self-protection, and avoidance respectively" (Calbris 2011: 199).

Kendon has documented two primary members of the palm-down gesture family: the Open Hand Prone Vertical Palm (OHP VP), in which the palm is oriented vertically, and the Open Hand Prone Horizontal Palm (OHP ZP), in which the palm is facing downwards. The Open Hand Prone VP gestures are known as *holding away* gestures and are used to "establish a barrier, push back, or hold back things" (Bressem & Müller 2014). With palm facing down, the OHP ZP gestures, by contrast, are based on actions of "cutting something through, knocking something away, or sweeping away irregularities on a surface" (Kendon 2004: 263). Others have referred to these as "finished" (Brookes 2004), "cutting" (Calbris 2003), and sweeping away gestures (Bressem & Müller 2014). Harrison (2018) has documented the use of palm-down gestures in contexts of negation in speech. He argues that there is a correlation of palm-down gestures with the node and scope of negation in speech.

Recent studies have shown that there may be a larger, more schematic family to which palm-down gestures belong, namely, the family of Away gestures (Bressem & Müller 2014). What unites this family is the action of removing or keeping things away. Rather than a kinesic core anchored in a particular handshape and orientation, as the families reviewed above have been, Away gestures are united by a motion away from the body, for the most part in a straight manner. The semantic theme for the Away family has been described as follows:

> […] the family is bound together by the themes of rejection, refusal, negative assessment, and negation, which are directly derived from the semantics of the underlying action scheme, in particular, from the effect that actions involving the clearing of the body space have in common: Something that was present has been moved away or something wanting to intrude has been or is being kept away from intrusion.
>
> (Bressem & Müller 2014: 1596)

Members of the away family include sweeping away, holding away, brushing away, and throwing away (see Bressem and Müller 2014 for a full treatment). To foreshadow the findings presented in this dissertation, the palm-down gesture is shown to be associated with the verbs STOP and QUIT in the study of aspect marking (Chapter 3), while more variants of the family of Away gestures are featured in Chapters 4 and 5, in which I investigate expressions of contrast and stanced expressions of discourse navigation.

Finally, the cyclic gesture family has received treatment in the literature on recurrent gestures. The cyclic gesture is characterized by a kinesic core involving continuous rotational movement away from the body and correlates with the semantic core of cyclic continuity (Ladewig 2011, 2014). Again, as with Away gestures, the cyclic gesture is not associated with a specific hand shape or orientation. When used descriptively, it depicts ongoing actions or events and also occurs in contexts of requests and when speakers are searching for a word or concept. Form variations that have been shown to affect semantic extension include the position in gesture space and the amplitude, or size, of the gesture. In this dissertation, cyclic gestures were found to play a role in gestures that co-occur with auxiliary verb constructions headed by the auxiliaries CONTINUE, KEEP, and START, as profiled in Chapter 3.

Like singular gestures, recurrent gestures can serve referential functions as kinds of iconic and metaphoric gestures, as well as pragmatic functions. In this dissertation, I will introduce gestures that mark aspectual information, that highlight opposition or contrast, and that co-occur with discourse-marking. Referential gestures are seen more in the marking of aspect, while pragmatic gestures occur predominantly in the marking of discourse navigation; however, the gestures that feature throughout the dissertation frequently serve both referential (e.g. content) and pragmatic (e.g. discursive) functions.

So far in this section, I have defined gesture and introduced typologies and principal recurrent forms of gesture. In the next section, I turn my attention to the many other ways the body contributes to the linguistic signal.

1.3.4 Multimodality: Beyond the hands

In addition to the focus on manual gesture and its role in conveying meaning in interaction, there has recently emerged a literature in cognitively-oriented fields that focuses on head movement, gaze, and other co-speech behaviours that are coordinated with speech in face-to-face conversation (Müller, Cienki, et al. 2013; Müller et al. 2014; Schoonjans 2014; Debras 2017; Jehoul et al. 2017). Feyaerts et al. (2017) note that this development in cognitive fields is late to the game. Earlier work in psychology on facial expressions examined them as markers of emotion and conversation analysis has long included intonation in its analysis of interaction (Sacks et al. 1974; Ekman 1993). Here, I introduce central findings on the functions of head and shoulder movements, gaze, and postural shifts such as leans.

Speaker head movements have been shown "to pattern predictably and have semantic, discourse, and communicative functions" (McClave 2000: 855). In her description of the multifunctionality of head movements in the discourse of speakers of North American English speakers, McClave (2000) identifies the following general functions of head movements and their associated forms:

(i) side-to-side shakes indicate inclusivity and intensification
(ii) lateral movements indicate uncertainty and lexical repair
(iii) locating referents in abstract space
(iv) shift in head posture marks shift from direct to indirect discourse, and
(v) head nod functions as backchannel request.

Similar to the role of the head to mark direct vs. indirect discourse listed in (iv), head shifts have been shown to mark quotatives in American Sign Language (ASL) (Shaffer 2012). Head movements have also been shown to be systematically correlated with specific expressions in speech. For example, in studies of co-speech behaviours that occur with German modal particles, the intersubjective downtoning particles *einfach* ("just") and *schon* (which marks truth despite counterarguments[6]), are regularly expressed with a headshake and nod respectively (Schoonjans et al. 2013; Schoonjans 2014).

Gaze also regularly marks specific functions in interactions, for example, speakers use gaze to direct their addressee's attention. We know that young children are sensitive to this effect and that adults most frequently look at the referent for the subject of their sentence (Ibbotson et al. 2013). Gaze has also been shown to play a critical role in structuring social interaction functions such as alignment and turn-taking. For example, speakers shift their gaze away to communicate to their interlocutor they are holding the floor while trying to recall

[6] An example of *schon: So the concert yesterday was ("schon") great, although actually I don't really like Free Jazz.*

something (Chovil 1989; Bavelas & Gerwing 2007). More recently, studies have investigated the role of gaze patterns in relation to other signal systems in dialogue management. For example, with regard to turn-holding functions, gaze and speech appear to be coordinated, with verbal fillers occurring in synchrony with gaze aversion in turn-holding contexts (Brône et al. 2017). Furthermore, when a speaker looks at her interlocutor while uttering a word, the interlocutor is significantly more likely to use that word, whereas this lexical alignment effect is not mirrored for gestural alignment (cf. Oben & Brône 2015 and Feyaerts et al. 2017). In addition to marking memory access, gaze has also been shown to impact how speakers mark narrator and character viewpoint, with gaze shifts marking alternation between enacted characters (Sweetser & Stec 2013).

Brow movement is also known to have linguistic properties. In ASL, brow movement is known to be a marker of topic-comment constructions. Topic phrases are marked with raised eyebrows and a backward head tilt over the entire phrase. The signer then returns to neutral for the comment clause. Thus, in ASL these topic markers signal a topic shift, rather than topic maintenance (Janzen 1998, 1999). There is no literature to my knowledge on co-speech brow movement, other than as a part of the shrug composite (Debras 2017).

While there is not much literature specifically on leans and movements of the torso, leans have been shown to be part of the enacted profile, for example, in viewpoint and accompanying inclusive-cooperative manual gestures and argumentative ones. Regarding viewpoint, Parrill (2009, 2010) found that when viewing videos and then retelling the events, speakers were more likely to adopt a character viewpoint if the character whose actions they are retelling used his or her hands and torso in the action. With regard to the discourse functions of leans and postural sets, Wehling (2017: 257) suggests that "stepping or leaning away from an interlocutor or bystander in a friendly conversational setting may signal an invitation to speak or to join a discourse. […] In an argument, however, the stepping or leaning away from an interlocutor […] may signal the ceding of space that has beforehand been (illegitimately) occupied."

In this section thus far, I have surveyed some of the known functions of head movement, gaze, brow, and torso movement. The majority of the literature on these kinesic elements has examined only one articulator at a time in combination with speech, looking at speech and manual gesture or speech and gaze in eye-tracking studies; however, many co-speech behaviours co-occur not only with speech, but with other articulator movements. For example, gaze and gesture together play a role in alignment and viewpoint and both gestures and postural leans can play a role in inclusive and argumentative body actions in discourse. Only recently have researchers begun to integrate different articulators into single studies. One exception to this has been the work on shrugs, which integrates multiple articulators in its full profile.

Long viewed as a "densely communicative human behaviour" (Givens 1977), the shrug typically involves a bilateral lift of both hands and forearms into a supine orientation, along with a raise of both shoulders, and it can include raised eyebrows (Streeck 2009; Debras 2017). More recent studies show that both head tilts and head shakes can also be a part of the shrug (Schoonjans 2014). In early work in gesture studies, the shrug was classified as an emblem, a gesture with a stable form-meaning convention that can be used and understood independently from speech (Ekman & Friesen 1969; Efron 1972; Morris 1994). More recent investigations of the shrug have suggested that specific sub-categories of stance (including attitude of helplessness, affective distance, and epistemic expressions of ignorance or uncertainty) are associated with specific shrug postures (Debras 2013: 233). For example, mouth shrugs or shoulder lifts alone tend to express epistemic ignorance or indetermination (e.g. "I don't know" shrugs), while shrugs that indicate common ground ("you know" shrugs) tend to involve forearm and open palms supine. Debras argues that, far from being fixed emblems, the shrug is composed of a variety of features which vary in terms of the number of components that are used in the enactment in each context of use. As such, it "constitutes a more complex and dynamic network of related forms and functions" (Debras 2017: 5). Given its complex and compositional nature, the shrug has been described as a "compound enactment" (Streeck 2009; Debras 2017), a "shrug complex" (Morris 1994), and "shrugging composite" (Givens 1977). These terms align with the term *composite utterance* as involving not only composites of speech and gesture, but composites of multiple body articulators including vocal intonation and eye gaze, in addition to other modalities (e.g. diagrams) (Enfield 2013).

In this section, I have introduced the basic forms and associated functions of conventionalized body movements. These pairings can include not only speech and one articulator, but speech and many articulators, as in the case of the shrug complex. Some instantiations may feature only a few of the conventionalized features, while others may feature more. As Janzen states: "instantiations incorporating speech, prosodic features, gesture, eye-gaze, etc., thus lead us to consider multimodality as a many-faceted expression of language; individual utterances may incorporate some or all of these features variably" (Janzen 2017: 519). Earlier, I surveyed the literature that views the basic unit of language as a construction (generally considered in a monomodal sense until recently. Integrating a view of language as featuring composite utterances involving multiple modalities and articulators invites us to consider to what extent every instance of language use constitutes an instantiation of a *multimodal construction*. The notion of a multimodal construction is the subject of the next subsection.

1.3.5 Multimodal constructions

The recent focus on conventionalized gesture forms and families within the field of gesture studies and the adoption of gesture and multimodality into cognitive linguistics have prompted a swell in discussions of how, exactly, one can account for the nature of co-speech behaviour in a theory of language, let alone in a descriptive grammar of a single language. Given the increasing role of utterances from actual interaction (rather than decontextualized syntactic structures) and the centrality of the construction as the basic unit of language, the question has centered on the extent to which a construction can be considered to be inherently multimodal. In this section I preview some of the discussion on this issue to date.

A key component for construction-hood is conventionalization: a construction must occur with "sufficient frequency" (Goldberg 2006: 6). Over the last two decades, scholars have relied heavily on quantitative corpus methods to establish the degree of conventionalization, or entrenchment, of specific speech-only linguistic constructions (Stefanowitsch & Gries 2003; Gries et al. 2010). For example, (Gries 2006: 57) develops a "complete behavioural profile" of the English verb *to run* using the International Corpus of English-Great Britain and the Brown Corpus. In the profile, the sense of run as 'fast pedestrian motion' accounts for 25% of the data. Only much more recently have there been corpus-based studies of multimodal data that have included a measure of the rate of co-occurrence of co-speech behaviour with particular utterances. This approach is still in its infancy (Zima & Bergs 2017b). Schoonjans (2014) found a 23%-53% correlation between German modal particles and certain forms of head movement. In a study comparing English spatial constructions and motion constructions, co-occurrence of manual gestures was reported to be 37%-75% for English motion verbs (Zima 2017a), while in a study of the partially filled English construction [all the way from X PREP Y], manual gesture occurrence was much higher, at 80% (Zima 2017b). I will report on co-occurrence throughout the case studies presented in this dissertation. Although studies are now tracking the rate of co-occurrence, the main question remains: given that the degree of convention in a speech-only construction has been taken as a marker of its degree of entrenchment in the cognitive network, what is an adequate degree of conventionalization to take as evidence of entrenchment?

Lanwer (2017) problematizes the issue of co-occurrence and its relation to entrenchment as follows:

> Defining what is meant by 'in a sufficient number' seems to be substantial if we want to determine the unit status of a given pattern in the light of frequency measurements. However, this only seems to be necessary if we want to establish a clear-cut distinction between patterns which are entrenched and patterns which are not.
>
> (Lanwer 2017: 3)

Citing Langacker's observation that "linguistic structures are more realistically conceived as falling along a continuous scale of entrenchment in cognitive organization" (Langacker 1987: 59), Lanwer concludes that "it seems to be unnecessary, if not impossible to define something like a frequency-threshold" (Lanwer 2017: 3). In their review of numerous studies, Feyaerts and colleagues similarly conclude:

> The co-occurrence rates of verbal and gestural patterns range in frequency, depending on the perspective one takes, but they (obviously) never reach a 100 percent match. Where one needs to draw the line, however, still is very much a point of debate and is considered by some to be impossible to operationalize.
>
> (Feyaerts et al. 2017: 147)

Even in studies that investigate the degree of recurrence of a gesture over different contexts of use, the issue of what constitutes conventionalization in gesture is far from clear. Manual gestures have been described as hybrids of idiosyncratic and conventional elements

(Müller 2017: 293) and as an "evolving set of largely improvised heterogeneous, partly conventional, partly idiosyncratic, partly culture specific, partly universal practices of using the hands to produce situated understandings" (Streeck 2009: 5). This begs the question, of the elements enacted in a token of composite utterance, which elements of a gesture need to be the 'same' in order to be considered to be an instantiation of a particular prototype? If an utterance is articulated in one moment by one speaker with a head shake and in a different moment by a different speaker with a head shake and a shoulder shrug, or even more differentially with a head nod, how does a grammar account for this variation?

The issue of accounting for different levels of convention and schematicity is not unique to multimodal behaviour. When using a text corpus, for example, data are only rendered alike by looking through a narrow lens (e.g. focusing on only a small set of linguistic variables) and ignoring factors such as prosody that a text-based rendering completely obscures. No two samples of language use are ever entirely alike. It has simply been the habit of basing language analysis on idealized and text-based representations of data to remove much of the messiness from the picture. The result, I would suggest, has been to make language seem more systematic (and susceptible to prediction) than it is. That being said, the inclusion of co-speech behaviour in any linguistic model certainly multiplies the chance of variation, either across different speakers or different usage contexts.

1.3.6 Summary: Multimodality

In this overview of multimodality, I have presented current research in the field of gesture studies, which has more recently embraced an array of meaningful bodily behaviour beyond the hands, including facial expressions, gaze shifts, and shoulder and torso movement. In addition to surveying studies that have shown the specific (and heterogeneous) functions of each of these movement types, I have discussed how manual gesture forms inhere to gesture families in which clusters of recurrent forms can be associated with generalized semantic themes. Finally, I have introduced the notion of a multimodal construction and outlined the primary theoretical questions that multimodal constructions pose for cognitive linguistic theories of language.

1.4 Conclusion

In this introduction, I have laid the groundwork for the central questions that this dissertation seeks to address. I have outlined the primary elements of the embodied, cognitive, and interactional approach to language that underlie the case studies. I have introduced some of the pressing questions surrounding the implication of considering multimodality as a, if not the, primary form of communication. Below, I provide an outline of the structure of the dissertation and offer concluding comments.

1.4.1 Structure of the dissertation

The dissertation consists of six chapters. Following this introductory chapter is a methodological overview (Chapter 2), in which I introduce the methods of corpus-based multimodal analysis. This is followed by three chapters, each of which presents a series of case studies in a particular domain of conceptual and linguistic structure. In what follows, I briefly describe each of these chapters. The final chapter (Chapter 6) offers closing comments and directions for future research.

In the first study (Chapter 3), I present an investigation of five aspect-marking periphrastic constructions in North American English. Using English-language naturalistic interactional data from the Red Hen archive (introduced in Chapter 2), I investigate grammatical expressions that do the work of marking aspect, a linguistic term that captures the different ways in which the internal temporal contour of an event can be construed. I describe the multimodal behaviour of auxiliary constructions headed by one of five highly aspectualized auxiliary verbs: CONTINUE, KEEP, START, STOP, and QUIT, as in *The jackpot continued to grow* and *He quit smoking*. Findings suggest that the onset timing of the manual gesture compared to speech, the internal structure of the gesture stroke, and the gesture movement type, are variables that iconically and differentially represent distinctive aspectual conceptualizations. In other words, when there is an incidence of co-speech gesture accompanying a set family of expressions (i.e. phrases with a specific, grammaticalized verb acting as an auxiliary), the utterance is characterized by a conventionalized kinesic profile that mirrors the unique aspectual contour of the auxiliary verb in a specific utterance.

While the study of ASPECT consists of a small, fixed set of grammatical constructions that are shown to behave conventionally, in the second case study, I look at a wider range of expressions that fall into a different domain of verbal expression. In Chapter 4, I present a study of CONTRAST, which, schematically, could be said to involve the valuing of one position over another or the pitting of one position against another. The conceptual binariness inherent in CONTRAST is paralleled by a binariness in the movement that the co-speech behaviour can take: shoulders can move up or down, hands can move in and out, and the head can tilt from side-to-side or nod up or down. Even a head shake is a binary, side-to-side motion on the lateral axis. To explore the kinesic enactment of CONTRAST, I examine a set of contrast-marking linguistic expressions. The prototype for embodied expressions of CONTRAST in English is the fixed expression *on (the) one hand/on the other hand*, in which manual gestures performed sequentially on opposite sides of the body by the left and right hand depict the options inherent in the expression. English speakers also have a variety of other linguistic means for expressing opposition or CONTRAST without explicitly mentioning the hands. These include the logical operator *or;* phrasal units such as *by contrast* and *whether (or not)*; semi-fixed idiomatic expressions such as a *David and Goliath situation*; lexical pairs (e.g. *offense/defense);* and modal verbs (*should I/shouldn't*), and others. These expressions are often accompanied by the same bilateral embodiment of CONTRAST as seen with *on (the) one hand/on the other hand,* even though they lack an overt reference to the body. I treat CONTRAST as inherently stanced, given its basis in evaluation. Some expressions of contrast are more evaluative than others. For example, the fixed expression *on (the) one hand/on the other hand* is usually used in more objective contexts than is the modal expression *should I or shouldn't I.* While conventions associated with the conceptual notion of CONTRAST underpin each of these expressions generally, I show in this chapter that the specific ways in which CONTRAST is expressed in the body depends on a variety of factors. These include the degree of fixedness and the domain of the utterance, i.e. whether the contrasted elements belong to the real world or to a subjective, hypothetical one.

In Chapter 5, I examine a set of expressions that are used by speakers to navigate discourse. CONTRAST plays a role here as well, given that speakers often create junctures in their discourse that fall under the broad conceptual category of CONTRAST introduced in the previous chapter. Perhaps even more than CONTRAST, discourse navigation is inherently stanced, given that it involves the speaker's subjective view of the direction the discourse should take at any given moment. I examine a range of expressions speakers use to depart and return to the main conversation frame. There are a number of highly recurrent expressions in English that convey such digressions from a discourse path. As this chapter will show, many of these have recurrent kinesic signatures. Fixed expressions that I examine include *which is true; which is fine; which, by the way; how shall I put it?;* and *although, you know.* As 'toss-off' parenthetical comments, these (inter)subjective qualifiers are sign posts in the discourse and are frequently matched by co-speech behaviors that depart from the bodily actions that precede or follow them. I also examine the kinesic enactment of discourse markers that speakers use to return to the main path of the conversation, such as *however* and *but anyway(s)*. Again, in these cases, the bodily articulators that speakers recruit (i.e. the hands, head, shoulders, brows, torso) and the form that these articulators take (e.g., palm-up and palm-down gestures, head tilts and nods, brow raises, shoulder shrugs, and torso leans) are reliably associated with particular expressions.

The dissertation concludes in Chapter 6 with a discussion of multimodal constructions in light of the findings of the case studies. I also offer some concluding remarks about the contributions that linguistic multimodality can make across a range of disciplines and in society (and industry) at large.

1.4.2 Closing comments

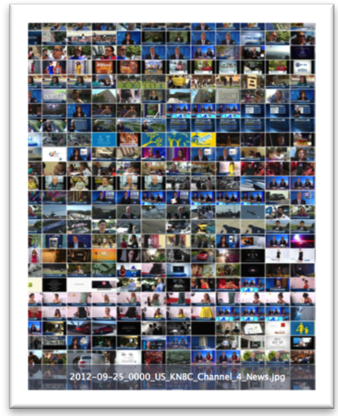
Throughout this dissertation, one of my primary aims is to provide new and convincing evidence to support the claim within cognitive linguistic circles that the construction is the primary unit of linguistic analysis and needs to be conceptualized as a multimodal entity with verbal *and* kinesic form, at least for spoken language. Since the earliest days of CL, researchers have been motivated to explore the complex nature of actual language usage. The usage-based ethos of cognitive linguistics aligned with the development of large databases of language and computational tools to yield the "quantitative turn" in the field (Janda 2017). Though textual language still dominated, corpus linguistic analyses have focused on patterns of collocations or 'chunks' of language that prevail in usage but are not restricted to well-behaved syntactic patterns. When applied to spoken language particularly, corpus analyses capture "[a] realization that languages may be less productive and compositional than we thought, with speakers

opting to rely on a smallish set of lexically frozen and semantically opaque (though pragmatically rich) collocations" (Rice 2017a). That is, oral language is characterized by holistic, non-compositional units and a high degree of synthesism. The set of utterances that I explore in this dissertation exhibit this highly conventionalized characteristic of oralcy.
Oral language is – by its very nature – embodied language and with the onset of the recent "interactional turn" in the field, CL has begun to prioritize face-to-face interaction as the primary locus of language use (Feyaerts et al. 2017). This turn has placed both oral language and its embodied nature at the forefront of linguistic analysis. As the case studies presented here show, multimodal language exhibits some of the same characteristics as oral language, including a high degree of synthesism, which can be seen across modalities (e.g. speech and the body) as well as channels (e.g. head vs. hands for co-speech behaviour). This holistic, synthetic characteristic of spoken and embodied language has been captured in the terms *compound enactment*s as for a shrug (cf. Streeck 2009; Debras 2017) and *composite utterance* (cf. Enfield 2013), taken here to be relatively synonymous.[7] Far from being idiosyncratic and ad hoc in nature, composite utterances exhibit a multitude of semiotic complexities that motivate the conventionalized patterns speakers use to express themselves. Findings from the three conceptual domains featured, namely ASPECT, CONTRAST, and DISCOURSE NAVIGATION, strongly suggest a high degree of conventionality in both the speech signal and the co-speech behavior.

---

7 Enfield's *composite utterance* refers to synthesis across even more semiotic channels: "Their composite nature is widely varied in kind: they may consist of a word combined with other words, a string of words combined with an intonation contour, a diagram combined with a caption, an icon combined with another icon, a spoken utterance combined with a hand gesture" (Enfield 2013: 689).

In describing the oral and embodied behaviour of a wide range of utterances across a range of conceptual domains, this dissertation quantitatively captures both the convention and variation inherent in the language phenomena at hand. Neither the synthetic and idiomatic nature of multimodal constructions nor the variation that they feature should relegate these structures to the unanalyzed periphery of language. Rather, this dissertation makes the case that they, too, are linguistic phenomena that contribute to the linguistic signal and are thus worthy of analysis and theorizing. Moving beyond text-based analyses to prioritize oralcy and interaction yields a more robust understanding of the way language works and of the importance of including these phenomena in language analysis.



Chapter 2 | Methods of multimodal corpus analysis

2.1 Introduction

One of the primary challenges of working with natural discourse lies in its messy complexity and high degree of variability. In fact, much of the messiness is what formalist approaches such as generativist linguistics proposed should be excluded from the study of language:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogenous speech community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interest, and errors (random or characteristic) in applying his knowledge of the language in actual performance.
>
> (Chomsky 1965: 3).

As a result of the prevalence of the tradition of studying idealized language, until recently, most of the research on conversation has taken place outside the bounds of the discipline of linguistics (Clark 1996; Sidnell 2010; Enfield 2013, 2017). While the field of conversation analysis takes a micro-sociologist approach to the study of natural interaction, there has been little focus within linguistics on the "seemingly messy back-and-forth of conversation" (Enfield 2017: 6). Of course, this messy back-and forth includes the type of co-speech behaviour that is under investigation in this dissertation.

One challenge for the linguist interested in natural interaction lies in the collection and annotation of large amounts of conversational data. The challenge is multiplied many times over for multimodal data, in which both the speech stream and the kinesic signal are fodder for detailed analysis. With the technological advances of recent years, the availability of multimodal data has increased. This has removed one barrier for linguists seeking to move beyond monomodal data, which is usually just a textual rendering of an invented and therefore decontextualized sentence. The data sourced for this dissertation stem from the Red Hen, a massive archive of programming from broadcast television, which I describe in more detail below. I collected video samples of specific utterances by searching for text strings in the closed-captioning aligned with the television programs. I then applied corpus methods to annotate and systematically analyze the data, which includes both features of the linguistic utterances and the co-speech behaviour that typify specific domains of language use. The linguistic phenomena I investigate range from grammatical constructions that mark aspect (Chapter 3), to expressions that mark contrast (Chapter 4), and stanced utterances that function as markers of discourse junctures (Chapter 5). In this chapter, I describe the corpora (Red Hen and the Corpus of Contemporary American English [COCA], which I describe below) and the methods for the collection and annotation of the data that apply to all the studies presented here. Methods specific to each study are provided in their respective chapters.

2.2 Corpora

Two sources of data were used for this dissertation. To gather multimodal data, I used the archive and facilities of the Distributed Little Red Hen Lab (hereafter Red Hen), an online multimedia archive co-directed by Francis Steen and Mark Turner (Steen & Turner 2013). Red

Hen is an archive – a vast collection of data, rather than a corpus. A corpus is a tagged and sampled collection of (usually monomodal) text. For my purposes, using the closed caption transcriptions in Red Hen, I created a series of mini-corpora containing the target utterances for each study, which I then tagged for a series of features. When I wished to examine larger samples, I used the Corpus of Contemporary American English (COCA). COCA is a tagged corpus that features a distribution of genres of text and (transcribed) speech. It is tagged for part-of-speech and features tools to perform collocation and other types of searches. In what follows, I describe the Red Hen archive and the COCA corpus in more detail.

2.2.1 Red Hen

The multimodal data for this dissertation were collected using Red Hen. The core dataset for Red Hen is the NewsScape Library of International Television News, which contains digitized television news programs collected from cable and broadcast sources from 2005 to the present. At present, Red Hen consists of over 400,000 hours of audiovisual data from public broadcasts and grows at approximately 150 hours of programming daily. The 400,000 hours is aligned with circa five billion words of closed-captioned text that is produced by the broadcast networks and integrated into Red Hen alongside the videos. The closed captioning provides the text for all verbal utterances, rendering Red Hen searchable as a text archive. (See Appendix B2 for a screenshot of the search interface).

While the bulk of the archive focuses on varieties of English (mainly from the U.S., but also U.K. and Indian programming), the archive now captures 18 other languages. Red Hen contains a wide variety of genres, from news broadcasts to talk shows, comedy, and advertisements. Some of these are scripted (e.g. sitcoms), while others feature spontaneous, unscripted discourse (e.g. talk shows). For all the studies presented in this dissertation, only spontaneous and interactional contexts were included; scripted shows, prepared news reports, political speeches, and other non-interactional genres were excluded. The data were collected from TV programs broadcast between November 2012 and November 2018. Where applicable, date spans within that range are given for each case study.

In using the Red Hen archive, I am analyzing conversational data that has been produced and recorded for television broadcast. This raises important questions regarding the distance between what might be considered completely naturalistic interaction, i.e. speaking to a friend or spouse in your kitchen, and speaking to another person in front of a studio audience or even directly to a television audience. The questions, then, are: Are these data representative of naturalistic, spoken interaction? Would the speaker make the same utterance in the same way if she were in a less public environment?

There may well be effects of the made-for-television genre on the language use of the speakers as represented in the Red Hen archive. For example, talk show hosts and broadcast personalities are accustomed to appearing on television. They may adapt their behaviour to suit their on-air persona and motivations, such as a need to be compelling, persuasive, funny, etc. Some speakers, such as politicians, may even receive training with regard to message delivery, including gesture use, for example. These potential genre effects cannot be swept aside when using broadcast television as primary data; however, there are several factors that lead me to conclude that the findings in the studies presented here *can* be applied beyond the genre of broadcast television. I would argue that the findings can tell us something about language behaviour more generally for several reasons. Firstly, many of the speakers who speak in the broadcasted interactions are not celebrities, but rather private citizens. They are the guests on the shows, rather than the hosts. In these cases, it would seem that the likelihood of an effect of training, for example, would be substantially reduced. Of course, they may still be presenting themselves differently than they would in their kitchens. They may be self-conscious or adopt mannerisms they would not normally display; however, I would suggest that these genre effects also exist between, say, a kitchen conversation and a schoolyard or professional interaction. These genre effects are an inherent part of the variability of language in its natural environment, conversation.

In considering the impact of using programs recorded for television as primary data, it is important to consider the potential limitations of the data in light of the specific issues I attempt to shed light on here. I am specifically interested in the degree to which certain verbal expressions are aligned with patterns of co-speech behaviour, which contribute to an utterance's multimodal profile. From this perspective, it is most important to determine the degree of enactment of certain phrases as compared to other phrases. In this dissertation, I hold the genre and type of interaction relatively constant to compare behaviour across utterances. Moreover, given the way most of co-speech behaviour flies 'under the radar' of a speaker as they are producing it, while effects on speakers that result from being on TV may be to amplify or subdue enactments, I don't believe these genre effects ultimately change the linguistic, semiotic, cognitive, and embodied processes at play.

Finally, one of the chief advantages of the Red Hen archive is that it allows linguists and other researchers access to video and transcribed audio of actual samples of spontaneous language use; however, the disadvantage of such an archive is that researchers have little to no demographic information about the speakers and no ability to manipulate the environment. It is impossible to survey the speakers regarding any training they received, the possible impact of the camera, or even to assess this objectively by comparing the same speaker in conversation on television versus in their living room.

While acknowledging these constraints, I would maintain that the advantages of Red Hen and video samples from broadcast media outweigh the disadvantages. The fact is that there is no other large archive or corpus of spontaneously produced multimodal data – even of a world language such as English – that could facilitate this research. The disadvantages that are related to each individual instance in the data are mitigated by taking a quantitative and statistical approach in which I establish profiles based on 20-50 independent tokens of the same utterance. I have also attempted to reduce the impact of the genre by removing contexts that are less reflective of a conversational, naturally spoken discourse style. I therefore do not include genres such as news reports delivered in a newsroom. I provide more detail on the search constraints in §2.3 below. Before outlining the methods of data collection, I introduce another corpus that I used for reference and comparison purposes, namely the Corpus of Contemporary American English.

2.2.2 Corpus of Contemporary American English

In order to achieve a broader sense of the constructional patterns of the target utterances I investigate for each domain, I used the Corpus of Contemporary American English (COCA; Davies 2008)[8]. COCA is the largest genre-balanced corpus of American English. It consists of language from spoken conversation, fiction, magazine, newspaper and academic genres of American English. In order to most closely reflect the spoken English genre of Red Hen data, when comparing COCA and Red Hen data, I limited the searches to the spoken portion of COCA ($COCA_{sp}$). The spoken sub-corpus consists of transcribed telephone calls and

[8] https://www.english-corpora.org/coca/

transcripts of unscripted conversation on TV and radio programs from 1990 to 2017 and, thus, in large measure is comparable to the language sampled in Red Hen.

2.3 Data collection

2.3.1 Target utterance searches

For each of the linguistic and conceptual domains examined in this dissertation, my goal has been to investigate a range of conventionalized patterns of linguistic and kinesic behaviour that inhere to that domain. To operationalize each study, I selected a set of target utterances in each domain. The selection of target utterances will be explained in subsequent chapters. Here, I explain how I captured data in Red Hen featuring these target utterances.

There are several search engines available within Red Hen. The data for all parts of the dissertation were accessed online using the Red Hen's Communications Archive Edge Search Engine. The search interface is shown in Figure 2.1. The search functions I used most frequently were the topmost two: 'with all of the words/phrases' and 'at least one word/phrase'. For most searches, I used the first search box. The second search function allowed me to include more than one target item for target utterances that had multiple forms. I used this, for example, in Chapter 4 to search for *on one hand* and it's very near neighbor *on (the) one hand* at the same time, as shown in the figure. Similarly, for the study on discourse markers in Chapter 5, I used the 'at least' function to find search returns for both *but anyway* and *but anyways*, with an 's'.

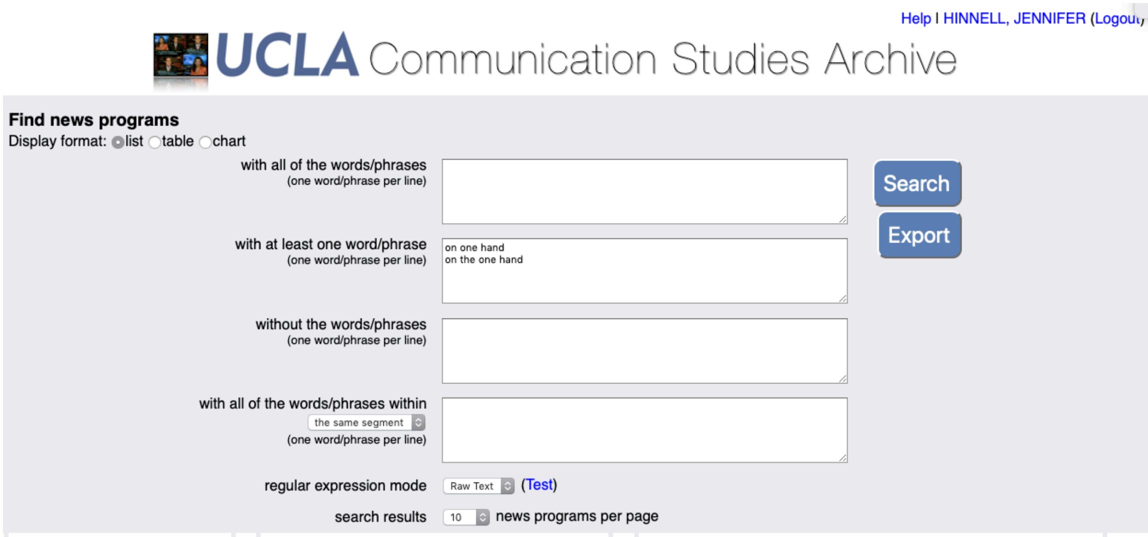


Figure 2.1 Screenshot of Red Hen search interface

I also manipulated other parameters in the search interface. Firstly, I set the date and time to retrieve the most recent data possible. The date range therefore differs for each set of studies depending on when the study was started. The time span, i.e. how far back I searched, varies based on the frequency of the target utterance in Red Hen. For lower frequency utterances, a wider date range was required to retrieve an adequate number of search returns. I also adjusted the networks and series to optimize the occurrence of search returns that were from interactional settings. This created a more efficient process in the annotation phase, with far fewer tokens having to be discarded due to not meeting the criteria of stemming from an interactional scenario (outlined below in §2.3.2). If the search was too restrictive (i.e. I didn't select enough programs), there was an insufficient number of search returns. Conversely, if I included all programs and networks, the data that were returned were overwhelmingly from scripted shows or news programs. The list of programs and networks that I used for all studies is included in Appendix B.

2.3.2 Search returns

Red Hen returns tokens for a queried string chronologically starting with the most recent. Figure 2.2 shows a search return screen for the fixed expression *how shall I put it*.

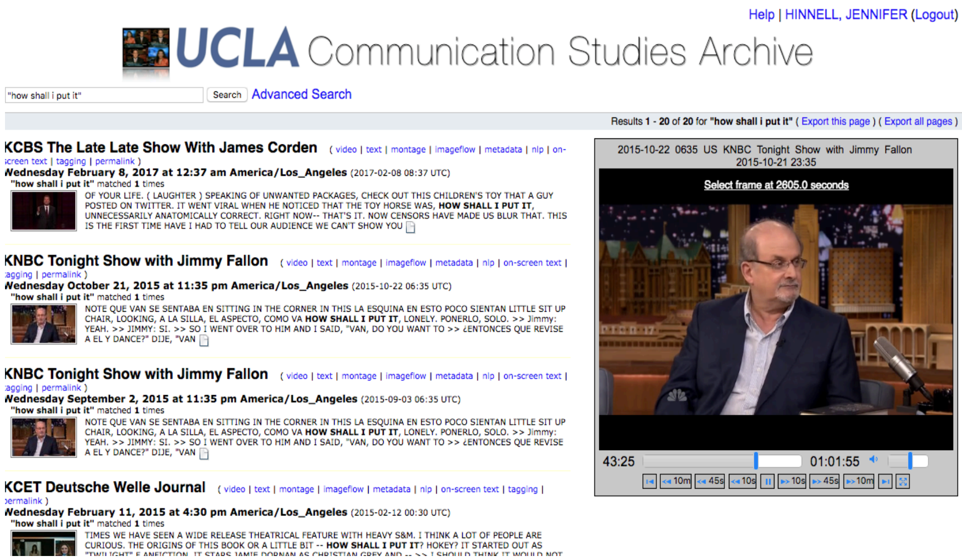


Figure 2.2. Screenshot of Red Hen search return page

For each target utterance, the search returns were first viewed and processed to determine which ones constituted valid data points. To be included in the study, the token had to stem from a context that was interactional and unscripted and in which the speaker's hands and upper body were visible. A studio audience was considered to meet the criteria for interactional, but I excluded news anchors in a newsroom based on the impression that this type of delivery was more scripted. Scripted or pre-recorded interactions such as sitcoms or advertisements were excluded. Depending on the requirements of each case study, between 20 and 50 interactional, visible tokens of each target utterance were collected. Once the required number of tokens was reached, the selection formed the target utterance data set, which was then annotated. I describe annotation methods in §2.4. First, I discuss some challenges in working with Red Hen and how I addressed them.

One of the mechanical issues of working with Red Hen is that there is no automatic way to filter out what I had determined were invalid data points for my purposes. In many cases, I had to view 4-5 times the number of clips in order to obtain the minimum desired number of tokens of an utterance. For example, for the case study of *should I/shouldn't I* presented in Chapter 4, Red Hen captured 174 search returns over 69 programs, yet this resulted in only 17 valid data points. These figures demonstrate the high number of 'hits' in Red Hen that are required to obtain valid data. Repetition of broadcasts is one cause of this issue. When an utterance occurs in public discourse, such as a U.S. presidential news conference, it is captured and broadcast multiple times by different networks, which results in seemingly unique tokens that are in fact repetitions of the same data point. Frequently, a search return appears in an advertisement, in which the advertisement appeared multiple times in the search returns. I found ways to work around these issues and remove some of the duplication in the search returns. For example, in the study of *quit Verb-ing* presented in Chapter 3, *quit* frequently collocated with the complement *smoking*. This meant that ads designed to help people quit smoking overwhelmed the other verbal complements in the search returns. Given that advertisements do not meet the criteria of being unscripted and interactional, I needed to omit all of these. To operationalize this, in the Red Hen search interface field 'without the words/phrases' (shown in Figure 2.1, above) I entered the brand name of the company generating the ad, whose name appeared in the transcript. This blocked the ad from being included in the search returns and reduced the noise in the search return list for this particular target utterance. Most searches did not offer such an easy way to sort through repetitions and I viewed each clip and culled irrelevant ones manually. Other frequent causes for omitting a data point include issues with the camera angle. In many cases, the camera angle meant the articulators were not clearly visible. In others, the camera was not on the speaker at the moment of the utterance. Finally, as mentioned above, manipulating the programs and networks searched also reduced the number of search returns that needed to be discarded as non-interactional or scripted speech.

Figure 2.3 exemplifies another difficulty in working with TV data, namely the running text across the bottom of the screen that characterizes network television programming, known as a *chyron*. The instance shown in Figure 2.3 stems from the study of the fixed expression *on (the) one hand*. While there is a head tilt that co-occurs with the target utterance, this data point was excluded as the chyron obscures the speaker's hands.

Figure 2.3. Data point excluded due to chyron

In order to track the valid vs. invalid data points, the search return pages for each search were exported to a comma-separated-values (csv) file in Excel using the Red Hen export function. (This function is visible at the top right above the video window in Figure 2.2). It exports the following metadata to Excel: filename, closed-captioning/transcript, timestamp, a unique identifier, and a permanent link, which people with access to Red Hen can use to locate the program for a given clip. This .csv file formed the basis for the dataframe I developed for each target utterance. I then added columns for further organizational purposes and the annotations. In a column created to track valid search returns, valid returns were marked 1, invalid were marked 0. When I reached approximately 1.5 times the number of search returns I was seeking to analyze, I sorted the Excel sheet to list the valid returns and discarded all those tokens in the dataframe that were not interactional and visible. The extra tokens were kept to ensure that I had enough data points to overcome any that needed to be removed at a later point due to not meeting criteria, or duplication, for example. At this point, I was ready to download and annotate the valid search returns for the target utterance.

2.3.3 Downloading clips from Red Hen

Red Hen facilitates viewing the video in the web interface, but it is not possible to slow down or stop the video replay, nor to make annotations within the interface (though this functionality is in development). For Chapter 2, I downloaded all clips. For Chapters 4 and 5, for which I did not measure gesture timing, I viewed clips in the Red Hen interface, downloading only those needed for presentation purposes or for viewing at slower speeds. For video downloads, I created a command line script to download a 30-second segment of each video, approximately 15 seconds on either side of the target utterance. To download one or two videos at a time, I also used Movavi Screen Capture[9] software to capture the videos from my screen, as this proved faster than using the batch download method at the command line.

2.4 Annotations

After acquiring a data set of 20-50 tokens for each target utterance, I annotated each return for linguistic and kinesic variables. I then analyzed the annotations using quantitative and statistical methods, which yielded a multimodal profile for each utterance type. The resulting

[9] https://www.movavi.com/screen-recorder-mac

profile captures both the prototypical form and frequent variations on this form for each utterance type. I return to a fuller discussion of the notion of multimodal profile to close this chapter in §2.5. First, I describe the variables I used to capture the verbal and kinesic signal for each set of utterances.

2.4.1 Degree of enactment

The starting point for kinesic annotation was a binary measure of co-occurrence, which I call *degree of enactment*. The degree of enactment is the rate of co-occurrence of any co-speech behaviour with a particular utterance. As introduced in the first chapter, the measure of co-occurrence of speech and gesture has been taken – not uncontroversially – as an indication of the degree of entrenchment of a construction (Lanwer 2017; Zima & Bergs 2017a, 2017b). In my annotations, for each token in which the audiovisual context met the criteria of being interactional, unscripted, and with articulators visible, the first pass captured whether there was co-occurrence of a signal in the body at the time of the target utterance. Co-occurrence was coded as positive if there was a change in state, either from rest to movement in an articulator, or from one movement to another movement. Co-occurrence was more complex if the speaker was already in motion in the previous utterance. In these cases, I noted a co-occurrence with the target utterance if there was a change in body movement related to the target utterance. Thus, if a speaker was continuously exercising a beat gesture (a vertical downward gesture stroke produced in time with intonation peaks) or a head shake, for example, for the full duration of a discourse segment, the target utterance would be assessed as not having co-enacted body movement. If there was no change in state, i.e. the gesturer maintained a preceding gesture form as a rhythmic marker, or if there was no correlating gesture, co-occurrence was coded as negative. An example of a pre-existing gesture form is shown in Figure 2.4. The palm forward bent finger form in Figure 2.4a spans over the target utterance *on the one hand* in Figure 2.4b as well. Since there is no change in gesture aligned with the utterance *on the one hand,* this would also have been coded as negative for degree of enactment. (Here, and throughout the dissertation, below each screenshot I provide a transcription of the speech (S) and of the co-speech behaviour (G, for gesture).)

(a)

S: *Very interestingly, let's take a moment. The pope seems to have given a mixed message there.*

G: left hand palm forward, bent fingers, across body

(b)

On the one hand, talking about the crimes of sexual abuse…

no change, gesture held from utterance in (a)

Figure 2.4. Example of pre-existing gesture handshape

Co-speech behaviour is defined as prosodically, semantically, and temporally aligned with an utterance. In this dissertation, I exclude gestures that are only prosodically related (such as the beat gestures described above) and focus on those that are semantically and temporally related to an utterance. This begs the question of how one knows that a gesture is related and aligned in these ways. The issue is complex. Gesture that is semantically related can precede speech by a significant (in gesture time) amount of time. For example, the cyclic gestures associated with *continue* in the aspect study presented in Chapter 3 average an onset prior to the speech signal of 387 milliseconds, while the *quit* gestures were nearly simultaneous with the utterance. Given the variation in acceptable onset times for gestures that are clearly semantically aligned with an utterance, I relied on general observation of time alignment rather than using a fixed onset window. While the semantic and time alignment is difficult to define, humans are excellent at agreeing on what co-speech behaviour is aligned to specific utterances in natural discourse. In the inter-rater agreement that was performed for Chapter 3 and which met a criterion of 86%, none of the disagreements were a result of timing.

In this dissertation, I refer to the co-occurrence variable as the *degree of enactment*. I settled on this term after publishing Chapter 3 (Hinnell 2018), in which I used the term *degree of co-occurrence*. Other terms that I have used and considered since beginning this work include *degree of attraction* and *degree of embodiment*. I have since settled on *enactment* as a more suitable term. Prior to 2013, there was very little precedence for this type of measure in linguistically oriented studies of manual gesture. Hofstetter and Alibali (2008) use *co-occurrence* to capture the rate at which gesture occurs with imagistic speech. Woodin subsequently adopted my term (Woodin 2018; Woodin et al. 2019) . The terms *attraction* and *embodiment* were seen to evoke *a priori* theoretical assumptions. *Attraction* suggested to some scholars that, from a speech processing perspective, speech comes first and drives the gesture production. I considered the literal interpretation term *embodiment* ('to express in the body') to capture my intended meaning quite precisely, but for many in the field it evokes a wide array of fundamental assumptions in cognitive science that I do not intend to engage with this co-occurrence variable.[10]

In the remaining sections in this chapter, I describe the annotation schemas for kinesic form (§2.4.2) and linguistic form (§2.4.3) that formed the basis of my analyses.

[10] Thanks to colleagues at MaMuD IV (2017) and the 2018 gathering of the Berkeley Linguistic Society, in particular Mandana Seyfeddinipur, for engaged discussion of the clearest term for this variable.

2.4.2 Kinesic form annotations

For all chapters, the starting point for kinesic form annotation after establishing the degree of enactment was the annotation of gesture variables. For the studies of CONTRAST and STANCE presented in Chapters 4 and 5, respectively, I also annotated upper body movement, including head movement, shoulder movements, eyebrow movements, and torso shifts.

Gesture annotations

Manual gesture has dominated the debate about what constitutes meaningful co-speech body movement and there has been much attention paid to the description of the highly expressive capacity of the hands (McNeill 1992; Kendon 2004). Kendon's widely accepted typology of a gesture unit introduced in Chapter 1 includes five distinct phases: rest, preparation, stroke, hold, and recovery. The stroke is considered to be the meaning-carrying element of the gesture. The rest and preparation phases precede the stroke and there can be multiple preparation/stroke sequences prior to a final hold and recovery, which is a return to a rest or neutral position. The form features described in this section relate predominantly to the stroke phase of a gesture, which is considered to be the meaning-carrying element of the entire excursion. The other phases are essentially kinesic requirements for the execution of the stroke and are not relevant to the research questions I pose.

The original basis for the gesture annotation system I have developed was Bressem's (Bressem 2013) notation system for form features in manual gesture. The notation system is based on the four parameters of sign language: hand shape, orientation, movement, and position in gesture space.[11] As Bressem states:

> The [notation] system is grounded in a linguistic semiotic approach to gestures, assuming a heuristic separation of form, meaning, and function in the analytical process […]. Accordingly, the present system differs from others existing systems in three essential aspects: (i) It concentrates solely on a form description of gestures. (ii) It proposes a form description independent of speech. (iii) It avoids gestural form descriptions including paraphrases of meaning.
>
> (Bressem 2013: 1080)

In the course of observing gestures and beginning annotations, it was clear that there were variables I wanted to capture that were not part of Bressem's notation system. I therefore adopted Bressem's primary parameters of hand shape, orientation, movement, and position in gesture space and captured further parameters of my own. The full set of parameters that I have annotated for are given in Table 2.1.

[11] The system was developed in the context of research at the Berlin Gesture Centre and that center's project *Towards a Grammar of Gesture* (TAGOG), an approach to gesture studies based on form features (Müller, Bressem, et al. 2013).

Table 2.1. Manual gesture annotation parameters by domain

Parameters	Levels	ASPECT (Ch.3)	CONTRAST (Ch.4)	DISCOURSE (Ch.5)
Spatio-temporal				
Gesture space	center, left of body, right of body	✓	✓	✓
Symmetry	symmetrical, asymmetrical		✓	
Laterality	lateral: right-left, left-right, centre-left, left-centre, centre-right, right-centre not lateral: right-right, left-left, centre-centre		✓	
Onset asynchrony	Gesture onset prior to target utterance onset in msec	✓		
Movement				
Type	straight, arced, circle, none/static	✓		
Axis	vertical, lateral, sagittal, static (e.g. wrist turn only)	✓		
Direction	up, down, left, right, diagonal, towards body, away from body, none/static	✓		
Action phases	number of separate articulations within a gesture stroke	✓		
Hand shape				
No. of hands	left, right, both	✓	✓	✓
Palm orientation	palm-up (PU), palm-down (PD), palm-lateral (PL), palm-vertical (PV), palm-diagonal (di)		✓	✓
Palm orientation in gesture space	towards-center (TC), away-center (AC), towards-body (TB), away-from-body (AB)		✓	✓
Recurrent form	container, palm-down, PUOH, holding away		✓	✓

In the next subsections, I describe the parameters listed in Table 2.1 in the following groupings: spatio-temporal parameters, movement parameters, and hand shape. Given that the annotation of the spatio-temporal parameters of symmetry, laterality, and onset asynchrony depend on some of the movement and hand shape parameters, I describe these three variables at the end of the section rather than in the order listed in Table 2.1.

Gesture space

To annotate gesture space, I used McNeill's (1992) model of gesture space, shown in Figure 2.5. The model represents an idealization of a speaker-gesturer's position and, as such, has been criticized for being static (i.e. assuming the speaker's torso remains stationary) and only considering gestures from the speaker's perspective (Priesters 2012). For the studies presented here, many speakers in the data set were standing, or seated in a different configuration than the model (e.g. at an angle with legs crossed on a couch). These differences notwithstanding, the general positioning of the gesture relative to the speaker's body was assessed as shown in Figure 2.5. For all three studies, only center, right, and left were annotated.

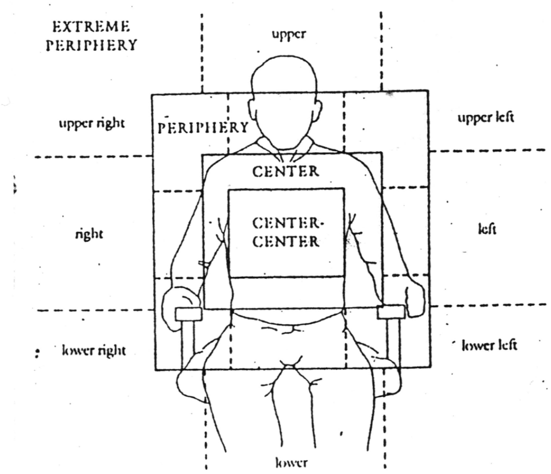


Figure 2.5. Model of gesture space (McNeill 1992: 89)

Movement

Three principal movement variables characterize gesture, namely, movement type, axis, and direction. Together, these are said to form the *movement trace*, which is the contour of a gesture in gesture space (Mittelberg 2008). To annotate the movement trace of the aspect-marking gestures in Chapter 3, I used the playback feature in ELAN to slow the gesture down, often to 50% of the original speed. I annotated for type and direction of movement (omitting Bressem's 'quality of movement' parameter, which qualitatively captures size, speed, and flow of movement). Movement direction consisted of the axis and the direction of movement along that axis. As shown in Figure 2.6, axis was coded as moving along one of three axes: the vertical, horizontal, or sagittal axes. The diagram on the left shows the vertical axis (movement up or down) and the lateral axis (right or left), while the diagram on the right shows the sagittal axis (movement away from or towards the body).

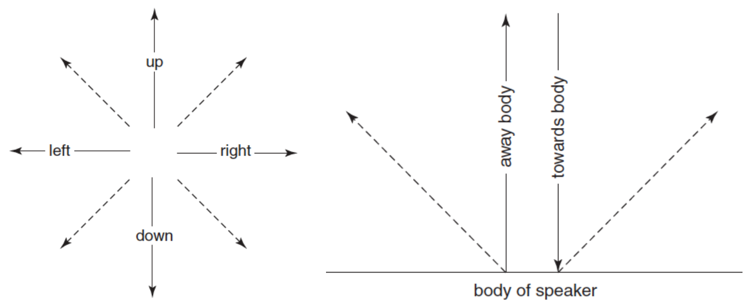


Figure 2.6. Gesture movement annotation: Axis and direction (Bressem 2013: 1090)

The second movement parameter, movement type, was coded according to the images in Figure 2.7; however, these distinctions proved to be idealized and difficult to implement. For example, repeated arc forms can be difficult to distinguish from spiral gestures. Therefore, as noted in Chapter 3, both spiral and arc gestures were coded as arc in the ASPECT study, with spiral forms coded as iterated arc gestures (the number of iterations was captured in the *action phase* variable described below). Movement type was only coded for the ASPECT study. Given the types of gesture associated with aspect, zigzag and s-line movement types did not occur in the data. This left straight, arced, and circle as potential movement types.

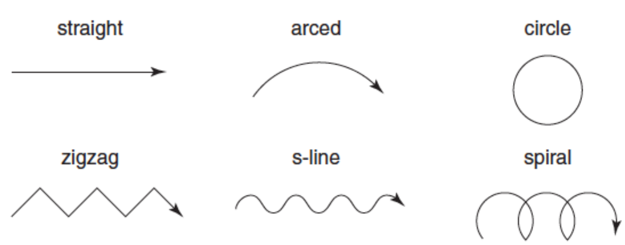


Figure 2.7. Gesture movement annotation: Movement type (Bressem 2013: 1088)

When there was no appreciable displacement in space, for example, only a change in handshape, the axis and direction were coded as static. This occurred, for instance, when there was a change in handshape only, e.g. from a neutral shape to a palm-up open-hand shape in roughly the same spatial position.

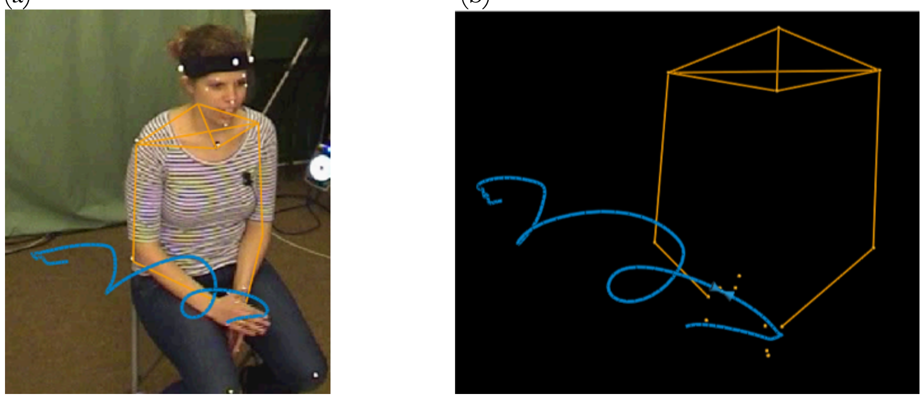
What I have termed *action phase* has not been captured in previous annotation schemas. Ladewig (2011) annotates for the number of rotations in a cyclic gesture; however, these and other studies seem to capture this only for curved movement forms, i.e. spiral, circle, and arc. Instead, I wanted to capture segmentation and repetition which can occur across all movement types. For this I used the term *action phase*. The term refers to the number of separate articulations of a movement type within a stroke. For example, in a cyclic gesture, if the gesturer performs the cycle only once, this would count as one action phase. If the cycle is repeated three times, that constitutes one stroke divided into three action phases. I found that the action phase variable was an important way of capturing aspectual information within a gesture phrase.

Unlike recurrent gesture forms, which are frequently held for a brief time at the end of a stroke, movement traces are very fleeting. 3D motion capture technology (MoCap) is an excellent tool for visualizing and ultimately transcribing gesture movement traces.[12] Figures 2.6 and 2.7 show motion captures of a speaker in the process of gesturing. The images in (a) shows the speaker in the experimental setup with the movement trace mapped onto the speaker, while the images in (b) show an abstraction of both the speaker (in orange) and the movement (in blue).

[12] Motion capture (MoCap) data stems from an experiment I ran with Dr. Irene Mittelberg, Natural Media Lab, RWTH Aachen University, investigating native English speaker gestures that mark aspect (in prep).

The movement trace proved to be an important part of the multimodal profile for the aspect-marking constructions investigated in Chapter 3, while handshape proved more important for the forms the occurred in contrast- and discourse-marking data presented in Chapters 4 and 5. I turn now to a description of hand shape annotations.

(a) (b)

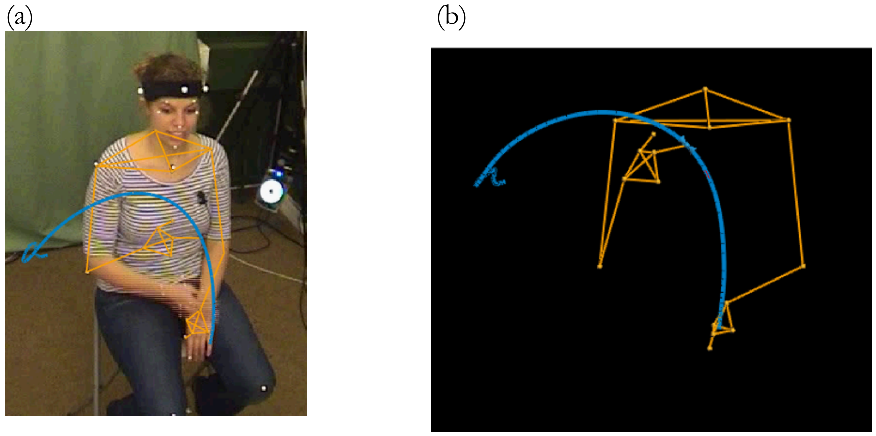


S: *I do run*

G: spiral movement trace

Figure 2.8. Spiral movement trace as captured by MoCap

(a) (b)



S: *if I were to continue watching*
G: arc movement trace

Figure 2.9. Arc movement trace as captured by MoCap

Hand shape

The LASG does not formally contain a separate variable for capturing whether a gesture is performed by one or two hands; however, it does note that the parameters must be assigned for each hand if they are both involved. For all the studies presented in this dissertation, gesture was annotated as being performed by one hand or both hands, and for one-handed gestures, I also captured which hand was involved.

The remaining parameters for hand shape in my annotation schema are based on the LASG annotation for palm orientation, which describes hand orientation in terms of the orientation of the palm itself and how the palm is oriented in gesture space related to the speaker's body. Palm orientation refers to the angle of the palm, namely, upwards, downwards, lateral, vertical, or diagonal. These palm orientations can be articulated differentially in the gesture space. For example, a lateral palm orientation could be expressed towards centre, towards the body, or, rarely, away from the body (perhaps referentially to signal the location of two exits to either side of a speaker). A vertical palm orientation is generally expressed in pragmatic gestures as away from the body, although it could also be expressed towards the body or towards centre. These latter two orientations are most likely in a referential gesture, e.g. someone might display their hands in a vertical orientation facing towards them to indicate a mirror in front of them. A palm-up gesture has no further notations for its placement in gesture space.[13]

Hand shape is frequently associated with certain recurring gestures. After coding the palm orientation (for the CONTRAST and DISCOURSE chapters only), I therefore also coded any recurrent hand forms. These forms included the palm-up open-hand gesture (PUOH)

[13] Bressem (2013) provides several more form parameters for the annotation of hand shape that I found were not relevant to the studies presented here. These include basic categories of hand configuration (fist, flat hand, single fingers, combination of fingers), the shape of digits (stretched or bent fingers), the character of movement (reduced or enlarged), the speed of motion (decelerated or accelerated), and the flow of movement (accentuated ending).

(Kendon 2004; Müller 2004), palm-down gesture, container gesture (Mittelberg 2014), and holding away gesture (Bressem & Müller 2014). I describe these common recurrent forms here.

In a PUOH gesture, the palm faces upwards in a flat or slightly curved configuration and it is often introduced with a wrist turn rather than directional movement (Müller 2004). Figure 2.10a and b show a one-handed and two-handed PUOH gesture, respectively.

(a) (b)

S: *We started talking to this woman who is walking her dog…*
G: PUOH, right hand

S: *But the big thing is it stops, it stops moving for 48 hours ('it' = a storm)*
G: PUOH, both hands

Figure 2.10. One-handed and two-handed PUOH gesture forms

Two other frequent forms that I encountered in my searchers were the container and holding away gestures. A container gesture is characterized as one in which the speaker-gesturer depicts the holding of a manipulable object, taken to iconically represent a physical object or to metaphorically represent an abstract object or concept. Figure 2.11a shows one variation of a container gesture with palms lateral towards-centre. Lastly, a conventionalized gesture known as the holding away gesture features one or two hands in a vertical palm orientation with palms facing away from the body, as shown in Figure 2.11b.

(a) (b)

(a) S: *It is in one way very exciting that this is finally a matter of…*

G: container form

(b) S: *but anyways*

G: holding away

Figure 2.11. Recurrent gesture forms

Symmetry

For the chapter on CONTRAST, I was specifically interested in possible symmetries in hand shape and movement. Therefore, I made use of the variable *symmetry* to capture whether the same handshape was apparent for a pair of target utterances. For example, if *on (the) one hand* and *on the other hand* or *on (the) one hand* and *but at the same time* were paired over the course of a discourse segment, the symmetry variable captured whether each utterance was co-articulated with the same gesture (although possibly in different gesture spaces). If the speaker enacted a PUOH for both utterances, this was coded as symmetrical, whereas a sequence in which the handshapes were different or one part of the utterance featured a handshape and the other featured no enactment were coded as asymmetrical.

Laterality

For the multimodal profiles of CONTRAST markers (Chapter 4), pilot data suggested that gestures of CONTRAST frequently used lateral space, whether through the use of a right or left hand or through placing objects in left and right gesture space. CONTRAST is a domain in which two objects or concepts are being evaluated and thus can convey positive or negative valency. Some studies have suggested that speakers gesture with their dominant hand when assessing something positively and their non-dominant hand when assessing something more negatively (Casasanto & Jasmin 2010; Casasanto & Henetz 2012). While it is not possible from Red Hen data to know the handedness of speakers, I used a combination of several variables to capture preferences for laterality. These variables included which hand (and side) was used first, direction of movement, and use of gesture space. For example, if a PUOH gesture (described under handshape) was performed first with the right hand and then with the left hand, laterality was coded as 'right-left' (R-L). This was applied for direction of movement as well.

Onset asynchrony

Lastly, the parameter I termed *onset asynchrony*, which was measured for the aspect study only, refers to the difference between the onset of the gesture phrase and the onset of the target utterance. I measured this in ELAN, video annotation software which allows the user to slow down video playback. I marked the onset of the gesture as compared to the onset of the target utterance at slow speed and exported the data to Excel to subtract the utterance onset from the gesture onset, which gave the onset asynchrony value. The target utterance was considered to begin directly at the onset of the auxiliary verb (e.g. the inflected or infinitival form of *continue, keep, start, stop,* and *quit*), regardless of upstream predicates in the verb phrase. Since the gesture phrase generally preceded the onset of the target utterance in the speech signal, the asynchrony on average is a negative value. For example, if the gesture phrase began at 18.60 seconds into the clip and the target utterance at 19.00 seconds, the asynchrony value was calculated as 18.60-19.00, or -400 milliseconds.

In this section, I have previewed the gesture annotation schema used in each chapter. Within each chapter, I describe in more detail the annotations most relevant to each set of case studies. In the next section, I describe the annotation system for movements of the upper body, namely, of the head, shoulder, eyebrows, and torso.

Upper body annotations

For the CONTRAST and DISCOURSE expressions presented in Chapters 4 and 5, respectively, I also annotated upper body movement. There is no widely-accepted schema for the annotation of co-speech behaviour in the upper body. Paul Ekman's early work on non-verbal behaviour (Ekman & Friesen 1981; Ekman 1993) focused mainly on facial expression tied to the expression of emotion, rather than on how it may or may not be conventionalized as part of the linguistic signal. The purpose of my annotation scheme was to capture the contribution of each upper body articulator to the kinesic signal, i.e. to assess its participation in the multimodal profile. I thus took a form-based approach similar to Bressem's annotation system for manual gesture to develop a schema for annotating the upper body. For each modality (head, shoulder, brows, torso), I captured form features based on each articulator's kinesic affordances – at least those used in co-speech behaviour. For example, while I may be able to move my head at unusual angles for the purposes of stretching, such actions are rarely part of intentional, and, in the case of co-speech behaviour, meaningful movement. Thus, for head movement, I used common terms to capture how the head usually moves along each axis. A nod is recognized when the head moves up and down; a tilt is a rotation of the head in which the ear moves towards the shoulder on one side; a turn is a rotation in which the chin turns towards the shoulder; and a shake is a repeated turning motion from side-to-side. Similarly, shoulder movement that occurs in co-speech contexts usually involves an up (and then down) motion of the shoulders. (One rarely sees shoulders waggle forward and back in co-speech contexts, despite the kinesic ability to move in this way.)

The annotation schema used for the upper body is shown in Table 2.2. Symmetry and laterality were coded according to the definition given above, transposed onto upper body articulators. For example, if the full expression *on (the) one hand/on the other hand* featured a head tilt on the first part of the utterance and no movement on the second part of the utterance, this was annotated as asymmetrical but lateral (a head tilt is by definition a lateral use of space). A head tilt on each side was annotated as symmetrical and lateral.

Table 2.2. Upper body annotation by domain

Modality	**Variable**	**Levels**	**CONTRAST (Ch.4)**	**DISCOURSE (Ch.5)**
Head	Movement type	Nod, tilt, turn, shake	✓	✓
Shoulders	Direction	Up, down	✓	✓
Eyebrows	Direction	Raise, lower	✓	✓
Torso	Shift	Turn/twist, lean + direction	✓	✓
All	Partitioning	G (gesture only), G&UB (gesture and upper body), UB (upper body only)	✓	✓

Partitioning

In order to capture the ways in which the hands and upper body form part of a composite utterances (a term introduced in Chapter 1 to signify the multiple sources of information or multiple modalities that are part of an utterance), I coarsely grouped the participation of the hands versus the upper body using a *partitioning* variable. I use this term to assess the participation of each set of articulators in an enacted utterance. I group the head, shoulder, eye, and torso movement under the heading upper body (UB). The levels of this variable can thus be gesture only (G), gesture and upper body (G&UB), or upper body only (UB). If an utterance cooccurs with movement of one, some, or all of the upper body articulators, partitioning is coded as UB. If the utterance consists of a manual gesture at the same time, it is coded gesture and upper body (G&UB).

Upper body movements such as shoulder shrugs and head movements have been shown to be strongly recruited in the expression of more subjective, stanced utterances (Schoonjans et al. 2013; Schoonjans 2014; Hinnell & Rice 2016; Debras 2017; Rice & Hinnell 2017; Hinnell & Rice 2019). Thus, particularly in the more stanced expressions examined in the chapters on CONTRAST and DISCOURSE markers (Chapters 4 and 5), the partitioning variable forms an important part of the multimodal profile. It allows a coarse-grained analysis of the involvement of manual gesture vs. the upper body in the enactment of a multimodal construction. Given that only manual gesture was annotated in the aspect chapter (Chapter 3), partitioning is not part of that study.

I initially called this variable *body partitioning*, a term borrowed from and inspired by Paul Dudis' work on the division of labour between different body parts in the expression of viewpoint in American Sign Language (Dudis 2004; Wulf & Dudis 2005). Dudis' use of the term specifically captures how different articulators can simultaneously serve different functions. For example, the face of a signer can show the viewpoint of an observer towards an event that is being enacted in the hands of the signer. I use the term somewhat differently. Rather than emphasizing the differential contribution of the partitionable 'zones' of the hands vs. the face, I am grouping the body articulators involved in co-speech behaviour by upper body (i.e. shoulder, head, facial and torso movement) vs. lower body (hands). To reflect the difference from Dudis' term, I have settled simply on the term *partitioning* to capture my meaning.

To close this section on kinesic annotations, I now introduce the notion of multimodal profile.

Multimodal profiles

As I have described in this section on kinesic form annotations, I capture specific aspects of the kinesic signal by coding a set of parameters for manual gesture and upper body movements. These kinesic movements often occur in concert as part of a compound enactment (e.g. the shrug, as described in Chapter 1). I therefore also take the results of the individual aspects of the annotations (e.g. gesture trace and hand shape, head movement, etc.) and examine them together as a composite signal. In this way, the annotations and quantitative and statistical analyses provide a *multimodal profile* for each target utterance. The multimodal profile for an utterance type captures the degree of enactment, the most frequent and/or statistically significant form features that emerge in the gesture and upper body annotations, and the body partitioning pattern for that utterance. The multimodal profile provides an aggregate, or prototype, form for each utterance; however, there are many forms that feature for an utterance type that lie outside this prototype. For example, though the most frequent hand shape for an utterance may be a palm-up form, in other tokens of the same utterance, a palm-down or palm-lateral gesture may be used. In cases in which two different forms occur to the same degree, I included both in the final multimodal profile. To capture conventionalization as well as variation for a given construction, in reporting the findings for each utterance, I describe some of the principal form variants alongside the prototype forms.

To complete this section on annotation methods, I now introduce the annotation scheme for the linguistic signal.

2.4.3 Linguistic form annotations

Corpus methodologies have allowed linguists to investigate the impact of contextual features on constructions. Concordancers that are built into corpus engines usually have a keyword in context (KWIC) function and the corpora often have a range of metadata associated with each search return that pertains to general usage facts, such as modality, genre or sub-corpus, and, at times, speaker demographics. The field has thus gained significant insights into the idiosyncrasies of meaning and form and of variation within constructions. For example, usage patterns vary across genres. Particular collocations or constructions may be widely used in spoken registers but far less frequently in written forms or vice versa. Usage patterns can also vary based on other types of contexts, e.g. by inflectional categories such as tense, aspect, or mood (TAM) or person of the subject. For example, in a corpus study of subject marking in the British National Corpus, Rice and Newman (2005) showed that the verb *think* occurs in the first-person singular between 65% and 93% of the time, depending on TAM inflection (cf. Rice and Newman 2018). By annotating a subset of linguistic contextual features as described in the remainder of this section, I could explore the specific contributions of linguistic material. For each chapter, I annotated variables in the surrounding linguistic context. In addition to information about the subject (e.g. person and number), I also captured whether the subject of each side of the [O1H-OOH] construction was non-human, as in (3), or human, as in (4) (both examples are from Red Hen).

(3) *On one hand, <u>the value</u> of their investments is falling. On the other, <u>there are indicators</u> that might suggest our economy's doing just fine.*[14]

[14] Red Hen: 2015-08-26_0130_US_KNBC_NBC_Nightly_News,58

(4) *On the one hand his team is doing everything they can to make sure he's in place if he decides to run. But on the other hand, it sure seems to me that he has truly not made up his mind whether he's going to get it.*[15]

As each domain featured target utterances characterized by different morphosyntactic patterns, the linguistic variables that were annotated also varied. The aspect-marking expressions investigated in Chapter 3 were the most highly variable in their morphosyntactic context. Thus, I annotated a fuller set of morphosyntactic variables related to the verb phrases. Given the semantic-pragmatic context of evaluation that characterized the study of contrast in Chapter 4, I also annotated semantic prosody to encode any discernible positive or negative valuation towards the proposition. For example, expressions such as *benefit, amazing, exciting,* and *I like it* were coded as positive, while phrases such as *worrying* and *curse* were coded as negative. An overview of the full annotation schema is provided in Table 2.3.

[15] Red Hen: 2015-09-20_1500_US_KABC_This_Week,3145

Table 2.3. Linguistic annotations by domain

Variable	Levels	ASPECT (Ch.3)	CONTRAST (Ch.4)	DISCOURSE (Ch.5)
KWIC	key word in context	✓	✓	✓
Subject		✓	✓	✓
Person	1st, 2nd, 3rd	✓	✓	✓
Number	singular, plural	✓	✓	✓
Animacy	human, non-human	✓	✓	
Identification	name or descriptor of speaker	✓	✓	✓
Verb				
Role in VP	inflected verb, past participle, imperative, future, infinitival complement, modal complement, do-support complement	✓		
Tense	past, present future	✓		
Mode	realis, irrealis	✓		
Semantic prosody	positive, negative		✓	

In the study of ASPECT presented in Chapter 3, I annotated linguistic variables for all data points, i.e. the full data set of 250 tokens, regardless of whether the tokens were enacted multimodally or not. In the CONTRAST and DISCOURSE chapters (Chapters 4 and 5), I annotated the linguistic context only for those tokens that included co-speech behaviour. Since the degree of enactment ranged from 70% and 90% and only 20-25 examples of each utterance were collected, there were too few examples that were *not* enacted to undertake any meaningful analysis for the tokens that did not feature co-speech behaviour. Lastly, in addition to linguistic information, I captured the name (if it was known or given) or other identifying information of the speaker in order to track the possible over-representation of one speaker within a dataset, which I defined as a speaker occurring more than twice.

2.5 Citing multimodal examples

All multimodal examples are given in figures that are numbered by chapter and figure number within the chapter (e.g. Figure 3.2 is the second screenshot in Chapter 3). The caption includes the key phrase and a brief description of salient features in the form. In each figure, the speech transcript is given in *italics* in the first row directly below the image, with the label S: (for speech). The target utterance in the transcript is indicated in the screenshot by underlined text. A description of the co-speech behaviour is given in the second row of text, with the label G: (for gesture). Figures are numbered consecutively throughout the dissertation. In cases in which the progression of the co-speech behaviour is shown over a sequence of screenshots within one figure, each screenshot is listed alphabetically, e.g. (a), (b), etc. An example is given in Figure 2.12.

(a) (b)

(a) S: *Apparently, they like him on both sides. On the one hand he believes in climate change so they consider him a Liberal.*

G: right hand PUOH

(b) *On the other hand, he doesn't believe in abortion so they consider him a Republican…*

left hand PUOH

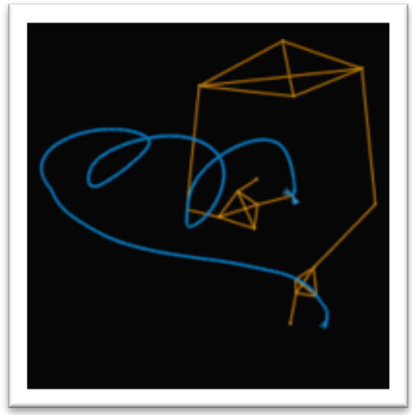
Figure 2.12. *on the one hand/on the other hand*, PUOH gesture

For each screenshot from Red Hen in this dissertation, I have made the original video available (as .mp4 or .mov file) in an online repository. Please see Appendix A1 for instructions on viewing the video clips. The repository also includes a spreadsheet containing all the source metadata, as outlined in Appendix A2. Each entry in the metadata file includes the date and time of the broadcast, name of the program, network, the timestamp (i.e. how many seconds into the program the instance occurred), and the permalink that allows anyone who has access to Red Hen to view the videos within the Red Hen archive.

Examples of speech utterances listed as numbered items are sourced from COCA or Red Hen. The metadata for each utterance is given in a footnote (e.g. Red Hen: 2015-09-20_1500_US_KABC_This_Week,3145).

2.6 Summary

In this chapter, I introduced the source of primary data for the case studies presented in this dissertation, the Red Hen archive, as well as the COCA_{sp} corpus, which I use for comparison. After discussing the assumptions and implications of using data from broadcast television as a proxy for natural interaction, I described in detail the kinesic and linguistic variables that form the basis of my annotations. These include a measure of co-occurrence of speech and co-speech behaviour for each utterance, its *degree of enactment*. Finally, I introduced the notion of *multimodal profile*. I use this term to describe the composite signal, or prototype form, of a construction. In the following three chapters, I investigate constructions belonging to the domains of ASPECT, CONTRAST, and DISCOURSE NAVIGATION and build multimodal profiles for each construction.



Chapter 3 | *Keep on talking* — and gesturing: The multimodal marking of ASPECT

3.1 Introduction

As discussed in the introductory chapter, co-speech gesture has been shown to be related to abstract levels of language processing (Matlock et al. 2012; Huette et al. 2014). One such abstract domain is at the centre of the study presented here, namely the linguistic-conceptual domain of ASPECT[16]. Broadly defined, linguistic aspect refers to "different ways of viewing the internal temporal constituency of a situation" (Comrie 1976: 3; Chung & Timberlake 1985; Talmy 2000b; Croft 2012). Cross-linguistically, aspect can be manifested grammatically or lexically. Grammatical aspect is marked through inflectional or periphrastic distinctions (e.g. the *-ing* morpheme to mark progressive aspect in English), while lexical aspect reflects the inherent aspect of a situation (e.g. the state of being Canadian is naturally an enduring one). The multimodal study presented here examines aspect-marking verb constructions in North

[16] I use small caps (ASPECT), to refer to the broad conceptual domain of aspect as I use it in this dissertation. The convention in CL literature is to use regular case to refer to grammatical and lexical aspect more specifically.

American English. It contributes to the current discussion within cognitive and usage-based linguistics, as well as in the field of gesture studies, that the study of linguistic meaning should be based on language as it occurs in everyday interactive contexts and both linguistic and kinesic (including prosodic) patterns should be included in linguistic analysis.

In this chapter, I present a case study of five periphrastic auxiliary constructions in North American English and the manual gestures that are co-produced with them. Each construction comprises an aspect-marking auxiliary, i.e. a lexical verb that has been co-opted to function as a 'helping' verb, in this case to mark aspectual force on the main verb, which is the complement of the auxiliary verb. The auxiliaries – CONTINUE, KEEP, START, STOP, AND QUIT[17] – are followed by a complement verb, as in (5) to (9) from the Red Hen multimodal archive used for this study:

(5) *Things start <u>going</u> good in life and they don't think they deserve it.*[18]
(6) *My parents never said 'Stop <u>making up</u> stories'.*[19]
(7) *I quit <u>doing</u> Botox about five years ago.*[20]
(8) *… to find out whether or not it continues <u>to grow</u>.*[21]
(9) *The reason I kept <u>getting married</u> was I wanted to feel like I was normal.*[22]

I investigate the extent to which each [AUX VERB] construction invites co-speech embodiment and then to what extent these embodiments are conventionalized. Principal

[17] In keeping with common practice in corpus linguistics, I use small caps to indicate the lemmatized form of the verb, i.e. KEEP subsumes the inflected forms *keeps, keeping, kept.*

[18] Red Hen: 2013-08-29_1700_US_KABC_The_View,2198

[19] Red Hen: 2013-08-13_1700_US_KABC_The_View,730

[20] Red Hen: 2012-08-21_0635_US_KCBS_Late_Show_with_Dave_Letterman,1829

[21] Red Hen: 2013-06-03_1200_US_FOX-News_Fox_and_Friends,1676

[22] Red Hen: 2013-08-12_1700_US_KABC_The_View,1171

research questions are: (1) Is the aspectual nature of the auxiliary verb expressed in co-speech gesture? (2) If so, in what way? That is, in which parameters of the gesture profile can aspectual construal of the event be observed? (3) A number of form parameters have not traditionally been part of gesture form annotations, such as gesture co-occurrence, gesture/speech asynchrony, and segmentations of the gesture phrase. To what extent do these parameters also capture the marking of aspectual information? The study takes as its data source natural, unscripted interactions, and uses corpus methods to analyze 250 search returns of [AUX VERB] constructions (50 for each of the five aspectualized auxiliaries).

3.1.1 Aspect

The constructions investigated here are headed by aspect-marking auxiliaries. As per the preliminary definition given above, aspect presents different ways of viewing the internal temporal constituency of a situation (Comrie 1976: 3). That is, it expresses whether an event is complete or ongoing, punctual or extended in duration, and continuous or repeated, among other distinctions. Crucially, aspect allows a speaker to convey different temporal construals of the same event (Comrie 1976; Frawley 1992; Croft 2012). For example, both of the following sentences depict the reading of a story in the past tense. Note, however, that in (10), the reading of the story is bounded and presumably complete, whereas in (11) the past progressive signals that the event continued for a while and may have been interrupted or abandoned before the story was finished.

(10) *Charlotte read a story.*
(11) *Charlotte was reading a story.*

English is a language known to have impoverished grammatical marking of aspect, i.e. it marks very few aspectual distinctions morphologically (Comrie 1976: 1). Its aspect-marking morphemes are limited to the *-ing* verbal suffix signaling progressive and the *-ed* suffix signaling past tense. Instead, English recruits a wide range of grammatical and lexical constructions to convey aspectual information, from adverbial phrases (*again and again, over and over, at once*) and directional particles (*he ate it up, she was singing away, they drove on*; Rice and Newman 2004), to periphrastic verb constructions (e.g. *John started talking*), such as those featured in this study. The auxiliary verbs in these constructions have moved along an 'auxiliation path' from their initial lexical meanings towards a grammaticized aspectual meaning that partially echoes the semantics of the original lexical root (Kuteva 2001; Heine & Kuteva 2002). That is, as auxiliaries, they influence the resulting construction, bringing their intrinsic Aktionsart or inherent aspect to the new V-V construction to allow for a re-construal of the aspectual structure of the original event. For example, the verb *sneeze* is inherently punctual and bounded, having a definite on-set and finish, and refers to a single event, as in (12). The progressive in (13) can give a reading of Mackenzie being in the middle of either a single sneeze, as if a photographer caught her in mid-sneeze, or of a multiple-sneeze event; however, only the multiple event reading is available in (14).

(12) *Mackenzie sneezed.*
(13) *Mackenzie was sneezing.*
(14) *Mackenzie kept sneezing.*

Thus, in (14), the inherent aspect of the auxiliary *keep,* which means 'to hold or retain' in its basic, lexical sense, imbues the *keep sneezing* construction with a stronger, iterative meaning (Talmy 2000b; Heine & Kuteva 2002: 19).

Finally, aspect has been categorized into a three-way taxonomy: open, closed, and phase (Chung & Timberlake 1985; Brinton 1988; Frawley 1992; Talmy 2000b).[23] The distinction between open and phase aspect forms one of the organizing principles for this study. Open aspects include all those that extend an event over a timeframe, namely the imperfective, durative, progressive, and habitual. These aspects capture events that are ongoing or happens routinely; they focus on the internal time of the event and are unbounded (e.g. *Claire lives in Europe*). In the present study, CONTINUE and KEEP constitute open aspect auxiliaries. By contrast, closed aspect events are viewed as complete, unitized events, and are bounded.[24] (There are no closed aspect constructions featured in this study.) Phase aspect completes the classification of aspects by capturing how events change status inside or outside the time frame, by their beginnings, middles, and ends (Frawley 1992: 328). Phase aspects includes inceptive aspect (*She started to talk)*, prospective aspect (*They are about to leave),* and terminative aspect *(She quit smoking).* In this study, the auxiliary START denotes inceptive aspect, while STOP and QUIT denote terminative aspect.

Open aspect focuses on the internal structure of an event. This raises the important point that there are different *ways* for an event to be open. An open aspect event can be structured as one continuous event, in which case an iterative reading means a repetition of the whole event. An open event can, alternatively, develop in phases, or incrementally. Talmy's term *plexity* (Talmy 2000b) is helpful in capturing this potential ambiguity in open

[23] Brinton (1988) uses imperfective and perfective to denote open and closed aspect respectively.

[24] *Boundedness* can refer to a 'natural endpoint' in the case of an accomplishment or achievement, as in *write a letter*. In the case of grammatical aspect, and as used here, boundedness "refers to an action that is finished, whether it has a natural endpoint that has been reached, or simply terminates" (Croft 2012: 77).

aspect. Plexity captures a notion of quantity for both objects and for time. Should an object or event consist of only one component, it is considered *uniplex*, whereas an event or object that is complex in some way is referred to as *multiplex*. For example, in English the noun phrase *a bird* has a single referent, which can be 'multiplexed', or pluralized, by adding *-s* to the noun, as in *birds*. In the example in (12), above, the verb *sneeze* inherently predicates a single or 'uniplex' relation consisting of one sneeze event (i.e., a closed event that is bounded and highly punctual). This relation can be coerced into a multiplex interpretation by using it in the [KEEP VVG] construction (in which VVG is a standard corpus tag denoting the progressive inflected form), as in *She kept sneezing* (Talmy 2000b: 69). In the gesture annotations in this study, the variable that I label *action phases* (described in §3.2.3 below) correlates with plexity: gesture phrases that contain only one articulation in the stroke are considered uniplex gestures, while those containing two or more are considered to be multiplex.

In addition to aspect, another semantic category that plays a role in language structure Talmy's (2000a) notion of *force dynamics*, which captures how entities interact with respect to force. Force dynamics refers to the internal or external force that activates an event, for example the exertion of force, resistance to a force, overcoming a force, etc. Johnson (1987: 42) prompts us to consider that "our bodies are clusters of forces and that *every* event of which we are a part consists, minimally, of forces in interaction." Despite only beginning to receive attention in gesture research, force dynamics has been acknowledged to occupy "a central place in embodied approaches to meaning" (Mittelberg 2013b: 772) and, as such, needs to be considered in light of the target utterances of this study. Consider the minimal pairs given in (15) and (16). While the first sentence offers an objective construal of a ball rolling (the ball is unimpeded), in (16) the addition of the auxiliary *keep*/*keep on* invokes one of two possible force dynamic patterns: the ball has a tendency to roll but is being impeded (e.g. long grass), or the ball has a tendency to rest but is being propelled (e.g. by the wind). Thus, the *kept/kept on* connotes a countervailing force being overcome.

(15) *The ball was rolling along the green.*

(16) *The ball kept (on) rolling along the green.*

(Talmy 2000a: 412)

In this study, CONTINUE and KEEP differ in their force-dynamic profile in a similar way to these two examples, with CONTINUE being the more neutral of the pair, force-dynamically speaking, since there is no real expectation of countervailing pressure or contra-flow to stop the action, only persistence in time.

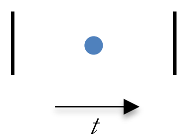
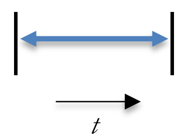
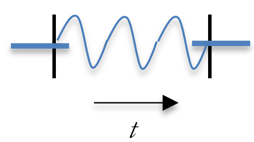
Although STOP and QUIT are not characterized by Talmy, I suggest that they can also be considered a force-dynamic minimal pair: QUIT signals a more forceful termination of the event than STOP and seems to be associated with a higher degree of intention (under what Talmy refers to as the psychological realm). For example, in (17) and (18), the first utterance suggests that Charlie will come back to his homework later, while in the second there is an inherent finality to the event – the homework will remain unfinished.

(17) *Charlie <u>stopped</u> doing his homework.*

(18) *Charlie <u>quit</u> doing his homework.*

In Table 3.1, I summarize the major types of aspect and force dynamics and their relation to the auxiliaries profiled in this study. In the image schema diagrams, the side bars on each diagram indicate the scope of predication. The bold dot in the case of closed aspect and the bold lines (straight or wavy) in the case of open and phase aspect mark the aspectualized event that is profiled within that scope of predication, for which time is passing (marked by the *t* in the diagram). Thus, for closed aspect, the dot in the schematic diagram signifies that the event takes place and its internal structure cannot be analyzed. For open aspect, the bolded straight horizontal line indicates that the focus is on the internal structure of the event. For phase aspect, the line crossing the boundary at the beginning or end of the event indicates that the focus is on the change of state either into or out of the event.

Table 3.1. Characteristics of types of aspect and their target auxiliaries

Category of aspect	Image Schema	Characteristics	Auxiliaries
CLOSED ASPECT perfective punctual telic semelfactive		complete, unitized, bounded	(not profiled in this study) *Charlie read the book.* *Mac ate her lunch.*
OPEN ASPECT imperfective habitual/durative progressive iterative/incremental		extension over time; modifiable, unbounded	CONTINUE, KEEP* *It <u>continued</u> to grow.* *She <u>kept</u> getting married.*
PHASE ASPECT inceptive terminative		change of status within or outside time frame of event; focus on endpoints: onset/ offset	START, STOP, QUIT* *Things <u>started</u> going well.* *<u>Stop</u> making up stories.* *I <u>quit</u> smoking.*

* Denotes the force-dynamically strong auxiliary in each set of verbs.

3.1.2 Iconicity in gesture

This chapter also aims to contribute to a broader understanding of the ways in which gesture use is iconic. Within gesture research, discussions of iconicity have largely involved references to hand shape and the use of gesture space, as in the studies of the PUOH gestures and the use of gesture space in the metaphorical expression of time and space described in the introduction. By contrast, in the analysis presented here, I hope to demonstrate that iconicity in gesture is represented in more abstract embodiments. An examination of a variety of quantitative and qualitative variables in combination with a larger data set ground this approach. By examining variables such as the difference in onset synchrony between speech and gesture, the number of segmentations in a gesture stroke (which I capture with the parameter action phases), and other parameters of movement such as axis, direction, and movement type, a gestural profile emerges for the aspect-marking constructions investigated here. As the multimodal findings show, some of the distinct aspectual information imparted in the event construal articulated in the speech stream is conveyed in the equally distinct gesture profiles of the five target catenative auxiliary constructions.

The study of conventionalization within gesture has focused extensively on the use of gesture form and gesture space to convey metaphorical meaning. Referential gestures (Müller 1998; Cienki 2005), which include McNeill's metaphorical gestures, "iconically represent and refer to a physical object or relationship in the source domain of the metaphorical concept" (Casasanto 2008a: 2). An example of referential gesture representing a physical object is the palm-up open-hand (PUOH) family of gestures (Müller 2004), characterized by a specific hand shape and orientation, namely palm open and turned upwards. The PUOH gesture is metaphorical in that it "presents an abstract, discursive object as an inspectable one" (Müller 2004: 233). The hand shape in this gesture form is key to this metaphorical mapping. Gesture movement type has also been found to play a role in other metaphorical mappings. For example, in Cienki's (1998) study of metaphorical mappings of spatial metaphors, he found that speakers accompany a mention of honesty or 'doing the right thing' with a straight line in the gesture phrase. Similarly, in their gestures, speakers mark up for morality and down for immorality, and up for a good grade and down for a bad grade (Cienki 1998). The metaphorical mapping of time in gesture space has shown the iconic use of gesture space in the embodied marking of time. Casasanto and Jasmin (2012) discuss the sagittal axis as predominant in English spoken metaphors, such as the deictic expressions of time *deadlines lie* *ahead of/behind us,* and *looking forward/looking back*. The lateral axis has also been shown to be important in the mapping of time such that time flows from left to right, generally represent time passing from past (left) to future (right) (Cooperrider & Núñez 2009; Núñez & Cooperrider 2013), despite the fact that very few expressions in English contain this metaphorical mapping. Gesture research is rife with discussions of iconic representations of a wide range of metaphorically construed concepts. In addition to the iconic representation of objects and the metaphorical mappings of space and time mentioned here, conventionalization in gesture has been studied in areas as wide-ranging as English grammatical terminology (Mittelberg 2006) and the embodied, metaphorical expression of balance in dance lessons (Müller & Ladewig 2013); however, the iconic representation of higher-order grammatical notions such as aspect have not yet been examined. Given the abstract nature of such conceptual categories, it is expected that the mapping of conceptual features onto gestural representation will include a wider range of variables (i.e. rather than focusing exclusively on hand shape or direction of movement as in the examples given here).

3.1.3 Studies of aspect in gesture

In Chapter 1, I introduced experimental studies that have explore the role of the body in cognition and language processing. In §1.2.2 on embodied cognition, I described several studies that suggested that both gaze and gesture patterns differentiate between different types of aspect (e.g. longer gesture strokes characterize event descriptions in progressive aspect). In this section, I return to this body of literature and specifically focus on studies that have explored how the linguistic-conceptual representation of aspect has been shown to affect co-speech channels of gesture and gaze.

In one of the few studies to include both grammatical and lexical aspect in its purview, Duncan investigated spoken and gestured expressions in Mandarin and English using an elicitation task to examine if differences in how aspect was encoded in the linguistic utterance were associated with differences in conceptual representation as operationalized by differences in co-speech gestures (Duncan 2002). The study found systematic, within-language co-variation of the choice of aspect and/or *Aktionsart* and the features of co-occurring gesture. For example, when speakers produced imperfective-marked speech, their gestures were longer in duration than for perfective aspect speech contexts, as if people were conceptualizing an ongoing event as enduring longer in time. Gestures accompanying imperfective verbs were also found to be more complex in nature, featuring longer and repeated movements (Duncan 2002: 183). These findings have been supported by other research, for example in the ACE experimental paradigm. For example, when speakers use sentences in the progressive aspect that describe hand motions (e.g. *Sally is closing the drawer)*, manual action in the same direction is facilitated; however, when they utter the same sentence but in the perfect aspect (*Sally has closed the drawer)*, the effect on manual actions is absent (Bergen & Wheeler 2010). Similarly, when people viewed videos of car accidents (Matlock et al. 2012), they gestured differently according to the aspectual force of the instructions they were given. Participants were asked what *had happened* (perfective framing) or what *was happening* (imperfective framing). Findings showed that "the form of aspect used in the question differentially influenced the way people conceptualized and described actions" (ibid.: 699). Aspect has also been shown to affect gestures accompanying event descriptions based on texts containing simple past or progressive marking (i.e. only grammatical aspect). In a task in which participants were presented with a series of texts featuring events in either the past progressive or past perfect, when recount the depictions in the text, participants' gestures were longer-lasting and more complex for events that were both recounted in the progressive aspect and had been presented in the original text in the progressive (Parrill et al. 2013).

Aspectual distinctions influence eye movement as well as gesture. As I described in Chapter 1, in one study, as participants listened to the storytelling of events in the simple past and in the progressive, patterns of eye-movement fixations differed (Huette et al. 2014). In a simple past description (e.g. *He went*), eye-movement fixations showed the greatest density in the central region of a screen and saccades were of a longer duration, as if participants were "staring at a static object or scene" (2014: 2). By contrast, fixations aligned with the past progressive showed a more diffuse distribution on the screen and shorter saccades, which the authors suggest reflect the dynamic nature of the event.

With the exception of Duncan (2002), most experimental studies of gesture and aspect restrict themselves to the distinction between the past progressive (*was driving)* and simple past (*drove*), typical English proxies for the nearly universal imperfective/perfective distinction. Furthermore, since auxiliation involves the grammaticalization of hitherto lexical verbs, it been characterized as a highly imagistic process, in which auxiliaries "reflect general conceptualization capacities crucially involving imaginative, or rather *imaging*, aspects of human cognition" (Kuteva 2001: 19). Yet, little is known about how auxiliary constructions with strong lexical aspect such as those featured here are gestured and there have been no other corpus-based studies of natural discourse and gesture use. This study aims to fill some of this gap in the literature.

3.2 Methods

The methods used in this study follow the template outlined in Chapter 2. As described there, I used the Red Hen archive to collect data. In this section, I describe the set of target utterances I searched for and how these were collected and annotated.

3.2.1 Target utterances

The auxiliaries targeted in this study occur in one of two constructions: the infinitival [AUX to VVI], as in *continue to go*, or the progressive [AUX VVG], as in *stop talking*.[25] Whereas in English, KEEP, STOP, and QUIT are available only in the progressive construction, CONTINUE and START are available in both constructions (i.e., *start talking* and *start to talk).* I used frequency of construction type in The Corpus of Contemporary American English (COCA) to determine which construction to use for the study. A search of the spoken portion of COCA ($COCA_{sp}$) revealed that for CONTINUE, the infinitival [CONTINUE to VVI] was most frequent, while for START, the progressive [START VVG] was most frequent.

Finally, there are other verbs within a similar semantic frame that have undergone the same path of grammaticalization from main verb to catenative auxiliary, such as *begin, finish,* and *carry on*. The target items included in this study were chosen over these near-synonymous counterparts due to their considerably higher overall frequency in the spoken register of the COCA corpus (e.g. spoken COCA featured 28,531 examples of [STOP VVG] and only 917 examples of [FINISH VVG]).

[25] VVG and VVI are standard tags used in the British National Corpus (BNC) and COCA: VVG is the progressive inflected form and VVI is the infinitive verb form.

3.2.2 Data collection

The data were collected from TV programs broadcast between September 1, 2010, and September 1, 2013. The search string consisted of a two-part verb phrase: all possible inflected forms of the auxiliary (e.g. *keep/keeps/kept*) plus the ten most frequent collocating verbs for that auxiliary construction according to COCA. Search returns in Red Hen are given in chronological order starting with the most recent. The search returns were first viewed and processed to determine which constituted valid data points. As described in Chapter 2, valid video clips were those that featured interactional, unscripted discourse in which the speaker's hands were visible and unencumbered (e.g. not holding a microphone). The first 50 valid clips formed the dataset for each auxiliary, yielding a total of 250 clips.

This set of 250 video clips was then viewed again for gesture co-occurrence. This measure of how frequently the expression is gestured speaks to the degree of conventionalization of gesture for each auxiliary. Gesture co-occurrence was coded as 'yes' if there was a new gesture at the time of the target utterance, and 'no' if there was no gesture, or if there was no change in gesture (i.e. the gesturer maintained a preceding gesture form as a rhythmic marker or maintained a hold of a previous gesture form). Of the 250 valid clips, 147 co-occurred with gesture. Annotations were completed by the author. A second coder annotated 20% of the dataset composed of the 50 clips for QUIT for gesture co-occurrence. Inter-rater reliability for gesture co-occurrence was 86% (i.e., there was disagreement on 7 clips, all of which were resolved by discussion).

3.2.3 Data annotation

The 147 instantiations of the target constructions that featured gesture, produced by 126 unique speakers, were downloaded from Red Hen for annotation in ELAN.[26] The approximate length per clip was 40 seconds, allowing for 15-20 seconds of context on either side of the target utterance. The variables and levels included in the gesture annotations are presented in Table 3.2. Most variables are based on Bressem's form-based annotation schema as presented in the Methods chapter (§2.4) (Bressem 2013). As described in Chapter 2, gesture movement patterns have been characterized by movement type, axis, and direction. Together these movement variables form the *movement trace* – the lines and contours of objects and abstract concepts that leave "imaginary traces in gesture space"' (Mittelberg 2010: 367). When there was no appreciable displacement in space (e.g. a wrist turn), the data point was coded as 'static'. Two variables – action phases and onset asynchrony – are not part of Bressem's schema and I defined them separately in Chapter 2. Here, I review the descriptions and give further examples to explicate the role of these two variables in aspect-marking gestures.

[26] ELAN is annotation software developed by the Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, that allows users to create, edit, visualize, and search time-aligned transcriptions of multiple annotation layers. https://tla.mpi.nl/tools/tla-tools/elan/

Table 3.2. Gesture annotation schema

Variable	Levels
Hands	left, right, both
Movement type	straight, arced, circle, static[27]
Movement axis	vertical, lateral, sagittal, static (e.g. wrist turn only)
Movement direction	up, down, left, right, towards body, away from body, static
Action phases	number of separate articulations within a gesture stroke
Timing	asynchrony of gesture onset vs. onset of target auxiliary in msec

Action phase refers to the number of separate segments within the gesture stroke. For a cyclic gesture this aligns with the number of rotations.[28] While straight gestures have been described as having 'singularity of form' and being one unit (Cienki 1998). As the data in this study show, there is no reason to assume that gestures that proceed with a straight movement type are likely to be executed in one distinct phase. Thus, action phase can also capture segmentation within a straight stroke. The gesture shown in Figure 3.1, for example, does not feature a smooth trajectory, but rather features a stroke that is articulated in a segmented, jerky fashion along the vertical axis.

[27] Bressem (2013) also features spiral as a movement type. Given that repeated arc forms can be difficult to distinguish from spiral gestures, both spiral and arc gestures were coded as arc in this study.

[28] Ladewig (2011) annotates for the number of rotations in a cyclic gesture; however, these and other studies seem to capture this only for curved movement forms, i.e. spiral, circle, and arc.

S: *Ever since then [fame] has just continued to grow and grow and grow*

G: Gesturer holds a metaphorical object and moves it in a straight trajectory upwards in five action phases

Figure 3.1. *continued to grow,* segmented gesture on vertical axis

Iterative and incremental development over time are two ways in which open (unbounded, progressive, durative) events can unfold. The *action phase* parameter captures repeated rotations of the arm within the same cyclic gesture stroke, indicating repetition or iterativity. It also, however, can capture a phased or articulated expression of a straight trajectory, as shown in Figure 3.1. This form is associated with incrementality. Iterative and incremental development over time are two ways in which open (unbounded, progressive, durative) events can unfold. The number of *action phases* is closely tied to the conceptual category of *plexity* described above. A singular circle gesture or a single-phase straight gesture would constitute uniplex events, while any gesture with an action phase greater than one was considered to be multiplex.

Lastly, to capture gesture timing, the target utterance was considered to begin directly at the onset of the auxiliary, regardless of upstream predicates in the verb phrase. For example, in (19), the onset was measured from *continue*, rather than from the beginning of the inflected verb phrase:

(19) *We're going to have to <u>continue to be smart</u> about the way we do business.*[29]

The beginning of the preparation phase in the gesture was used to calculate onset timing. The LASG relies on a frame-by-frame marking procedure based on Seyfedinnipur (2006) to establish the onset time of a gesture and the same procedure was followed here. The timing of the onset of both the gesture and the target utterance was recorded in milliseconds. The difference between the gesture and speech onsets gave the asynchrony, with a negative value indicating that gesture preceded the related utterance.

Inter-rater reliability was measured for gesture timing, number of action phases, and gesture type based on a subset of the full data set. I randomly selected 20% of the 147 videos which featured gesture (30 video clips) for a second coder. For gesture timing, the asynchrony (onset time of the gesture minus onset time of the auxiliary in speech) was measured and compared to the asynchrony value captured by the author for the same clip. For example, if the author found the onset asynchrony for a video clip was -0.135 and the coder -0.163, the difference was calculated as .028, or 28 milliseconds. The average difference in asynchrony values for the 30 clips was 180 milliseconds. For action phases (AP), the coders agreed in 25/30 clips. The author had an average AP value of 1.90 while the coder had an average value of 2.14 APs, a difference of 13%. For gesture type (arc, circle, straight, none/static), there were five disagreements that resulted in straight gestures in the original annotations being marked as arc gestures by the second coder. Once definitions were clarified, these instances were resolved. Four differences remained in the coding, yielding an inter-rater reliability of 87% for gesture type. Remaining disagreements were settled by consensus.

[29] Red Hen: 2013-08-11_0700_US_FOX-News_Huckabee,2858

To close this section, in Table 3.3, I summarize the predicted associations between the variables and the features of aspect introduced in this section.

Table 3.3. Predicted relationships between gesture variables and aspect type

Variable	Predictions
Movement type	*open aspect*: more cyclic gestures to reflect ongoing process *phase aspect*: more straight gestures to indicate event onset/offset
Axis/direction	*open aspect*: horizontal to reflect event unfolding over time *phase aspect*: vertical to reflect focus on event onset/offset
Action phases	*open aspect*: 2+ APs to reflect multiplex event (iterative or incremental); 1 AP to reflect uniform extension *phase aspect*: 1 AP to reflect uniplex onset/offset of event; no multiplex strokes as event extension is not available in phase
Gesture timing	*open aspect*: longer onset to reflect ongoing event *phase aspect*: shorter onset to reflect focus on event boundary

3.3 Findings

To give an impression of the speech and gesture patterns that characterize each auxiliary, I begin this section with a qualitative account of each construction, including syntactic and semantic characteristics and example of their multimodal enactments (§3.3.1). In §3.3.2, I provide quantitative results for each auxiliary and a statistical comparison of differences between auxiliaries by open and phase aspect.

3.3.1 Semantic overview by auxiliary

Open aspect auxiliaries

Continue

The dataset for the [CONTINUE to VVI] construction was characterized by sentences such as:

(20) *The big story continues to be North Korea.*[30]

(21) *I will work every single day to make sure that America continues to be the greatest nation on earth.*[31]

(22) *There is no money […] that is going to keep them in office that will allow them to continue to go against it.*[32]

(23) *I remember it for -- for the first one, and ever since then, it's just continued to grow and grow.*[33]

The most frequent complement verbs were *be* (with a token frequency of 16), *have* (7), *do* (6), and *grow* (4). Third person subjects were dominant (82%) and the vast majority (90%) of the utterances expressed abstract subjects, as in examples (20) and (21) here, and/or metaphorically-construed events, as in (22) and (23) (in which the 'it' referred to is fame). The [CONTINUE to VVI] construction was gestured in one of two prototypical forms: (a) with a stroke consisting of one action phase on either the horizontal or vertical axis, and (b) with a stroke consisting of multiple action phases on the vertical axis. Examples of these two prototypical forms are given in Figure 3.1, above, and Figure 3.2. below. In Figure 3.2, the panelist begins with both hands lax and palms facing inward toward the centre. The gesture moves outward on the horizontal axis in one fluid motion over the course of *continue to grow*. This is iconic of an event contour that is progressive and unfolds in a constant manner. The gesture expands evenly over its duration. By contrast, in the example for *continued to grow* in Figure 3.1 described above, the speaker is holding a metaphorical object (in this context the fame of a movie), and his hands ascend vertically in five distinct segments, depicting an incremental unfolding of the event.

[30] Red Hen: 2013-04-06_0635_US_KNBC_Tonight_Show_with_Jay_Leno,64

[31] Red Hen: 2013-08-16_2300_US_KNBC_The_Ellen_DeGeneres_Show,677

[32] Red Hen: 2013-06-26_2300_US_Current_Young_Turks_With_Cenk_Uygur,1359

[33] Red Hen: 2012-07-16_2300_US_KNBC_The_Ellen_DeGeneres_Show,688

S: *Who's going to make sure that what the mayor has invested in over the last few years will <u>continue to grow</u>?*

G: Palms face inwards in centre of body and move outwards at even rate

Figure 3.2. ***continue to grow*****, singular continuous gesture on horizontal axis**

CONTINUE also featured several examples with cyclic gestures. These cyclic movements consisted of multiple stroke segments, as exemplified in Figure 3.3. Here, the movement is a smooth circle trace that is iterated twice to indicate an uninterrupted event progression.

S: to find out - ah - whether or not it <u>continues to grow</u>

G: two full cyclic movements with right hand

Figure 3.3. ***continues to grow,*** **cyclic gesture**

Keep

The [KEEP VVG] construction was characterized by sentences such as (24) to (27). As we saw with [CONTINUE], the majority of [KEEP] utterances were abstract (in contrast to concrete), as in *heartbreaking* in (24), and metaphorical, as in (26) and (27). Frequent complement verbs were *say* (14 tokens), *go* (12 tokens), and *get* (10 tokens), and there was near equal distribution of first person (40%) and third person (38%) inflection, exemplified in the first two and latter two examples below, respectively.

(24) *I keep saying it's heartbreaking.*[34]

(25) *It's like god said, maybe this once I'm just going to keep going.*[35]

(26) *So when there is no winner, that jackpot keeps getting higher.*[36]

(27) *But the potential candidates just keep coming.*[37]

Gestures for [KEEP VVG] prototypically consisted of repeated straight strokes on the vertical axis and repeated cyclic gestures. In Figure 3.4 and Figure 3.5, the same speaker presents each of these forms over the span of the same utterance. In the first, the speaker produces a cyclic movement iterated four times for *keep coming back*. Here the rotation is counter-clockwise, which is unusual for a right-handed cyclic gesture. This appears related to the semantics of *coming back*; gestures indicating past time have been shown to be gestured leftward from the speaker's perspective on the speaker's horizontal access (Casasanto & Jasmin 2012). The speaker in Figure 3.5 then produces a multiplex gesture. The hand shape is a point, and this form is repeated four times downwards on the vertical axis aligned with *keeps getting paid*.

[34] Red Hen: 2013-08-08_2300_US_KNBC_The_Ellen_DeGeneres_Show,2610

[35] Red Hen: 2013-08-14_0635_US_KCBS_Late_Show_with_Dave_Letterman,2367

[36] Red Hen: 2013-08-07_1200_US_FOX-News_Fox_and_Friends,1608

[37] Red Hen: 2013-08-23_0300_US_ComedyCentral_Daily_Show,96

S: …*because if you keep coming back you keep getting paid.*

G: cyclic movement, counter-clockwise, iterated four times

Figure 3.4. *keep coming back*, iterated cyclic gesture

S: …*because if you keep coming back you keep getting paid.*

G: downward stroke with pointed hand shape iterated four times

Figure 3.5. *keep getting paid,* iterated downward stroke

While the vertical axis overall prefers downwards gestures, Figure 3.6 shows a gesture for [KEEP VVG] that moves upward on the vertical axis. The movement type was coded as spiral (arc) given the gesture's displacement in space. Both the direction and the incrementation – there are six action phases – iconically reference the metaphorical growth of the jackpot to a *higher and higher* number.

S: *So when there is no winner, that jackpot keeps getting higher and higher*

G: upward moving bimanual arc (spiral) gesture with six action phases

Figure 3.6. *keeps getting higher*, spiral movement on vertical axis

These examples show the variation within the gesture profile of the open aspect auxiliaries CONTINUE and KEEP in [CONTINUE to VVI] and [KEEP VVG] constructions, respectively. I now outline the semantic and gestural profiles for phase aspect auxiliaries.

Phase aspect auxiliaries

Start

The inceptive phase aspect of START imbues the [START VVG] construction with a conceptual focus on event onset. The utterances for [START VVG] (indeed, for all phase aspect constructions) tended to reflect concrete subjects and non-metaphorical events, as in (28), though abstract and metaphorical uses still occurred, as in (29). The most common complement verbs for the auxiliary START were *talk* (12), *get* (9), *go* (5) and *do* (5), and the construction preferred inflection in the third person (44%), which tended to be impersonal (e.g. *water, the KKK, blood, things*, etc.) rather than human subjects, as in (30). (1st person accounted for 32% of the START utterances and 2nd person for 24%).

(28) *The water started coming over the edge.*[38]

(29) *Everything just started creeping up on me.*[39]

[38] Red Hen: 2013-08-29_0635_US_KCBS_Late_Show_with_Dave_Letterman,2812

[39] Red Hen: 2013-08-30_1400_US_KNBC_Today_Show,1927

(30) *And then he <u>started talking</u> about my family.*[40]

(31) *I got up in his face and <u>started talking</u> to him.*[41]

Uniplex cyclic gestures, as well as uniplex straight gestures on the vertical and sagittal axes, characterize the gesture profile for [START VVG]. An example of a cyclic gesture for [START VVG] is shown in Figure 3.7, in which the speaker gestures in a circular motion, starting with a movement towards his body and then upwards in a circular motion and ending with a presentational palm shown in the image on the right.

S: *When the pope comes on Twitter […] and says 'I'm going to <u>start making blessings</u> on Twitter'….*

G: cyclic movement on sagittal axis; one action phase

Figure 3.7. *start making blessings,* single phase cyclic gesture

Stop and Quit

Sample utterances for the two remaining phase aspect constructions, [STOP VVG] and [QUIT VVG], both of which express terminative aspect, are given in in (32) to (39), below. The most common complement verbs for STOP were *do* (11), *talk* (9), *use* (6), and *work* (5), while for QUIT they were *smoke* (20), *drink* (8), *do* (6), and *use* (5). [QUIT VVG] is the only auxiliary construction to prefer the first person (42%): *quitting* appears to be something that one speaks

[40] Red Hen: 2013-08-29_0635_US_KCBS_Late_Show_with_Dave_Letterman,952

[41] Red Hen: 2013-08-02_1600_US_KABC_Live_With_Regis_and_Kelly,2552

of doing oneself. STOP, by comparison, is more often used in the third person, or as a command (second person imperative).

(32) *Nobody really wants to stop doing this.*[42]
(33) *If you own one, stop using it and contact them for a free repair.* [43]
(34) *Maybe you should just stop talking altogether …*[44]
(35) *500,000 people stopped looking for jobs.* [45]

(36) *I quit doing Botox about five years ago.*[46]
(37) *And you began with the first major decision, to quit drinking at the age of 40.* [47]
(38) *And so he's told the republican candidates to quit using the word capitalism.* [48]
(39) *Our loudest opponents on the left are never gonna like us so let's quit trying to curry favour with 'me.* [49]

[QUIT VVG] was the only auxiliary to prefer lateral strokes. The profile that emerges is of a two-handed, lateral gesture, exemplified in Figure 3.8: a sweeping sideward gesture made with both hands moving outwards in a flat hand shape. This form has been discussed in previous literature as the 2-palm-down-across and is characteristic of utterances which contain negative force (Harrison 2013). It features a relatively higher velocity (qualitatively observed) in its stroke and an abrupt end to the stroke phase. Importantly, in examples in which this bilateral stroke occurs with multiple action phases, the segments are incremental, i.e. the

[42] Red Hen: 2013-07-25_2300_US_Current_Young_Turks_With_Cenk_Uygur,2890
[43] Red Hen: 2013-07-24_1600_US_KNBC_Today_Show,1454
[44] Red Hen: 2013-08-01_0300_US_ComedyCentral_Daily_Show,803
[45] Red Hen: 2013-04-08_1000_US_FOX-News_Fox_and_Friends,768
[46] Red Hen: 2012-08-21_0635_US_KCBS_Late_Show_with_Dave_Letterman,1829
[47] Red Hen: 2011-03-08_2300_US_KABC_Oprah_Winfrey, 486
[48] Red Hen: 2011-12-07_0200_US_CNN_Piers_Morgan_Tonight,2020
[49] Red Hen: 2011-06-21_0100_US_FOX-News_Hannity,129

stroke is segmented into phases along one outward lateral trajectory, rather than the gesture stroke returning to its starting point for a second outward motion. The gestures produced with the [STOP VVG] construction also display this outward moving, bimanual gesture pattern, as shown in Figure 3.9, although the end of the stroke phase was observed to be less abrupt qualitatively in its execution.

S: *I lit a cigarette […] and put it out, like, I'm gonna quit smoking.*

G: bimanual outward gesture

Figure 3.8. *I'm gonna quit smoking*, bimanual, horizontal, outward gesture stroke

S: *Remember when I stopped talking to you for six months?*

G: bimanual outward gesture on lateral axis

Figure 3.9. *stopped talking,* bimanual, horizontal, outward gesture stroke

In this section, I have shown a range of contexts of use of the five auxiliary constructions and the gesture forms that accompany them. It is clear that some form features are related to the semantic representation of individual events, in addition to the aspectual construal of the event type. In the next section, I abstract away from these individual differences and present quantitative and statistical findings that reinforce the qualitative differences across the five aspectual auxiliaries described and illustrated in this section.

3.3.2 Quantitative findings

Quantitative findings are presented by aspect type, beginning with open aspect auxiliaries followed by phase aspect auxiliaries. Table 3.4 gives an overview of the results for each auxiliary. Open aspect verbs constitute the two columns on the left and phase aspect verbs the three rightmost columns.

Table 3.4. Summary of annotation results

	Open Aspect		**Phase Aspect**		
Variable	**[CONTINUE]**	**[KEEP]**	**[START]**	**[STOP]**	**[QUIT]**
Gesture co-occurrence absolute freq. (n=50)	37	29	25	27	29
relative frequency	74%	58%	50%	54%	58%
Gesture timing in msec mean	-386	-356	-269	-293	-171
SD	485	443	278	365	253
Action phases, mean	2.1	2.6	1.5	1.5	1.6
SD	1.3	1.5	.77	.64	1.1
Movement axis*	vertical (47%)	vertical (61%)	vertical (46%)/ sagittal (46%)	vertical (56%)	horizontal (52%)
Movement type*	straight (75%)	straight (64%)	straight (63%)	straight (88%)	straight (74%)
Hands (preferred hand and relative frequency)	1 hand (62%)	1 hand (52%)	1 hand (60%)	2 hands (52%)	1 hand (62%)

* Denotes most frequent. Relative frequency given in brackets.

Open aspect auxiliaries

Continue

With a gesture co-occurrence rate of 74%, the [CONTINUE to VVI] construction makes CONTINUE the most gestured auxiliary. Although a chi-squared test yielded no significant differences between gesture co-occurrence across auxiliaries, a Pearson residuals posthoc test does show the observed rate of gesture for CONTINUE exceeded the expected rate. CONTINUE also features the greatest mean asynchrony between the onset of the gesture phrase and the onset of the target utterance (-386 msec). Table 3.5 shows the distribution of movement type by axis, plexity, number of action phases and gesture timing for CONTINUE. The distribution of axis by movement type is significant (p<.05).[50] A post hoc test of Pearson residuals reveals that both straight gestures on the vertical axis and arced and circular gestures on the sagittal axis occurred more frequently than expected (these cells are noted in bold italics in the table).

Table 3.5. [CONTINUE to VVI] – Summary table (n=37)

	Axis			Plexity			APs	Asynchrony
	Lateral	**Sagittal**	**Vertical**	**Uniplex**	**Multiplex**			
Arced	2	***2***	0	2	2		1.8	-435 msec
Circle	0	***4***	1	0	5		3.8	-1.06 sec
Straight	5	6	***16***	13	14		1.9	-259 msec
Static	n/a	n/a	n/a	1	0		1	-230 msec
TOTAL	7	12	17	16	21	MEAN	2.1	-386 msec
						SD	1.3	485 msec

NB: bold italics indicate significant cells

While circle gestures are predominantly multiplex – indicating iteration, straight gestures are almost equally distributed between uniplex and multiplex strokes. These findings indicate a fairly even distribution of aspectual structure: straight uniplex gestures show a smooth (i.e. unsegmented) extension over time, such as illustrated in Figure 3.2 above, while the multiplexed straight gestures, as in Figure 3.1, provide additional aspectual detail, such as an incremental unfolding of the event.

[50] To avoid 0-cells in the chi-squared test, arced, and circle were reduced to a non-straight category here.

Keep

With a gesture co-occurrence rate of 58%, KEEP is less gestured than CONTINUE. The distribution of axis, plexity, mean action phase and onset timing by movement type are shown in Table 3.6. KEEP features an interaction of axis and movement type (chi-squared test, *p*<.005); a Pearson residuals posthoc test confirmed that straight movements on the vertical axis and arced movements on the lateral axis occur more frequently than expected.

Table 3.6. [KEEP VVG] – Summary table (n=29)

	Axis			Plexity			APs	Asynchrony
	Lateral	**Sagittal**	**Vertical**	**Uniplex**	**Multiplex**			
Arced	***2***	1	1	1	3		3.3	-217 msec
Circle	2	3	1	1	5		2.3	-679 msec
Straight	0	3	***16***	7	12		2.5	-283 msec
TOTAL	4	7	18	9	20	MEAN	2.6	-356 msec
						SD	1.5	443 msec

NB: bold italics indicate significant cells

As Table 3.6 shows, all gestures for KEEP display a strong preference for multiplex. This suggests that gestures for KEEP are more frequently used to add additional information about the internal aspectual structure – i.e. whether an event develops iteratively or incrementally – than simply depicting the unboundedness (or openness) of the event in a uniplex stroke.

Phase aspect auxiliaries

The three remaining catenative auxiliaries – START, STOP, and QUIT – are phase aspect auxiliaries; as such, they mark movement into or out of the time frame of the event (see the image schemas in Table 3.1 above).

Start

With a gesture co-occurrence rate of just 50%, [START VVG] has the lowest incidence of gesture use across the five auxiliary constructions. Table 3.7 shows the distribution of movement type by all variables. There is no significant interaction of movement type by axis; however, given the general preference across the other auxiliaries for vertical strokes, the relatively higher frequency of movement along the sagittal axis for START is worth noting. START also features a higher frequency of cyclic gestures (24%) than either of the open aspect auxiliaries.

Table 3.7. [START VVG] – Summary table (n=25)

	Axis			Plexity			APs	Asynchrony
	Lateral	**Sagittal**	**Vertical**	**Uniplex**	**Multiplex**			
Arced	1	2	0	2	1		1.3	-300 msec
Circle	0	4	2	2	4		2.2	-377 msec
Straight	1	6	9	13	4		1.3	-224 msec
TOTAL	2	12	11	16	9	MEAN	1.5	-269 msec
						SD	.77	278 msec

Stop

The terminative, phase aspect construction [STOP VVG] featured 27 search returns (54%) that were aligned with gesture. Results are summarized in Table 3.8. There is an interaction of axis by movement type (p<0.05); the Pearson residuals posthoc test shows that the cell-wise difference was in straight movements along the vertical axis, which again were dominant. Of the eight lateral movements, five of them feature both hands moving outward (as shown in Figure 3.9 above). Of note is that STOP is the only auxiliary not to feature any cyclic gestures.

Table 3.8. [STOP VVG] – Summary table (n=27)[51]

	Axis			Plexity			APs	Asynchrony
	Lateral	**Sagittal**	**Vertical**	**Uniplex**	**Multiplex**			
Arced	2	1	0	1	2		1.7	-73 msec
Circle	0	0	0	0	0		n/a	n/a
Straight	6	2	**14**	13	9		1.5	-332 msec
Static	n/a	n/a	n/a	2	0		1.0	-200 msec
TOTAL	8	3	14	16	11	MEAN	1.5	-293 msec
						SD	.64	364 msec

NB: bold indicates significant cells

Quit

QUIT features a gesture occurrence rate of 58%, making it the most gestured of the three phase aspect auxiliaries. As shown in Table 3.9, movement along the lateral axis is dominant, representing a marked difference from other auxiliaries and indicative of the profile shown in example Figure 3.8. QUIT also features a higher proportion of uniplex strokes than all other auxiliaries, at 72%. Furthermore, removing the outlier (a circular motion with six APs, which I return to in the discussion in §3.5) would result in QUIT having the lowest mean action phases per stroke, at 1.4 APs.

Table 3.9. [QUIT VVG] – Summary table (n=29)[52]

	Axis			Plexity			APs	Asynchrony
	Lateral	**Sagittal**	**Vertical**	**Uniplex**	**Multiplex**			
Arced	6	0	0	4	2		1.3	-150 msec
Circle	0	1	0	0	1		6.0	-300 msec
Straight	8	4	8	15	5		1.5	-332 msec
Static	n/a	n/a	n/a	2	0		1	-35 msec
TOTAL	14	5	8	21	8	MEAN	1.6	-171 msec
						SD	1.15	253 msec

[51] Two gestures had no value for the movement variable (e.g. wrist turn); therefore, the axis columns total 25 rather than 27.

[52] Two gestures featured no movement along an axis; thus, the axis columns total 27 rather than 29.

3.4 A statistical comparison across auxiliaries

I now offer statistical results across open and phase aspect auxiliary constructions. That is, taken together, I examine how the results for open aspect auxiliaries (CONTINUE and KEEP) compare with the results for phase aspect ones (START, STOP and QUIT). I also discuss other interactions in the data that require attention.

There were two significant effects by aspect type, as shown in Table 3.10. Interactions were determined by one-way ANOVA. Although there is no effect of gesture co-occurrence by aspect, there is a main effect of action phase by aspect ($F(1,144)=18.86$, $p<.0001$) and a main effect of gesture asynchrony by aspect ($F(1,145)=4.21$, $p<.05$). These findings indicate that open aspect gestures are reliably produced further in advance of their target utterances and feature more action phases than their phase aspect counterparts.

Table 3.10. Interactions by aspect type

Variable	Open aspect	Phase aspect
Action phases*	2.4 APs	1.5 APs
Asynchrony**	-373 msec	-243 msec
Gesture co-occurrence***	45%	55%

*$p<.0001$, **$p<.05$, ***not significant

There were no significant effects of axis and direction by aspect; however, when compared across auxiliaries rather than aspect type, there were significant correlations. A chi-squared test showed an effect of auxiliary on axis ($p<.001$) and a Pearson residuals posthoc determined that QUIT has a preference for the lateral axis and an aversion to the vertical axis, while START prefers the sagittal axis and has an aversion to the lateral axis. Thus, although gestures across all auxiliaries most frequently features movement along the vertical axis as shown in the summary tables above, axis does appear to play a unique role in the profiles for START and QUIT.

An analysis of gesture direction contributes further information to the individual auxiliary profiles. A chi-squared test of gesture direction by auxiliary was significant ($p<.05$). A Pearson residuals posthoc showed that QUIT and STOP both featured lateral movements (these were bimanual gestures with each handing moving outwards), and KEEP had a greater than expected occurrence of upwards movements on the vertical axis.

The distribution of open aspect and phase aspect by gesture movement type (straight, circle, or arc) is shown in Table 3.11. While there was no significant effect, relative frequency shows that straight gestures dominate for both open and phase aspect, while circle gesture traces are more characteristic of open aspect. Arc gestures are more frequent for phase aspect auxiliaries.

Table 3.11. Relative frequency distribution of movement type by aspect

Movement type	Open aspect	Phase aspect
Arced	12%	15%
Circle	17%	9%
Straight	68%	70%
Static	3%	6%

Interactions not related to aspect type include an effect of movement type on both gesture timing and mean action phases, as shown in Table 3.12. This table shows that circular movements have the most asynchronous onset timing and the greatest number of action phases; however, it is worth noting that the mean onset asynchrony for circle gestures with open aspect auxiliaries CONTINUE and KEEP is -870 msec, while for phase aspect (START and QUIT; STOP had no circular gestures) it is -338 msec, which supports the finding of aspect type on gesture timing. The predictions stated at the outset were that the phase aspect auxiliaries would uniformly feature a more synchronous onset given their aspectual emphasis on the beginning or end of an event, which is indeed the case, even when the interaction of movement type on onset timing is taken into consideration.

Table 3.12. Movement type by mean action phases and gesture timing

Movement type	Mean APs*	Gesture timing**
Arced	1.9	-230 msec
Circle	2.9	-660 msec
Straight	1.8	-260 msec

*p<.05, **p<.001

In sum, the findings show that type of aspect plays a role in both plexity (number of action phases) and gesture timing. These are the two variables that were predicted to most abstractly represent the aspectual contour of an event. Since the auxiliaries CONTINUE and KEEP have an inherent aspect (*Aktionsart*) that draws out the various internal phases of an event (Talmy 2000b), it was predicted that the open aspect gesture constructions would involve gestures with more action phases within the stroke when compared to the inherently 'uniplexing' semantics of the phase aspect constructions START, STOP and QUIT. Findings suggest that this prediction is borne out. Phase aspect auxiliary constructions prefer one AP, while the open aspect constructions are characterized by two or more APs. The movement variables – type, axis, and direction – did not correlate significantly with aspect type, but did yield significant results within auxiliaries. These findings speak directly to the existence of a prototypical gesture form for each auxiliary.

3.5 Discussion

In this study, I have moved beyond the parameters traditionally considered to be the loci of meaningful content in gesture and included stroke segmentation into action phases and gesture onset timing in an examination of aspectual constructions. The findings strongly suggest that both gesture timing and action phases are variables that reliably and iconically represent aspectual information in a multimodal manner. The qualitative and quantitative data reveal that each aspectual auxiliary construction has elements that are unique to it, and that the type of aspect – open or phase – differentiates the gestural profiles as well. Here, I discuss possible semantic roots of the profiles that emerged and conclude with a discussion of how this study provides support for considering these and similar expressions as multimodal constructions.

Firstly, given the central place of force dynamics in embodied meaning (Mittelberg 2013b), it befits us to consider these findings in relation to the force-dynamic pairs identified earlier. As discussed, CONTINUE and KEEP could be regarded as constituting a semantic minimal pair vis-à-vis their force-dynamic characteristics: CONTINUE signals an unencumbered continuation of an event, while KEEP conveys that an exertion of force is required to shift an object or event into motion or for it to continue in motion. Here, CONTINUE has a longer onset timing asynchrony at -386 msec compared to -356 msec for KEEP. I propose that the correlation of longer onset with CONTINUE is due to the semantics that signal an unencumbered continuation as compared to [KEEP VVG]. For the action phase variable, the higher mean number of action phases for KEEP (2.6 APs) vs. CONTINUE (2.0 APs) may iconically represent the additional effort required to maintain the motion of the event, which the gesturer performs (though not consciously) through continued stroke iteration.

The force-dynamic profile of STOP and QUIT form a similar minimal pair; [STOP VVG] refers to a neutral cessation of an event while [QUIT VVG] is inherently more forceful. The findings showed that, while STOP can be both uniplex (59%) and multiplex (~ 30% had two APs), QUIT significantly prefers a uniplex gesture. QUIT also has the shortest onset asynchrony of all auxiliaries, at -171 msec before the auxiliary, compared to -293 msec for STOP. The more synchronous onset and single action-phase gestures for QUIT vs. STOP could be accounted for in the force-dynamic contrast between these auxiliaries. Although a direct link is difficult to ascertain in a corpus study of this limited size, the impact of force-dynamics on the findings presented here certainly requires further investigation.

The study of gesture is inherently complex in part due to the fact that "gestures convey meaning in different ways than speech does", namely, in a global and synthetic way rather than a linear process (Zima & Bergs 2017a: 4). Thus, although core forms do emerge, there are outliers, such as the cyclic gesture with six action phases for QUIT shown in Figure 3.10. This example represents a notable departure from the dominance of bimanual, outward gestures prototypical of QUIT, but this could be attributed to the semantics of the complement verb.

S: … *You were waiting for me to quit talking…*

G: left-handed cyclic gesture with six action phases

Figure 3.10. *quit talking*, iterated cyclic gesture

In this instance, it is not the *quitting* event that is gestured, but rather the *persistence of the talking*. The construal of *persistence* affects both movement type and number of action phases and points to the ability of the speaker-gesturer to underline – through co-speech gesture – a specific element of a construal.

In her seminal work on cyclic gestures, Ladewig (2011) suggests that depictive cyclic gestures (which she differentiates from word-search gestures) refer to an ongoing process "in every instance" (ibid.: 6). Thus, for this study, cyclic gestures were expected to be most common in open aspect events involving CONTINUE and KEEP. The finding that was not expected was the high proportion of cyclic forms with START constructions. Ladewig proposes that there is a metaphorical extension of the image schema CYCLE that invokes the motion of a crank as a process in a machine, as in starting up an engine. The over-representation of cyclic gestures with START could be related to the gesture metaphorically representing the start, or 'cranking up' of a process, rather than the ongoing nature of the process itself. This, again, speaks to the ability of the speaker/gesturer to choose which aspect of an event to highlight in his/her construal.

The profiles presented here represent an aggregate of the gesture forms. For each auxiliary there are instantiations that are close to the core profile and instantiations that vary widely from it. Some gestures occurring with open aspect constructions have one action phase, while occasionally a gesture with a phase aspect construction features three APs. Indeed, the major operational problem with multimodal data lies in the multitude of variations that are possible. As outlined in the literature cited earlier on frequency and co-occurrence, and further elucidated in Mittelberg (2007), a person's gesture is influenced by a multitude of factors, including individual idiosyncrasy. In addition to noted individual preferences in gesture space, there are many contextual factors that may influence gesture that have only begun to be examined and which, furthermore, are difficult to control for in naturalistic data. These include linguistic variables, such as inflectional biases (Rice & Newman 2005, 2008; Hinnell & Rice 2014; Rice & Newman 2018), and differences based on concrete vs. metaphorical usage reported on in this study. Further complicating matters is the way that gesture can be used to highlight different salient elements of a construal in a specific context of use, as in the cyclic gesture for QUIT shown above.

This brings us to the major issues of frequency and recurrence that are central to the discussion of what constitutes a multimodal construction. This study provides an empirically grounded analysis of patterns that emerge for each auxiliary related to aspectual profiles and provides evidence to support a preliminary constructional profile for these five auxiliary constructions. With regard to the matter of frequency and co-occurrence as markers of construction-hood, no single study can resolve this issue. Indeed, in light of Langacker's view of entrenchment as continuously scalar (Langacker 1987), it may not be possible (Lanwer 2017), or necessary (Zima 2017), to determine a precise threshold frequency as a measure of its entrenchment.[53] Certainly the field as a whole is wrestling with this issue and further empirical studies are needed. The study presented here contributes evidence of a high degree of conventionalization of the gestural profiles associated with these five aspectualized auxiliary constructions.

3.6 Conclusion

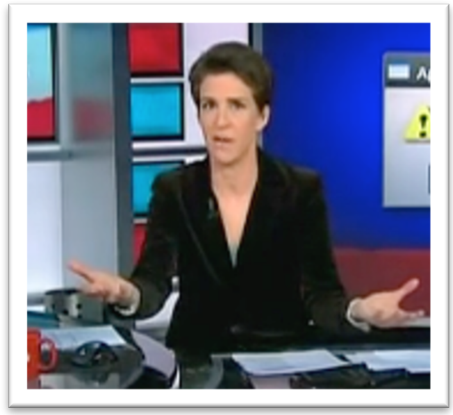
This study has shown that the five periphrastic auxiliary constructions examined here are differentially expressed in co-speech gestures and that the aspectual construal is iconically depicted in a range of gesture parameters. The degree of co-occurrence of the auxiliary constructions with co-speech gesture was 50%-74%. Open aspect auxiliaries (CONTINUE and KEEP) reliably correlate with longer asynchronous onset timing and a greater mean number of

[53] For a full discussion of the different "criteria for constructionhood", see Zima and Bergs (2017) and papers therein.

action phases per stroke, while phase aspect (START, STOP and QUIT) are correlated with more synchronous onset of gesture and fewer stroke segmentations. By including variables such as degree of enactment, onset asynchrony, and number of action phases, in addition to standard form variables, a more complete gesture profile emerges for each of these constructions. Correlations were seen for certain auxiliaries and particular movement directions and movement types. These were shown to parallel semantic distinctions in both aspectual and force-dynamic characteristics of the event construal.

Although co-speech gestures are not an obligatory part of the aspect-marking auxiliary constructions investigated here, when they are gestured, they certainly exhibit conventionalized forms. While the ultimate question remains as to what degree of conventionalization should satisfy a cognitive linguist, here I have shown that aspectual gesture constructions belong to the set of conventionalized gestures that are coordinated with linguistic processes. As Zima and Bergs (2017b) note, more empirical studies, with considerably more depth and breadth with regard to linguistic features, gesture form, and contextual variables, are required. Given the emphasis of seminal studies to date on the multimodal representation of stance and pragmatic markers (e.g. Debras 2017; Schoonjans 2014), further studies that explore abstract and grammatical elements of construal, such as aspect, are needed.

In this chapter, I explored five periphrastic aspectual constructions, each of which is headed by a lexicalized verb. While the constructions were phrasal in nature, as lexical items, the auxiliaries themselves are highly conventionalized. In what follows in the next two chapters, I investigate more open phrasal constructions that express CONTRAST (Chapter 4) and discourse navigation (Chapter 5). These studies explore fixed contrast-marking phrases such as *on (the) one hand/on the other hand* and *should I/shouldn't I,* and idiomatic expressions speakers use to signal directional changes in discourse, such as the parenthetical asides *which, by the way* and the return markers *however* and *but anyway(s))*. In turning my attention to these types of idiomatic 'chunks' that proliferate in spoken language contexts, I move beyond the realm of grammatical constructions to a wider range of stance-marking constructions. The contrast constructions mark a speaker's valuation of two options, while the discourse navigation constructions express how a speaker quite literally directs a discourse. As the following two chapters suggest, these highly idiomatic stance-marking constructions also present unique and conventionalized kinesic signatures.



Chapter 4 | *On the one hand*, *on the other hand:* CONTRAST in the body[54]

4.1 Introduction

Imagine a friend invites you to do something daring, like ride a motorcycle. You might say *Yes!* without reservation. If you are somewhat risk-averse, though, you might voice your indecision by saying *On one hand I'd love to,* and let your voice trail off, or you may offer an assessment such as *That looks like fun, but it might be dangerous.* Weighing options and making choices are part of the experience of being human. In fact, various sources estimate that an adult makes about 35,000 decisions every day and children about 3,000 (Sahakian & Labuzetta 2013). These staggering numbers aside, the variety of linguistic expressions at our disposal to communicate the options we face are also numerous. English has many expressions that are built to express two alternatives. The archetype is the semi-fixed expression *on the one hand/on the other hand.* The simple binary operator *or*, comparative adjectives like *better* and *worse*,

[54] Parts of this chapter have been published as: Hinnell, Jennifer. 2019. The verbal-kinesic expression of CONTRAST in North American English. *The American Journal of Semiotics* 35(1-2): 55-92. Unique to this dissertation chapter are the quantitative and statistical analyses of *on the one hand/on the other hand, better than/worse than* (in §4.3.3), and *should I/shouldn't I* (in §4.4.2)

demonstratives such as *this* and *that*, and adverbials such as *here* and *there,* also potentially contrast two options. Even pronouns could be construed as having opposites, as in *me* rather than *you*, and *us* vs. *them*. With this wide range of devices in English, we mark CONTRAST across a variety of domains, for instance, contrasting objects in the propositional domain (*Are you having the chicken or the fish?*), considering choices – an activity in the cognitive domain (*Should I do it? Should I not do it?*), and distancing ourselves from a previous discourse thread (*however* and *anyways*).

This robust array of verbal expressions of CONTRAST in English is matched by a wide range of body movements. For example, as you formulate a response to your friend's invitation to join her on the motorcycle ride, you might move one hand to the side with your palm facing upwards as you say *On the one hand, I'd love to*, or you might tilt your head first to one side and then the other as you say it might be *fun,* but you're worried it's also *dangerous*. As I demonstrate throughout this chapter, speakers use a variety of hand, head, and shoulder movements, as well as eyebrow raising and gaze shifts, to mark CONTRAST. The result is a robust marking of CONTRAST in the body. Some enactments are signaled very strongly, such as the highly iconic and imageable bimanual gesture frequently associated with the idiomatic expression *on (the) one hand/on the other hand.* Others enactments of CONTRAST, though, are more reduced and schematic, such as subtle head tilts from side-to-side, or a gaze shift.

Throughout this chapter, I present corpus-based analyses of several fixed and semi-fixed expressions, as well as lexical and pragmatic material, that set up CONTRAST in a variety of domains. I document the wide range of bodily behaviors that speakers co-produce with these linguistic cues. I argue that, while the level of abstraction differs between the recognizable iconicity of some enactments and the more schematic representations of others, the underlying iconic and indexical processes at play unify these speech-aligned behaviors as

CONTRAST marking. Furthermore, the data suggest that some of the variation in the enactments of CONTRAST is motivated by the domain in which the utterance inheres. Thus, in a contrastive scenario, propositional, real world content is most likely to be indicated with manual gesture and is more highly depictable in the hands, while in the cognitive domain (e.g. *Should I or shouldn't I?)* CONTRAST is more likely to be expressed through head movement. CONTRAST in the discursive domain, in turn, has yet another co-speech profile and is more schematically referential. In making this argument, I revisit the issue of convention and variation in co-speech behavior that I introduced in Chapter 1. As reviewed in that chapter, there is a robust literature emerging in cognitively oriented fields that investigate conventionalized patterns of co-speech behaviour such as manual gestures, head movement, gaze, and other body behaviours that accompany speech in interaction (Müller, Cienki, et al. 2013; Müller et al. 2014; Schoonjans 2014; Debras 2017; Jehoul et al. 2017). The findings presented in this chapter capture both the convention and variation inherent in expressions CONTRAST.

Previous studies using corpus linguistic methods have pursued the marking of *stance* – an umbrella term encompassing a speaker's viewpoint, attitudes, or judgment (Biber and Finegan 1989; Precht 2000), of which assessments of CONTRAST are an important sub-category. There are few corpus-based studies of spontaneous interaction related to specific stance-relevant predications from both a cognitive and a multimodal perspective and there has yet to be an investigation of how different form conventions contribute to embodied meaning in the construal of CONTRAST (cf. Parrill and Stec 2017 for an experimental study on the contrastive use of space to gesture abstract concepts). Throughout this chapter, I provide a series of case studies that document the articulation of CONTRAST in the body. Expressions of CONTRAST might set up opposing objects or ideas that belong to the everyday, 'real world', as in, *Are you having the chicken or the fish (for dinner)?*, or could be set up an opposition in a hypothetical world (e.g. *I don't know if she's coming to the party*, which invokes two possibilities, one in which someone will come and a second world in which she will not). The distinction between real and hypothetical scenarios is expressed in semantics by the term *mood*, with *realis* capturing events that are in or about the speaker's "actual world" and *irrealis* mood capturing events in the "nonactual world" (Frawley 1992: 387). In what follows, I consider realis and irrealis to be domains of expression. Since, in the realis domain, speakers compare two things in the real world, I consider it to be broadly propositional in nature. Conversely, the irrealis domain is characterized by utterances in which the speaker is thinking, imagining, or hypothesizing and I thus consider it to related largely to the cognitive domain.

4.1.1 CONTRAST background

Logically speaking, CONTRAST is the pitting of one position vis-à-vis another. In the study of rhetoric, this is captured by the notion *antithesis,* a figure of speech which brings out a contrast in the ideas by using highly contrastive words clauses or sentences, often within a parallel grammatical structure.[55] Historically, considerations of gestures expressing CONTRAST were included in manuals on rhetorical gesture, such as Bulwer's 17th century treatise *Chironomia* (1644). In it, Bulwer shows gestures for signaling types of antitheses, as exemplified by the phrase *on (the) one hand/on the other hand*, which he accompanies with the sketch in Figure 4.1.[56]

---

55 OED Online. Accessed October 23, 2019. https://www.oed.com/view/Entry/8893?redirectedFrom=antithesisand

56 I am grateful to Dr. Kensy Cooperider (Department of Psychology, University of Chicago) for making me aware of Bulwer's treatise.

Figure 4.1. Bulwer's rhetorical gesture for *antitheses* (Bulwer 1644)[57]

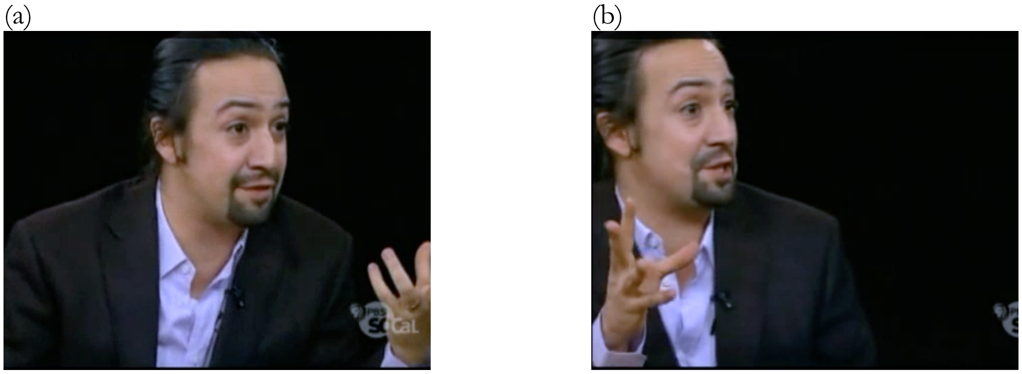
Aristotle's focus on logical dualism (Ogden 1932) grounded de Saussure's (1916) notion of *différence*, which in the 1920s led Prague School linguists to base their approach to language structure in opposition theory. For Saussure, *différence* was inherent to semiotic structure and captured the necessity of otherness. Most relevant to the study presented here is Saussure's notion of *difference* among concepts, which he described as follows: "Concepts are purely differential and defined not by their positive content but negatively by their relations with the other terms of the system. Their most precise characteristic is in being what the others are not" (de Saussure 1916: 117). Work within semiotics on opposition theory has been revived somewhat in recent decades. For example, Danesi and Perron's (1999) model of interconnectedness in culture and cognition includes implicit reference to opposition in their use of image schema theory (*up/down, closed/open).* In looking at how single binary opposition might be encoded in a cultural system, Danesi suggests that the binary opposition *right/left* "is derived, anatomically, from the fact that we have a left hand (and foot, leg, ear, and eye) and a right one" (Danesi 2009: 29). This aligns with original scholarship in cognitive linguistics

[57] This image is in the public domain and was accessed online here: https://publicdomainreview.org/collection/chirologia-or-the-natural-language-of-the-hand-1644

(Lakoff & Johnson 1980; Johnson 1987), and recent work in embodied cognitive science (Bergen 2012) and semiotics (Danesi 2009; Pelkey 2017) that concurs on the central role of the body and our embodied experience in the world in driving linguistic and conceptual structure.

I use the term CONTRAST broadly to capture the general conceptual notion of oppositionality, which should not be taken as restricted to just binary opposition. As I show here, however, the co-speech behaviour of the body frequently suggests that the speaker does generally perceive CONTRAST as binary, whether the difference is directly oppositional or not.[58] For example, take the utterance shown in Figure 4.2.

(a) (b)



(a) S: *Every biography either is insanely defensive of him*

G: left hand gestures to the left, palm-up container form

(b) *or vilifies him*

right hand gesture to the right, palm forward with outstretched fingers

Figure 4.2. Lateral gestures marking CONTRAST: *insanely defensive of him* vs. *vilifies him*

In this example, two options are given for the way that different biographies of a famous person reference the 'him' named here: they are either *insanely defensive of him* or they *vilif[y] him.*

---

[58] In his view of opposition, Jakobson (1962) delineated between contradictory and contrary opposition. An example of the former is *either/or,* whereas for him contrary opposition denoted two end points on a graded continuum such as *white* and *black*.

Each option fills a slot in the *either/or* construction. Now, *to be defensive of someone* and *to vilify someone* are not oppositional antonyms; however, due to their use in the *either/or* construction together with the manual gestures that anchor each of these options in opposing space on the speaker's left or right side, they are coerced into an oppositional reading. In other words, the linguistic and gestural signals together make manifest the conceptual structure in which the speaker views these elements as opposed. In this chapter, I examine how language and the body together manifest a conceptual comparison between two (in some way) opposing forces.

4.1.2 Expressions and domains of CONTRAST

In English, there are myriad ways of expressing a contrastive relationship. In the lexicon, antonyms express CONTRAST, for example antonymic polar adjectives (*good/bad*) and contrastive nominals (*defense/offense*). Quasi-grammatical conjunctive expressions such as *versus* in *Clinton vs. Trump* serve a similar function, as does the related fixed expression *vice versa*.[59] A range of semifixed to fully fixed idiomatic expressions also provide speakers with mechanisms for evaluating options, as in the archetypal *on (the) one hand/on the other hand*, as well as related expressions such as *on one side*.[60]

Just as the conceptual realm of CONTRAST can be instantiated linguistically through an array of expressions, it can hold in a variety of domains. The range of expressions listed thus far are typically used by speakers to express options in the real, actual, described world

[59] *Versus* and *vice versa*: both from Latin into late Middle English in 1400-50 and 1595-1605 respectively, meaning 'to turn from a vice' (*versus*: past participle of *vertere*, to turn). www.dictionary.com. Accessed February 21, 2019.

[60] See Lenk (1998) and Ford (2000) for text-based studies of English discourse markers and contrast marking in English, respectively, both from a functionalist approach to conversation analysis.

(denoted by the semantic notion of *realis*); however, the conceptual notion of CONTRAST can equally apply in the *irrealis*, or hypothetical, world. For example, the contrasting of two mental states, as in *Should I or shouldn't I?* has been described as *alternativity* (Dancygier & Sweetser 2005: 35). According to Dancygier and Sweetser, alternativity refers to two alternate spaces that are construed from the same base or reference space, but that are necessarily incompatible. Other space-building expressions in *irrealis* mood include pairs of conditional *if*-statements, such as the pair of utterances in (40).

(40) *If Gary Johnson gets to that level, he'll be on the stage. If he doesn't get to that level, he won't.*[61]

In the case studies presented throughout this chapter, I use specific expressions and domains in which they inhere as organizing principles. I focus largely on the realis domain and provide case studies of *on (the) one hand/on the other hand* and *better than/ worse than*. I then describe evidence of contrast-marking from irrealis domains. This section features several qualitative descriptions and a smaller case study of *should I/shouldn't I*. I explore enactments of discourse-level contrast in Chapter 5 as stance-marking discourse constructions.

4.1.3 Embodied CONTRAST

In the previous section, I introduced some of the expressions and domains in which CONTRAST is expressed. Whichever expression is used, and whether referring to events in the actual world or in possible worlds, CONTRAST fundamentally represents the balancing of two options. Balance, in turn, has been described as a fundamental image schema. As described in Chapter 1, an image schema is a "recurring, dynamic patterns of our perceptual interactions

[61] Red Hen: 2016-09-08_1400_US_MSNBC_The_Place_for_Politics_2016,2930

and motor programs that gives coherence and structure to our experience" (Johnson 1987: xiv). Regarding balance specifically, Johnson states:

> We almost never reflect on the nature and meaning of balance, and yet without it our physical reality would be utterly chaotic, like the wildly spinning world of a very intoxicated person. The structure of balance is one of the key threads that holds our physical experience together as a relatively coherent and meaningful whole. […] The *meaning* of balance begins to emerge through our *acts* of balancing and through our *experience* of systemic processes and states within our bodies.
>
> (Johnson 1987: 74-75)

The BALANCE image schema overlaps with the image schema of bilateral symmetry. As Turner notes, "the prototype of the balance schema is the schema of bilateral symmetry" (Turner 1991: 70).[62] The symmetrical and bilateral nature of the body is mirrored in the embodied experience of the speaker in her own body. As a three-dimensional entity in the physical world, the body has a front and back, two sides, two hands, etc. Thus, speakers will have an embodied sense of BALANCE, SYMMETRY, and, therefore, CONTRAST, from the way the body moves in the physical world they inhabit. This could be forward or backward, to the left or right. Similarly, individual articulators are most free to move in a constrained set of directions: hands move left or right, towards the body or away from the body, up or down; shoulders move up or down, eyebrows raise and lower, heads tilt left or right and nod up or down. The embodied experience of speakers in the world matched with the binarity inherent in each articulator's movement patterns make the body particularly adept at expressing the cognitive construct of CONTRAST.

---

[62] Both Johnson (1987) and Turner (1991) note that symmetry and balance are not identical, in that "physical balance can occur in cases where bilateral symmetry is absent" (Turner 1991: 70).

Recent studies on image schema and manual gesture underscore that "image schemas are readily available, indeed 'on hand' for recruitment as gestural forms" (Cienki 2005: 435; Mittelberg 2013b, 2013a). Indeed, both the BALANCE and SYMMETRY image schemas are observable in the archetypal fixed expression of CONTRAST in English, *on (the) one hand/on the other hand*. Through direct reference to the hands, this expression inherently invites the use of manual gesture to impart the pragmatic force of the expression. The setup is conventionally indicated with a palm-up open-hand (PUOH) gesture executed on one side of the body, with the resolution gestured with a subsequent PUOH form to the opposite side, as shown in the line drawing in Figure 4.3.[63]

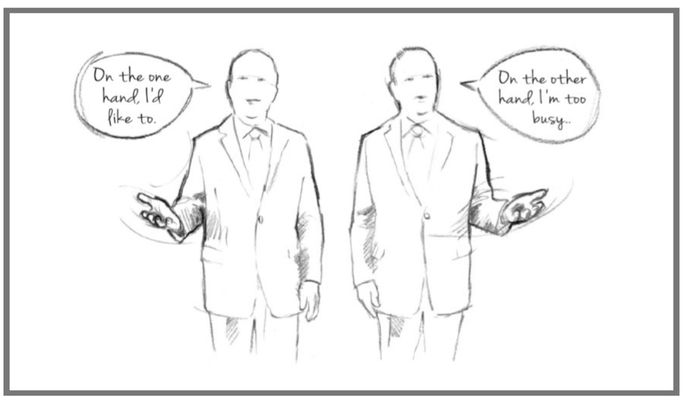


Figure 4.3. Line drawing of *on the one hand/on the other hand*

However, *on (the) one hand/on the other hand,* hereafter referred to as the [O1H-OOH] construction, is not the only expression to invite this type of symmetrical, lateralized pattern in the body. Rather, the conventional bilateral demonstration of alternatives is also seen with linguistic expressions that do not make explicit reference to the body, as in a comparison driven by lexical antonyms such as *good* and *bad* when referenced by co-speech gestures in the space to either side of the body (shown below in Figure 4.16).

---

[63] https://www.johnson-hunter.com/2015/12/15/gesturing-on-the-shelf/. Accessed Nov. 20, 2018.

Alongside the BALANCE and SYMMETRY image schemas that are seen to be at play in Figure 4.3, the use of space in manual gestures has been shown to be structured by other *conceptual metaphors*, in which one abstract domain is understood in terms of another (Lakoff & Johnson 1980). For example, it has been shown that gestures are frequently placed on a timeline situated on the horizontal axis, in which the past is gestured on the left and the future is on the right for users of left-to-right writing systems, such as those that predominate in roman-based orthographies for European languages (Casasanto & Jasmin 2012). This mapping reflects the conceptual metaphor TIME IS SPACE (Núñez & Sweetser 2006; Cooperrider & Núñez 2009; Winter et al. 2013; Casasanto & Bottini 2014). Finally, other studies of the ways in which abstract concepts, for example, that of *similarity,* map onto gesture space, show the role of conceptual structure in shaping the exploitation of gesture space. Speakers can bring their hands closer together to embody similarity and hold them apart to embody conceptual divergence (Sweetser 1998). In judgment tasks about the similarity of abstract entities or object properties, stimuli presented closer together were judged to be more similar, which the authors argue is "consistent with predictions based on linguistic metaphors linking similarity to physical closeness" (Casasanto 2008b: 1055).

There is evidence that evaluation also structures gesture space. Specifically, studies have shown that semantic prosody or valence (both terms capture positive or negative evaluation of an object, event, or idea), are associated with handedness and, more generally, use of lateral space (Casasanto & Jasmin 2010; Casasanto & Chrysikou 2011; Casasanto & Henetz 2012). These studies suggest that speakers gesture with their dominant hand when gesturing positive propositional content, while use of the non-dominant hand tends to co-occur with negative content.

Finally, speakers use gesture space to situate both concrete and abstract referents in discourse (McNeill 1992, 2005; Perniss & Ozyürek 2015). As Parrill and Stec put it based on their elicitation study of contrastive use of gesture space, "when speakers gesture about concrete referents, the locations of the referents within gestural space are consistent" (Parrill & Stec 2017: 34). They give the example of a speaker seeing a photo of a cat and mouse, and then in their gestures, placing the cat and mouse in isomorphic relation in gesture space to their positioning in the photo. Their study also found that these consistent patterns of reference in gesture space hold for abstract referents.

A final important contribution to our understanding of embodied CONTRAST comes from sign language research. Winston (1995, 1996) discusses the functions of spatial mapping strategies by signers. Beyond having a referential function, space is also used to structure discourse. Winston describes the discourse structure of comparative spatial maps in ASL as follows:

> The signer usually introduces the two entities to be compared without using a spatial map, then proceeds to build a spatial map to make the comparison. The signer accomplishes this by pointing first to the non-dominant side of the signing space to refer to the first entity. She then points to the dominant side to refer to the second entity. The second entity … is often the focus of the comparison. The signer continues to refer to the two entities by pointing to the two areas on the spatial map, comparing them throughout the discourse.
>
> (Winston 1996: 10)

Research on viewpoint and real space blends in ASL has further detailed the use of space in expressing grammatical and conceptual information in ASL. In his discussion on viewpoint and comparative spaces, Janzen notes that in using comparative space mapping, the signer "does not use body shifts towards each space for the purpose of portraying the vantage point of either referent on a scene, but rather to list and describe attributes of each while maintaining the role of 'informer' in the discourse" (Janzen 2012: 168). These uses of space in ASL are worth keeping in mind, as we see similar building and indexing of space in co-speech behavior aligned with CONTRAST.

In sum, we know that gesture space and sign space are exploited by signers and speakers, respectively, for communicative and cognitive purposes (see discussion in Parrill and Stec 2017). It is structured in ways that reflects image schemas and conceptual metaphors that are driven by our experiences of our bodies and the world around us. Furthermore, speaker-gesturers use gesture space to anchor and maintain both concrete and abstract referents in a specific location in gesture space so that they can consistently refer back to them in anchored spaces. In this chapter, I investigate how spatial mappings are conventionalized and in what ways they vary across expressions and domains of application of CONTRAST marking in English.

4.1.4 Conventionalization in the embodied signal

Like speech, co-speech behaviour is highly conventionalized, although gestures appear at first glance to be more *ad hoc* and less constrained when compared to speech. As discussed in the introductory chapter and the study of aspect-marking in Chapter 3, this has led to important discussions regarding the degree of conventionalization required to support the inclusion of the kinesic signal in a linguistic utterance. Here, I have already introduced some early claims and more recent findings that support the hypothesis that there is a degree of conventionalization in gestures of CONTRAST. Throughout this chapter, I will explore the conventionalization of the signaling of CONTRAST in the body by looking at a number of variables. Firstly, for the set of expressions I investigate, I capture the signal strength of the kinesic contribution. This is the degree of enactment I described in Chapter 2. Secondly, also described in the Methods chapter, I capture the set of articulators which are recruited to express CONTRAST by partitioning the body into gestures of the hands and gestures of the upper body. The body partitioning profile captures the observations that emerged from my analysis that speakers tend to gesture propositional contrast in the realis domain more in their hands (e.g. with a PUOH to each side), while in the more epistemic, imagined, or irrealis domain, CONTRAST tends to be expressed in the head (e.g. through a series of bilateral head tilts to each side). By capturing whether an expression is likely to be expressed in the hands, in the upper body, or in both gesture and upper body, I broaden the discussion of the locus for conventionalization and iconicity in co-speech gesture. As discussed in Chapter 1, and also shown in the study of ASPECT presented in Chapter 3, discussions of iconicity have tended to be restricted to specific parameters related to gesture handshape and movement. My approach moves the discussion beyond the established loci of iconicity by including upper body movement, as well as variables such as laterality and symmetry.

4.1.5 Summary

In this chapter, I examine the expression of CONTRAST in North American English in the real world (realis) and in more hypothetical scenarios (irrealis). I describe a range of bodily and linguistic resources that are recruited to mark CONTRAST. In distinction to the study of ASPECT presented in Chapter 3, in this chapter I include articulations beyond manual gesture, such as head nods, shoulder shrugs, and other meaningful movements in the upper body. In describing the enactments of CONTRAST, I aim to demonstrate that, while the marking of CONTRAST in the body is not obligatory, it is certainly not arbitrary.[64] As the findings show,

[64] Personal communication with Sally Rice (Department of Linguistics, University of Alberta)

gestures produced in the CONTRAST frame in many cases belong to highly recognizable iconic forms, such as the palm-up open-hand gesture. Enactments in the cognitive domain are more abstract. That is, in irrealis contrast marking, speakers often use shifts in the body (e.g. head tilts, gaze shifts, a shift from a palm-up to a palm-down), rather than the fully expressed weighing or presenting of options in both hands that is so characteristic of CONTRAST in the propositional domain. Thus, when analyzed across the propositional and cognitive domains, the ways in which CONTRAST is marked in the body can be viewed as a continuum of highly imageable to more schematic kinesic shifts, as well as shifts from the lower to the upper body in terms of dominant use of hand versus head and shoulder articulators. By placing the primary focus on the multimodal signal, this chapter hopes to expand our understanding of how speakers of North American English build meaningful environments around the construal of CONTRAST in their speech and body movements.

4.2 Methods

In this section, I introduce methods particular to the case studies presented here. The data for this study were collected using the archive and facilities of the Red Hen multimedia archive, described in detail in Chapter 2. The data set includes TV programs broadcast between November 2012 and November 2016. Only spontaneous interactions were accepted into the data. In cases in which a search result stemmed from a scripted show, prepared news report, political speech, or other prepared, non-dialogic genre, the data point was excluded from the study.

4.2.1 Target utterances

Search terms for the case studies presented throughout this chapter consisted of a range of lexico-syntactic constructions expressing CONTRAST collected from the literature and an exploratory pilot study. I began this research with a pilot study of the *on (the) one hand* and *on the other hand*, construction, or [O1H-OOH]. In that study, the data yielded further semi-fixed phrasal expressions that are frequently used in place of either the first or second segment of [O1H-OOH]. In some examples, a paraphrase replaced one element of [O1H-OOH], as in (41), in which *on the other side* completes the expression. Example (42) shows a different bipartite expression in use altogether. In these examples from the Red Hen, the setup and resolution phrases are underlined.

(41) *There is a sense in the two different pieces that I was trying to present there <u>on the one hand</u> a kind of doctrinal theological emphasis of this pope, and <u>on the other side</u>, a very practical question about the viability of the catholic church in the world.*[65]

(42) *<u>It is in one way</u> very exciting that this is finally a matter of high-level political discussion. <u>It is also at the same time</u> very disconcerting to see the Republican Party …*[66]

The data also featured many cases in which there is only overt marking at the lexico-syntactic or phrasal level on one side or the other of the fully fixed utterance. This phenomenon occurred much more frequently for *on the other hand* (i.e. the setup was a clause *not* introduced by *on (the) one hand*). Two examples are given in (43) and (44):

(43) *So, I hear your point about not overstating the importance of good music and good songwriting. <u>On the other hand</u>, the boundless and endless joy that we experience because of music and the place that we come to appreciate because of good music…*[67]

---

[65] Red Hen: 2015-09-05_1400_US_MSNBC_Melissa_Harris-Perry,359

[66] Red Hen: 2015-05-19 0100 US_MSNBC Rachel Maddow Show,676

[67] Red Hen: 2015-10-23_0600_US_KOCE_Tavis_Smiley,1136

(44) *So, you can look at that in two ways. You can go awesome, Paul Ryan values family and that means he has his priorities straight. <u>On the other hand</u>, some are rolling their eyes this morning. After all, he works 132 days a year, no five-day weeks…*[68]

Common semi-filled constructional frames that set up CONTRAST include *X and Y, both X and Y, X and Y alike, X as well as Y, X or Y,* and *either X or Y* (Murphy 2006). Murphy refers to these schematic frames as Coordinated Antonymy constructions. In her view, these constructions have two primary features: "(a) antonyms tend to co-occur in particular lexico-syntactic frames and (b) the frames themselves require or underscore the contrast between the X and Y elements" (Murphy 2006: 10). As an archive rather than an annotated corpus, Red Hen is not tagged for parts of speech. The search function is largely restricted to text strings and does not easily facilitate searching for semi-filled constructional frames such as those listed above. For the purposes of the antonym case study in this chapter, I therefore searched for specific antonymic noun pairs. That is, I used lexical string searches to find antonym constructions, rather than searching for the syntactic frames that Murphy identifies. For example, I searched for the utterance pairs *David/Goliath* and *good news/bad news*. I also included the comparative pair *better than/worse than.*

In the irrealis domain, I present a case study of the utterance pair *should I/shouldn't I*. A sampling of search returns are given in (45) through (48).

(45) *What should I wear? What shouldn't I wear?* [69]

(46) *Should I speak more loudly, should I speak more softly?* [70]

(47) *Should I or shouldn't I get the flu shot?* [71]

---

[68] Red Hen: 2015-10-21_1400_US_CNN_Newsroom,435

[69] Red Hen: 2018-06-21_0100_UK_KCET_BBC_World_News,1140

[70] Red Hen: 2018-06-21_0100_UK_KCET_BBC_World_News,1142

[71] Red Hen: 2018-02-20_2300_US_WEWS_News_5_at_6pm,1010

(48) *It's just like most stupid things. You didn't think about - I'm risking my life - should I or shouldn't I?* [72]

I also explored the enactments of *if*-clauses. Due to the heterogenous uses of *if,* for example to mark clausal conditionals such as *if… then…*, this was difficult to operationalize in Red Hen and I report here on only a few examples. I explain how I collected data in Red Hen for each utterance in §4.2.2.

4.2.2 Data collection in Red Hen

In this subsection, I explain the methods used to capture each target utterance in the Red Hen archive. In the realis domain, I systematically collected data for the bipartite fixed expressions *on (the) one hand/on the other hand* and *better than/worse than*. In the irrealis domain, the corpus study was for the target utterances were *should I/shouldn't I*. Other target utterances were collected on an ad hoc basis.

On (the) one hand/on the other hand

For the study of [O1H-OOH], each part of the expression was searched for independently rather than searching for both phrases within one segment. In natural discourse, the two segments can occur quite far apart in speech time and I did not want to limit myself to a short phrase between the two parts of the expression. Since the onset construction has two variations, *on <u>the</u> one hand* and *on one hand*, both were included in the same search using the Red Hen search function 'with at least one word/phrase' (see Figure 2.1 in Chapter 2). Thus, the results for [O1H] include examples of both variations. I searched the archive from November 1, 2014, to November 1, 2015. This resulted in 1,590 search returns for [O1H] and 2,783 for

[72] Red Hen: 2017-06-29_0500_US_KTLA_KTLA_5_News_at_10,1877

[OOH]. Working in reverse chronological order, I viewed the search returns in Red Hen until I had collected 100 viable search returns of each utterance.[73] I exported the Red Hen metadata to a spreadsheet and proceeded with annotations. As explained in Chapter 2, on the first pass through the target (collected and viable) utterances, each example was annotated for whether it featured any co-speech behavior aligned with the target utterance in the speech signal. This co-occurrence rate establishes the degree of signal strength for each of these utterances. For both [O1H] and [OOH], the rate of co-occurrence was 92% of 100. From these 92 enacted examples of each construction, 50 were randomly selected for complete annotation using the random selection function in Excel. The final data set for annotation for [O1H-OOH] thus comprised 50 tokens each of [O1H] and [OOH]. As for the makeup of speakers, this final set of 100 tokens comprised 59 male speakers and 41 female speakers. Each speaker was represented only once with two exceptions: talk show host Rachel Maddow appeared three times and Melissa Harris-Perry twice.

For the analysis of [O1H-OOH] presented in §4.3.3 below, I viewed the data from two perspectives: by phrase and by construction type. For data annotation by phrase type, I looked at each of the 50 tokens of just one part of the construction at a time, namely, [O1H] or [OOH], and annotated a subset of the linguistic and kinesic variables. Viewing the data by phrase type allowed me to investigate the ways in which speakers enacted each part of the phrase, regardless of whether it was part of the full construction [O1H-OOH] or one of the partial constructions, [O1H-exp] or [exp-OOH]. (Here, "exp" is a place-holder for myriad actual expressions that complete the binary contrastive statement). The analysis by

[73] As per the definition provided in Chapter 2, valid data points were taken from spontaneous interaction with all articulators visible

construction type, conversely, allowed me to investigate elements of the profile that were specific to each construction and those variables that examined both parts of the construction, such as symmetry and laterality, previewed in the Methods chapter and described in detail in the next section (§4.2.3).

The data set used for the analysis by construction type was a subset of the 100 tokens in the full data set. Recall that I collected the data using each part of the [O1H-OOH] construction separately, i.e. searching first on [O1H] and then performing a second search for [OOH]. Therefore, those utterances featuring both parts of the full construction occurred in both the [O1H] data set and the [OOH] data. Duplicates were removed for the analysis by construction type. In other examples, only one part of the construction was visible because the camera panned away from the speaker, which meant that the full enactment of the bipartite construction could not be annotated. Lastly, there were a few tokens in which there was no resolution of the bipartite expression (i.e. the speaker started with *on (the) one hand*, but over the next several discourse segments never resolved the contrast). These tokens were removed from the analysis by construction type as well as there was no second portion available for analysis. For these reasons, the number of tokens of the construction types in Table 4.1 totals 72 rather than 100. Table 4.1 shows each construction type with the corresponding number of tokens in the data set.

Table 4.1. Distribution by construction type

Construction	Total	Duplicates or not visible	Data set for annotation
[O1H-OOH]	31	10	21
[O1H-exp]	32	5	27
[exp-OOH]	37	3	34
TOTAL	100	18	72

Better than/worse than

For the case study of *better than/worse than* [BT-WT], I searched Red Hen for *better than* and *worse than* within 20 words of each other. The dates were set from November 2016 to December 2018. Again, I viewed all search returns in reverse chronological order and stopped once I had viewed enough search returns to collect 25 valid (i.e. spontaneous and visible) examples of the target utterance. I then proceeded as described above for [O1H-OOH] by first annotating for any co-speech behaviour to determine the degree of enactment and then proceeding with annotations as outlined in the next section. Consistent with the gender balance in the [O1H-OOH] data set, there were more male than female speakers, with 60% male.

Should I/shouldn't I

For the case study of *should I/shouldn't I* [SI-SNI], I searched Red Hen for the expressions *should I* and *shouldn't I* within ten words of each other. I constrained the search to 10 words for this expression after pilot studies showed that a wider search resulted in much more incongruent data. That is, the two expressions were more likely to be independent of each other rather than in a related, bipartite expression of CONTRAST. Preliminary searches showed that there were far fewer tokens of this construction in Red Hen. I thus set the dates to search a larger segment of the archive (November 1, 2009, to December 1, 2018) and included all programs and networks in the search to gain the maximum number of search returns. This yielded 179 search returns. I viewed all of them and had to discard all but 13 of them as they stemmed from ads, syndicated programs that were highly repetitious or involved segments in which *should I* and *shouldn't I* occur in close proximity but are not in a contrast frame or were speech errors in which a speaker first said *should I,* mistakenly, and then repaired to *shouldn't I*.

Lexical antonyms and bi-clausal if-statements

As mentioned above, for lexical antonyms in the propositional domain and for examples of CONTRAST in the cognitive domain expressed through bi-clausal *if*-statements, I discuss trends in individual search returns rather than presenting quantitative case studies.

4.2.3 Data annotation

Kinesic form annotations

Searches were conducted by entering linguistic target phrases into the Red Hen search engine. Red Hen then returns a link to a video that contains the target phrase. Each video clip was viewed for about 20 seconds (or more if warranted) on either side of the target utterance. After assessing that the context met the criteria for unscripted, interactional speech, each clip was viewed for whether there was body movement aligned with speech, i.e., was there co-speech behaviour that was temporally aligned with the utterance, and was the movement differentiated from the previous and following utterances. Following this, for all utterances that featured kinesic movement, two annotation passes were completed. On the first pass, the involvement of each articulator (hands, head, etc.) was annotated as *yes* or *no*. On the second pass, in all cases in which there was movement of an articulator, the movement was coded as per the annotation schema shown in Table 4.2 (repeated from Table 2.1 and Table 2.2 in Chapter 2).

Annotations for the contrast data were necessarily different than the annotations for the aspectual movement documented in Chapter 3. It was apparent from pilot studies that there was a great deal of symmetry and use of lateral space involved in the body enactments. I therefore created additional variables, *symmetry* and *laterality,* to encode these features, which I describe below.

Symmetry

Symmetry captured whether there was a correspondence in the form of co-speech behaviour used on both sides of the contrastive utterance. Take for example the bipartite construction, [O1H-OOH]. If the speaker enacted a PUOH for both [O1H] and [OOH], this was coded as symmetrical. A sequence in which the handshapes were different or one part of the utterance featured a handshape and the other featured no enactment were coded as asymmetrical. I also used *symmetry* to encode whether an upper body enactment shared the same profile for the first and second elements in the utterance. For example, if a speaker tilted her head to one side for *should I* and to the other side for *shouldn't I,* this was coded as symmetrical. If, instead, she returned her head to centre for *shouldn't I*, this would be coded as asymmetrical.

Laterality

While it is not possible from Red Hen data to know the handedness of speakers, several annotations captured preferences for laterality. The variables that were used to assess laterality included any that captured left/right movement. This included which hand was used, direction of movement, use of gesture space, head tilts or turns, and torso shifts. For example, if a PUOH was performed first with the right hand and then with the left hand, laterality was coded as 'right-left' (R-L). If a PUOH form was performed with both hands and executed first to the right and then to the left side of the body, laterality was also coded as R-L. Finally, when a gesture form originated or ended in the centre, but moved to or from the side, the movement was coded 'centre-left'/'left-centre' or centre-right'/'right-centre' depending on the direction of movement.

The full range of variables and levels considered for the data presented in this chapter is shown in Table 4.2.

Table 4.2. Variables and levels of kinesic movement annotation[74]

Modality	Variable	Levels
Gesture	Gesture space	core, left of body, right of body
	Hands	left, right, both
	Axis	vertical, lateral, sagittal, none (e.g. wrist turn only)
	Direction	up, down, left, right, diagonal, towards body, away from body, none
	Palm orientation	palm-up (PU), palm-down (PD), palm-lateral (PL), palm-vertical PV), palm diagonal (di)
	Palm orientation in gesture space	towards-center (TC), away-center (AC), towards-body (TB), away-from-body (AB)
	Recurrent form	PUOH, container, palm-down, etc.
Head	Movement type	nod, tilt, shake, other
	Direction	left, right, up, down (if relevant given movement type)
Shoulders	Direction	up, down
Eyebrows	Direction	raise, lower
Torso	Direction	turn, lean (+left, right, forward, back, other)
All	Symmetry	yes/no: 'yes' if body enactment was the same for both sides of contrast enactment; otherwise 'no'
	Laterality	lateral: right-left, left-right, centre-left, left-centre, centre-right, right-centre not lateral: right-right, left-left, centre-centre
	Partitioning	G (gesture only), G&UB (gesture and upper body), UB only

[74] The first seven manual gesture form variables and levels are annotated according to Bressem 2013. Symmetry, laterality, and partitioning are variables I have defined.

Linguistic form annotations

For all tokens with body movement, I annotated for a range of variables in the speech context. These are listed in Table 4.3.

Table 4.3. Annotation schema for target utterance context

Variable	Levels
Construction type	[O1H-OOH], [O1H+exp], [exp+OOH]
Subject	1st, 2nd, 3rd person + singular/plural
Semantic prosody	positive, negative, neutral
Subjectivity anchors	if yes, noted the expression

Construction type was only annotated for the case study of [O1H-OOH]. I also captured the expression that completed the construction if it was not [O1H] or [OOH]. Values for semantic prosody were positive, negative, and neutral, which I coded separately for each side of the contrast utterance. For example, for the utterance in (49), semantic prosody was coded as negative|positive. The phrase *disrupting the status quo…we don't understand* has negative evaluation, while *so convenient* has positive semantic prosody. The reverse order holds for the utterance in (50), which was coded as positive (*proud*) for the first part of the utterance (*proud*) and negative for the second part (*Vietnam)*.

(49) *Uber is bad. Uber is good. Poor people. Rich people. I don't know, this is a really complex issue. I don't understand it because I like it but then on one hand it's disrupting the status quo affecting workers and the economy in ways we don't fully understand. <u>On the other hand</u>, so convenient, tap the screen and boom you're a two hour drunken plane ride away from hugging Milky Mouth.*[75]

(50) *You must look at that with mixed emotions. On the one hand, a proud moment for him and for the country, not easy. <u>On the other hand</u>, then came Vietnam, and all that suggested and perhaps limited what he might have achieved.* [76]

[75] Red Hen: 2018-06-21_0100_UK_KCET_BBC_World_News,397

[76] Red Hen: 2015-08-01_0630_US_KOCE_Charlie_Rose,235

Finally, I was interested in the correlation between additional markers of subjectivity and changes in the kinesic enactments. I attempted to capture subjectivity by coding phrases in the surrounding utterance that express subjectivity (what I call 'subjective anchors'). I looked for highly stanced and subjective utterances in the surrounding context, such as first person with a verb of perception and cognition (*I feel, I think, I believe, I wonder*), fixed expressions of stance (*are you kidding me?)*, or stance-marking sentential adverbs (*honestly, clearly, importantly*). Even with this broad definition of subjectivity marking, this type of subjective marking proved rare; there were only eight such expressions in the dataset for [O1H-OOH].

4.2.4 Corpus comparison for *on (the) one hand/on the other hand*

Because construction type played a large role in the analysis and findings for [O1H-OOH], I wanted to explore whether the search returns in my sample were representative of American English speech. I compared the occurrence of *on (the) one hand* and *on the other hand* in Red Hen with the rate of occurrence for these utterances in the spoken portion of the Corpus of Contemporary English ($COCA_{sp}$). Table 4.4 shows the relative distribution of [O1H] and [OOH] in Red Hen and $COCA_{sp}$. (Given the disparate size of the two corpora, absolute frequency cannot reasonably be compared and is here for reference only).[77]

[77] Relative frequency here is the frequency of either [O1H] or [OOH] relative to the total for [O1H] and [OOH] together in that corpus.

Table 4.4. Frequency of [O1H] and [OOH] in Red Hen and COCA$_{sp}$

	Red Hen*		COCA$_{sp}$**	
	Tokens	Relative freq.	Tokens	Relative freq.
[O1H]	16,474	35.68%	1,720	27.82%
[OOH]	29,702	64.32%	4,463	72.18%
Total	46,176	100.00%	6,183	100.00%

*2010-2019, **1990-2017

I also wanted to confirm that the distribution of the utterances by construction types in my sample was representative of the larger corpus. I had already collected 100 of each of [O1H] and [OOH] to obtain the degree of embodiment. I then also then randomly selected 100 examples each for [O1H] and [OOH] in COCA$_{sp}$ and sorted the results by construction type. As shown in Figure 4.4, the relative distribution by construction type across both corpora and in my final sample of 100 was similar. The general trend is for [O1H-OOH] to be the least frequent and [exp-OOH] the most frequent construction type.

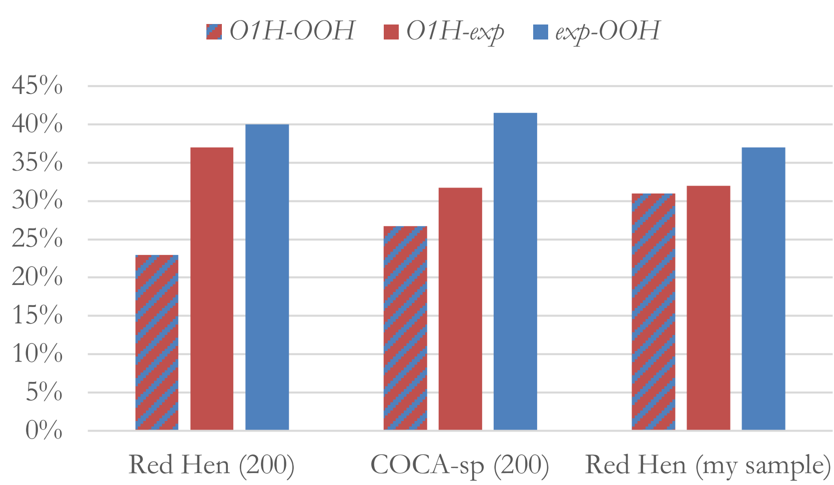


Figure 4.4. Distribution of construction types in Red Hen and COCA$_{sp}$

4.2.5 Summary: Methods

In this section, I have explained how I selected the target utterances, collected the data from the Red Hen archive, and annotated the search returns. I also defined the kinesic and linguistic variables that play a role in the analyses presented in this chapter. In the following two sections, I present and analyze the data for CONTRAST in the real world (§4.3) and in possible world scenarios (§4.4), or, put another way, in realis/propositional and irrealis/cognitive domains, respectively.

4.3 Real world CONTRAST

The expression of CONTRAST in the propositional domain is a very robust phenomenon. Given that *on (the) one hand/on the other hand* is the archetypal expression for signaling CONTRAST, both from a linguistic perspective and a gestural one, in this section I introduce the primary enactments that accompany this fixed and idiomatic expression. I also show that similar gestural and co-speech behaviours characterize the expression of other bipartite expressions, for example, *in one way/it is also* as given in (42) above. From there, I describe lexical expressions of CONTRAST, such as the antonym pairs *David* and *Goliath, offense* and *defense*, *Republican* and *Democrat,* and even *manicure* and *pedicure* as two options for spa treatments. What is most notable about the enactments observed in the propositional domain is that they are highly conventionalized and highly imageable. That is, the recurrent forms of these enactments have recognizable iconic motivations. CONTRAST gestures make use of lateral space and symmetrical hand forms to mark an evaluation of binary options. Following the investigation of propositional CONTRAST, in §4.4 I present kinesic enactments expressions that inhere to the irrealis domain and identify bodily enactments that dominantly recruit other articulators, e.g. head movement, in a more schematic means of representing CONTRAST.

4.3.1 *On (the) one hand/on the other hand*

I begin this section by presenting the gestural enactment found in *on (the) one hand/on the other hand* predications. The examples of [O1H-OOH] shown in Figure 4.5 and Figure 4.6 feature the palm-up open-hand gesture. In Figure 4.5, the host of the Late Late Show, James Corden, utters the following:

(51) *Like, <u>on the one hand</u>, Trump has made a lot of offensive statements. <u>On the other hand</u>, Trump is his party's only chance at winning.*[78]

Aligned with *on the one hand*, Corden raises his right hand in a PUOH gesture with relaxed, slightly bent fingers. When he articulates the second phrase, *on the other hand,* he raises his left hand in the same hand shape to yield the final gestural form shown in the rightmost screenshot. The PUOH gesture is performed in a similar fashion in Figure 4.6, although here the speaker lowers her right hand before proceeding to mark the contrast in her left hand.

(a) (b)

S: *Like, <u>on the one hand</u> Trump has made a lot of offensive statements …*

G: right hand raised in PUOH

<u>On the other hand</u>, Trump is his party's only chance at winning.

left hand rises to match right hand in PUOH

Figure 4.5. *on the one hand/on the other hand*, PUOH gesture

[78] Red Hen: 2016-07-12_0737_US_KCBS_Late_Late_Show_with_James_Corden,151

(a) (b)

(a) S: *Apparently, they like him on both sides. <u>On the one hand</u> he believes in climate change so they consider him a Liberal.*

G: right hand PUOH

(b) *<u>On the other hand,</u> he doesn't believe in abortion so they consider him a Republican…*

left hand PUOH

Figure 4.6.[79] *on the one hand/on the other hand*, PUOH gesture

Several elements in these gestural expressions contribute to the enactment of CONTRAST. Firstly, the gestural signs shown here are strongly iconic. The PUOH provides a metaphoric representation of ideas that are in or on the hand and are being presented or displayed. For the first part of the gesture in Figure 4.5a, the hand is open and relaxed and could be seen to be supporting (or ready to support) an imaginary object. In the second part of the gesture, both hands are flatter with fingers more extended. As such, they mimic the presentation of objects for inspection. Müller (2004: 233) describes one function of PUOH as "present[ing] an abstract discursive object as an inspectable one – an object which is concrete, manipulable, and visible, and it invites participants to take on a shared perspective on this object." In Figure 4.5, the PUOH sequence begins with a presentative gesture in the first part of the gesture phrase and proceeds to a final hold that simultaneously presents *both* alternatives for consideration in the final hold of the gesture. In Figure 4.6, by contrast, each option is only ever presented one at a time.

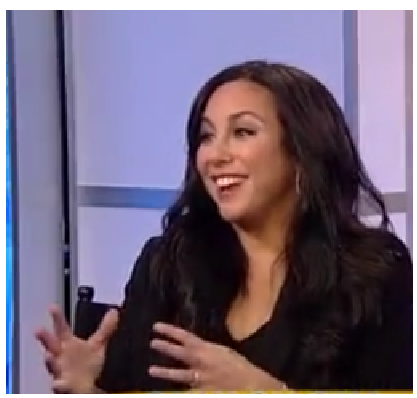
---

[79] Repeated from Chapter 2, Figure 12.

In combination with the iconic hand shapes in these two examples, the speakers index each contrasted notion in body space. Both speakers first place their right hands (in the PUOH form) to the right of the body and then the left hand to the left side of the body. In conjunction with the speech content, these gestures function to build two alternate spaces in the physical gesture space. These spaces then simultaneously reference the notions that are placed on each side. In Figure 4.6, in addition to the PUOH on each side, a head tilt is aligned with each gesture, underscoring the indexing of space on one side. Finally, the physical distance between the two spaces that are built iconically manifests the difference between the ideas, a mapping in which conceptual distance corresponds to physical distance.

Of course, given the nature of gesture formation as both spontaneously innovative and structured by convention, not all gestures aligned with [O1H-OOH] take the form of a bimanual PUOH gesture. A second dominant profile for [O1H-OOH] includes gestures that portray the form of a container. The use of the container gesture is in line with cognitive semantic work on primary metaphors such as CATEGORIES ARE CONTAINERS and IDEAS ARE OBJECTS (Lakoff and Johnson 1980). The container forms in Figure 4.7 and Figure 4.8a, below, show two enactments representing the metaphorical holding of an idea. The idea that is being held is the proposition that follows *on one hand*. Beyond their forms as container gestures, these gestures also index space. Both speaker-gesturers move their imaginary container to the right in Figure 4.7 and to the left in Figure 4.8, as they continue speaking. Thus, while the container form metaphorically represents the notion that is being contrasted, the displacement of hands in space serves an indexing function by placing the idea in an alternate physical space.

The use of eyebrows in Figure 4.8 is also notable. Eyebrow raises have been shown to co-occur frequently with topic marking and as emotive markers (Ekman 1993). Topic-marking is inherently a contrast-marking function; a new topic necessarily represents a shift from the previous topic. Given the wide-ranging possible functions of eyebrow raises, I leave open whether in this example the function of the eyebrow movement is an emotion marker or contrast marker. Importantly, this example illustrates the possibility that different types of CONTRAST are layered in the same enactment. For example, while in Figure 4.8 the bilateral marking of space in the hands is clearly indexical – building a space that the gesturer can then reference, the eyebrows mark the affective stance of the speaker as communicated by the hedge, *kind of exciting*. The eyebrow raise could also be co-indexing the alternativity in the proposition. Given the kinesic limitations of eyebrow movement to the vertical axis, the indexing is necessarily more schematic when compared to the hands, which can place an index in physical space to one side or the other of the body.



S: *Yeah. I mean, on one hand, I was trying to tell the hopeful story, because years ago, he couldn't even run.*

G: container, centre of body

Figure 4.7. *on one hand*, container gesture

(a)

S: *On the one hand*

G: container, centre of body

(b)

it's kind of exciting to be talking Iraq

hands move to speaker's left, relaxing of handshape

Figure 4.8. *on the one hand,* container gesture

The indexical function of CONTRAST gestures also manifests in a range of pointing gestures. Pointing gestures create a relationship between the tip of the articulator (the finger or hand, depending on the handshape used) and the real or imagined object, or target, of the point. They serve the function of a placeholder or "placing index" (Mittelberg and Waugh 2014: 1755). Point gestures feature a pointed finger in the hand form, as in Figure 4.9, while other gestures index space using more neutral hand shapes, as in Figure 4.10.

In the scenario captured in Figure 4.9, the speaker is describing a contrast between the principles of *civility and discourse* that the unnamed third person stands for and this person's actions, which go against those principles. The text of the full utterance is shown in (52).

(52) *You know, as we listen on the one hand, that he talks about civility and discourse and jumps into the Rush Limbaugh controversy, but takes a million dollars from Bill Marr, lectures on the one hand, his surrogates attack on the other hand, he attacks on the other hand. What are we to make of what this says about the president?* [80]

[80] Red Hen: 2012-03-09 1800 US FoxNews Hannity,410

For each item that the speaker lists (underlined in (52)), he uses a point hand shape and indicates to one side or the other. This results in a sequence of four shifts from right side to left side and back again – each aligned with an item that the speaker is contrasting. Here, again, there is a layering in the body that creates a composite signal. In addition to the gestural points, a head tilt aligns with each alternating point to clearly co-index and refer to the space that is built for each option throughout the sequence. The involvement of the head in co-indexing space is unsurprising, as the head is known to be used cross-culturally in points (Cooperrider & Núñez 2009; Cooperrider 2018).

(a) (b) (c) (d)

(a) S: …*takes a million dollars from Bill Maher,*
G: bimanual point to speaker's right

(b) …*lectures on the one hand,*
Point and head tilt to speaker's left

(c) …*his surrogates attack on the other hand,*
Point and head tilt to speaker's right

(d) …*he attacks on the other hand…*
Point and head tilt to speaker's left

Figure 4.9. ***on the one hand/on the other hand,*** **bilateral point sequence**

A more muted example of an indexical form is shown in Figure 4.10. Here, we see a bimanual open-palmed gesture that has moved to the left of the body, rather than a finger point hand shape.

S: Paula, I think, on one hand, we can expect if we found one piece of debris, that there are probably others with it.

G: palms lateral facing inwards to left side of body

Figure 4.10. *on one hand,* bimanual palm-lateral gesture

In addition to the full bipartite expression, speakers rely on other linguistic devices to create utterances that mirror the bipartite structure of [O1H-OOH]. Linguistic examples were given in (41) through (43), above. The analysis of such utterances showed that their bodily enactments mirror the parallelism and hand forms that predominate in the full [O1H-OOH] construction. This is not surprising given that both the fully idiomatic expression, [O1H-OOH] and expressions of contrast that follow the same bipartite model without the idiomatic forms share the same mental construct of CONTRAST, a parallelism in their constructional form, and a shared domain (most often realis). The next two figures illustrate two enactments that take the more generalized contrast construction form. In Figure 4.11, the utterance does not contain an overt marking of contrast in the onset expression, but the second part of the utterance is *on the other hand.* The enactment demonstrates a use of lateral space and symmetrical PUOH form that mirror the profile of Figure 4.5 and Figure 4.6, above, for the full [O1H-OOH] construction.

(a)

S: *Donald Trump's expected visit to Arizona tomorrow, it will spark protests tomorrow. It's already sparking consternation amongst some leading republicans in Arizona*

G: PUOH (cupped), left hand

(b)

But *on the other hand, it's the Maricopa County that invited him*

PUOH (holding pen) with right hand

Figure 4.11. ***on the other hand,*** **PUOH gesture**

Figure 4.12 features a bipartite expression in which neither element is the prototypical [O1H] or [OOH]. Rather, the bipartite form is *it is one way/it is also.* Its enactment begins with a container gesture indexed to the left of the body. This is then held with one hand while the alternate space is indexed on the right.

(a)	(b)	(c)	(d)
S: It is in one way very exciting that this is finally a matter of…	*…high level political discussion.*	*It is …*	*…also at the same time very disconcerting to see the Republican Party and*
G: bimanual container gesture	bimanual PUOH	moving to PUOH	PUOH with both hands

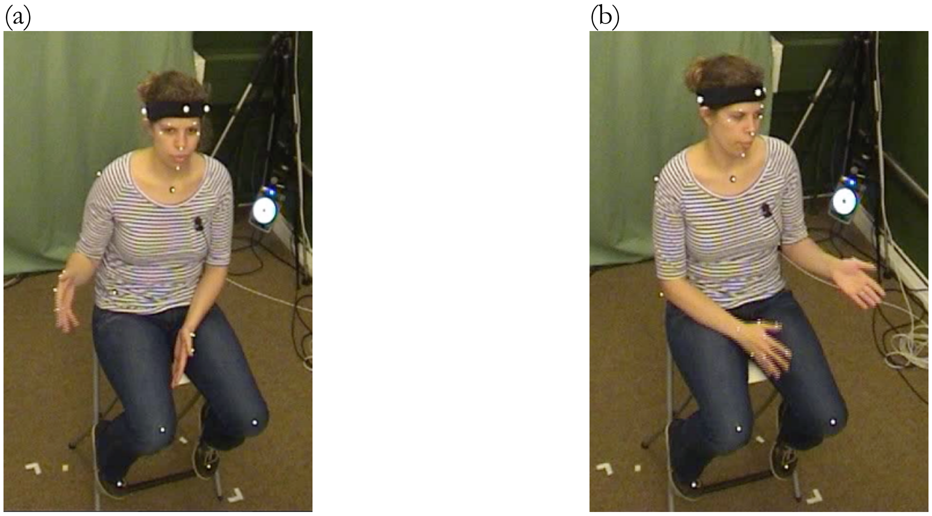
Figure 4.12. Bipartite expression, container and PUOH gestures

4.3.2 Other phrasal and lexical expressions of CONTRAST

Beyond the archetypal [O1H-OOH] and other similar phrasal expressions introduced thus far, the evaluation of options in English can also be expressed with a wide range of other fixed phrases and lexical items. As described in §4.2, lexical comparisons are often set up through specific syntactic constructions that do contrastive work, such as *X versus Y, X but not Y,* and *whether X or Y,* the comparative construction *better than/worse than*, and antonymic noun phrases *(David/Goliath, offense/defense)*. Contrast can also be conveyed with more open, but juxtaposed, syntactic constructions, such as negation (e.g. *is* vs. *isn't*). In the remainder of this section, I introduce the multimodal enactments of a variety of these expressions.

In data gathered in a 3D motion capture experiment that was designed to elicit discourse on the topic of habitual events, one participant compared her habit of nail-biting to the severity of other peoples' nail-biting habit.[81] This is shown in Figure 4.13. The participant indicates that she does bite her nails (*I think that that is something I still do*) and contrasts this with the degree to which other people do it (*there are lots of people who do it way worse*). At the same time as uttering these remarks, she indexes the space to the right of her body to mark her own habit, and subsequently places the nail-biting of other people on the left side of her body. In this construction, it is both the person (*I* vs. *other people*) that is contrasted, as well as the degree to which the speaker and these persons bite their nails. Similarly, in Figure 4.14, the lexical antonyms *offense* and *defense* are indexicalized on alternate sides of the body, both by means of a manual gesture and a head tilt (particularly to indicate *offense* in the screenshot on the right).

(a) (b)



(a)	(b)
S: *I think that that is something I still do…*	*… there are lots of people who do it way worse*
G: palms towards each other, arm movement to speaker's right	same motion to speaker's left

Figure 4.13. *something I do* vs. *people do it way worse,* bilateral gestures

[81] MoCap examples stem from a corpus of conversational data the author recorded in the Natural Media Lab of the RWTH Aachen University in April 2014 with Professor Dr. Irene Mittelberg.

(a)

S: …*Rick Petino playing some <u>defense</u>*

G: cupped hand gesture to left side

(b)

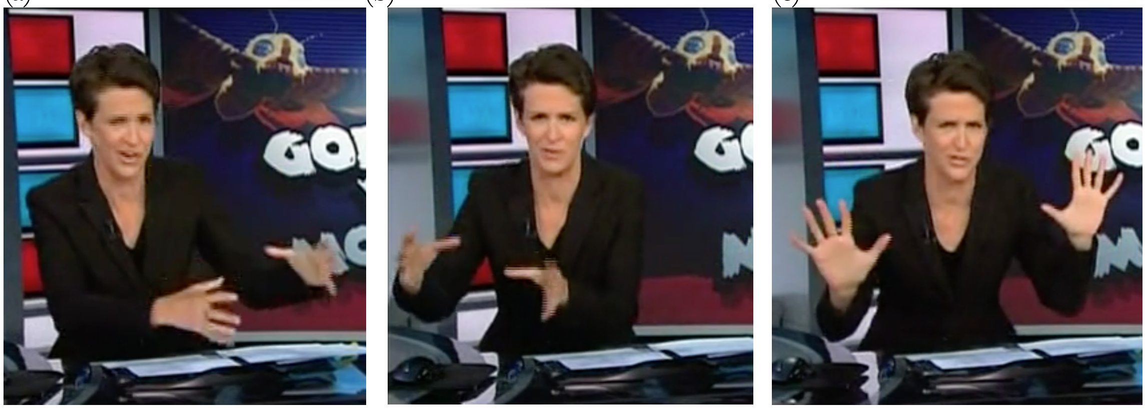
by <u>going on the offense</u>

same hand points, moves to right

Figure 4.14. *defense* vs. *offense,* antonymic contrast with bilateral gestures

The same pattern of space-building characterizes expressions of CONTRAST marked by noun phrases. In Figure 4.15, the noun phrases *Republican news network* and *frontrunner Republican presidential candidate* are placed in contrast to each other. Although these noun phrases are not inherently antonymic, CONTRAST is achieved through the use of the linking adverb *versus* in the linguistic expression. In the body, there is a concomitant anchoring of each noun phrase in the space on opposite sides of the speaker's body. Interestingly, in this example, after placing each noun phrase in its own space on each side of her body, the host closes the contrastive statement by saying "two monsters of roughly the same size", as shown in Figure 4.15(c). She holds both hands in the air and moves them up and down at the same time (i.e. one hand moving up while the other moves down, and vice versa). She thus maintains the indexicalized meaning she has placed in each hand and on each side; it is clear to the viewer which referent is referred to by each hand while she compares their size. (From a multimedia perspective, the comparison that the speaker articulates is highlighted by the projected text behind her referring to the 1992 Japanese science fantasy *Godzilla vs. Mothra*).

(a) (b) (c)

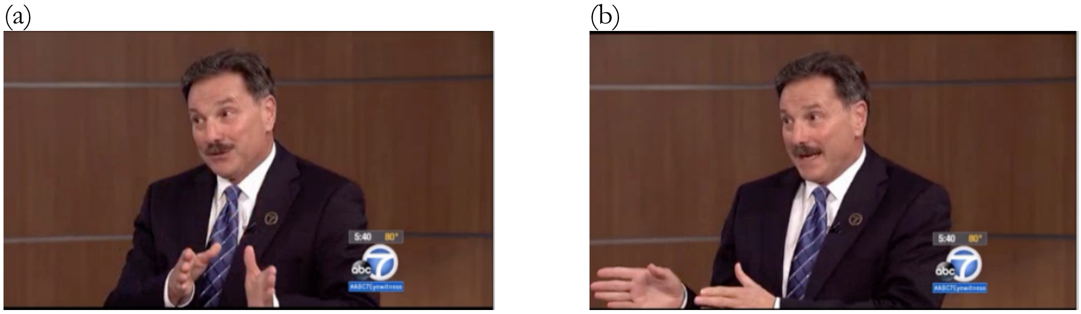


(a)	(b)	(c)
S: *Republican news network versus*	*frontrunner Republican presidential candidate*	*two monsters of roughly the same size*
G: bimanual container gesture left of body	bimanual container gesture right of body	palm forward moving up and down

Figure 4.15. antonymic noun phrases with *versus,* container gesture

The enactments in Figure 4.16 and Figure 4.17 show other antonymic expressions that are enacted using bilateral gestures. While the first tableau shows two contrastive elements, *good news* and *bad news,* in sequence with no other markers of contrast, the example in Figure 4.17 uses the logical operator *or* as the constructional frame for the antonymic nouns *manicure* and *pedicure.* Verb phrases also enact CONTRAST, as shown in Figure 4.18.

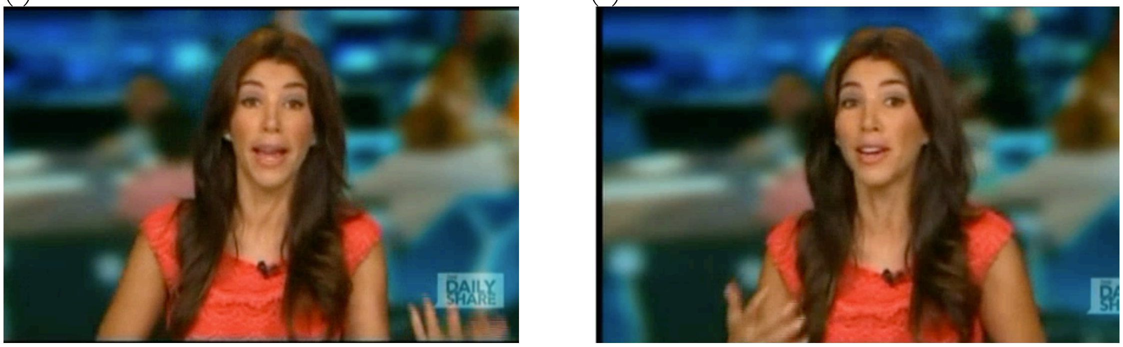
(a) (b)



(a)	(b)
S: *It's kind of a good news*	*bad news situation for them*
G: palms lateral facing inwards to left side of body	palms lateral facing inwards to right side of body

Figure 4.16. *a good news bad news situation,* bilateral indexical gestures

(a) (b)



(a) S: *If you're getting a manicure*

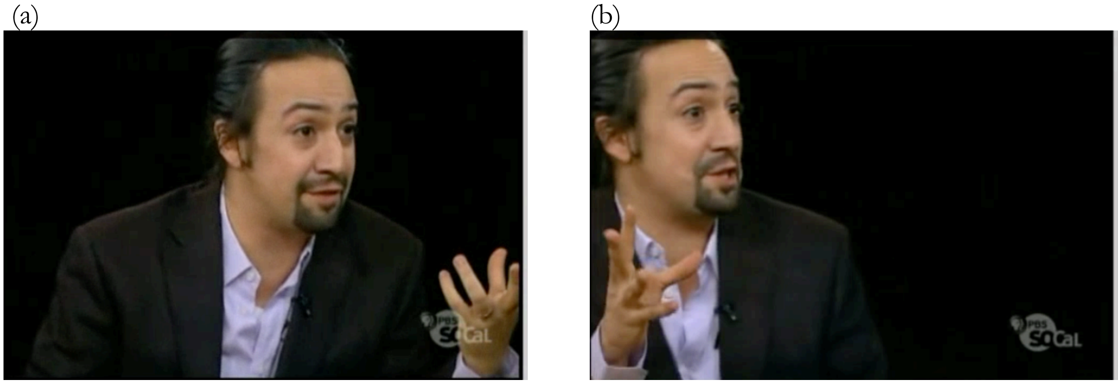
G: left hand palm-up open-hand with posture lean left

(b) *or pedicure or whatever it is*

right hand palm-up open-hand

Figure 4.17. ***manicure or pedicure,*** **PUOH gesture**

(a) (b)



(a) S: *Every biography either is insanely defensive of him*

G: left hand palm-up or container gesture

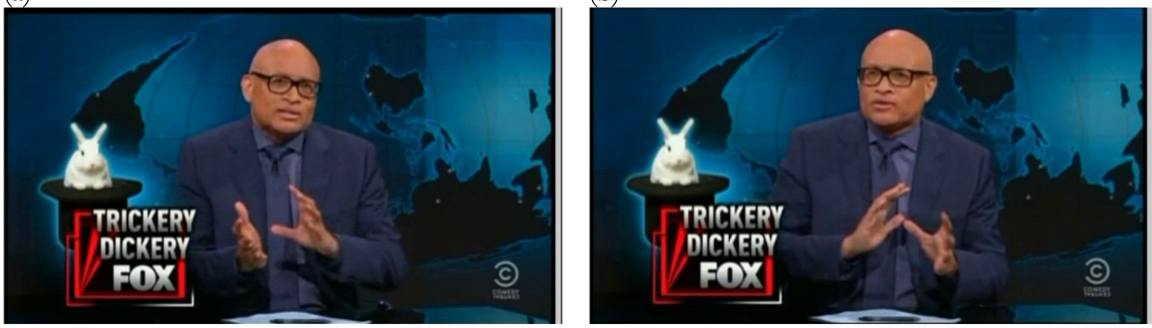
(b) *or vilifies him*

right hand palm forward outstretched fingers.

Figure 4.18. ***insanely defensive of him*** **vs.** ***vilifies him*****, bilateral gesture**

In Figure 4.19, I give an example of a gesture form that does not feature in literature on recurrent gestures, but which by now will look familiar to readers. It is a gesture of alternation characterized by a type of container form in which one hand is facing palm upwards and the other is facing palm downwards. On the expression of the second part of the contrast, the hands switch their vertical orientations. This form can feature a twist in the hands to a greater or lesser degree.

(a) (b)



S: *So how does Fox turn <u>something that isn't</u>* — *into <u>something that is</u>*

G: PTLC gesture tilted to one side — PLTC gesture tilts to the other side

Figure 4.19. *something that is* vs. *something that isn't,* PLTC gesture

Finally, the comparative construction *better than/worse than* completes this survey of contrastive constructions in English. Here, I provide two examples of the typically enacted profile of this expression. In Figure 4.20, the speaker-gesturer poses a question requesting an evaluation of the status of something in the future. He indicates first to his left side with his left hand in a palm-up form, when he asks if it will be *something a lot better than we've got now*, and subsequently raises his right hand in the same form (and reverses the form of his left hand from the palm-up to palm-down position), when giving the alternative possibility, that it will be *something a lot worse than we've got now*. In Figure 4.21, rather than a lateral use of the hands, the speaker-gesturer indicates the contrast expressed in [BT-WT] with a head tilt. In this example, the timing and use of space is slightly varied, in that the speaker gestures with her head to the right when saying *donor cells*, then returns her head to neutral for *equivalent to or better than*. She closes with *or worse than* and indicates to the right again with her head. What is not readily evident in the still shots is the very slight head movement to the left that occurs between *equivalent to* and *better than*. Thus, the major movements of the head are both to the right, with only a slight shift indicating the difference between the neutral, *equivalent*, option and *better than*.

(a) (b)

(a) S: *The question is, is it going to be something a lot <u>better than</u> we've got now*

G: PUOH with left hand, right hand lax

(b) *or something a lot <u>worse than</u> we've got now?*

reversal of gesture form in left hand to palm-down, right hand PUOH

Figure 4.20. ***better than/worse than*****, PUOH gestures**

(a) (b) (c)

(a) S: *We'll be able to answer several questions. One, do cells help compared to placebo and two, are <u>donor cells</u>*

G: rightward head tilt

(b) *equivalent to or better than*

no head movement

(c) *or <u>worse than</u> the child's own cells*

rightward head tilt

Figure 4.21. ***better than/worse than*****, head tilt**

In sum, these examples of the [O1H-OOH] construction and lexical and phrasal constructions of CONTRAST show how speakers enact contrast in speech and the body when they are speaking about topics in the propositional domain. The examples show the tendency for two polar elements to be placed in opposing gesture spaces with PUOH forms, a point handshape, lateral palms facing inwards and moving side-to-side, and indexicalized gestures using lateral space. In the next section (§4.3.3), I present the quantitative data from two case studies in the propositional domain.

4.3.3 Quantitative case studies

I now provide a quantitative analysis of the data. I begin with a comparison of the fixed expression *on (the) one hand/on the other hand* and then present a smaller case study of *better than/worse than.*

On (the) one hand/on the other hand

In introducing the methods in §4.2.2, I described the data analysis of the 100 tokens by both phrase ([O1H] and [OOH]) and construction type ([O1H-OOH], [O1H-exp], and [exp-OOH]). I begin this report on the findings of the case study by describing the data set in more detail, before reporting on findings for degree of enactment, body partitioning, symmetry and laterality, and handshape.

Description of construction types

Of the 50 examples of [O1H], 18 (36%) were paired with [OOH], while 13 (26%) of [OOH] examples were paired with [O1H]. This resulted in 31 tokens of the full [O1H-OOH] construction. The remaining 32 tokens of [O1H] were paired with a variety of morphosyntactic devices that resolved the expression, as follows (in descending order of frequency with number of tokens in brackets): *and* (e.g. *and two, and secondly, and here*) (10); *but* or *but then* (9); *at the same time* (4); and single occurrences of resolutions beginning with *also*, *still,* and others. Only seven examples featured no resolution marker. The 37 tokens of [OOH] were introduced by a much more heterogeneous set of expressions than was the case for the resolution of [O1H]. Aside from the 12 examples of [OOH], I found only one phrase more than once: *yes* introduced the first part of the utterance three times. An example is given in (53).

(53) *Yes, there were significant problems. On the other hand, if you talk to the major veterans organizations, the American Legion, the VA, at the end of the day they will tell you: 'Provide high quality health care'.* [82]

The majority of the tokens of [OOH] were not preceded with any marker in the upstream speech. The utterances in (54) and (55) show typical contexts of use.

(54) *Donald Trump is under water but on the other hand he's dominating polls around the country.* [83]

(55) *We think of issues and we want a more sophisticated approach. On the other hand, this is the second phase of the campaign. The summer is the fun part….* [84]

However, there were five utterances that marked epistemic stance in the upstream speech prior to [OOH], through which speakers expressed knowledge of an event or weighed in about an event's potential realization. The upstream stance markers included the epistemic adverbials *absolutely, certainly, of course,* and phrasal expressions featuring verbs of cognition such as *I believe* and *I can well understand.* Two examples are given here, with the evidential phrase(s) underlined in each:

(56) *I mean, I think that certainly the negotiators have the experience of the past deals and so that has been incorporated into these deals. On the other hand, when I hear the president talk about driving American cars in the streets of Tokyo, I wonder what planet he is on.* [85]

(57) *I believe our federal government has a role to play in preventing that. On the other hand, I don't believe the federal government should nationalize our banks.* [86]

---

82 Red Hen: 2015-10-28_2100_US_CNN_Situation_Room,2559

83 Red Hen: 2015-09-07_1000_US_FOX-News_Fox_and_Friends,1097

84 Red Hen: 2015-09-17_0200_US_MSNBC_The_Last_Word_With_Lawrence_Odonnell,534

85 Red Hen: 2015-10-08_0330_QA_AlJazeera_Inside_Story,1025

86 Red Hen: 2015-10-10_1900_US_CNN_Newsroom,3300

Degree of enactment

As reported above and summarized in Table 4.5, 92% of the utterances for both [O1H] and [OOH] were enacted through manual gesture, upper body movement, or both.

Table 4.5. Degree of enactment for [O1H] and [OOH]

	Enacted	Not enacted
[O1H]	92	8
[OOH]	92	8
Total	100	100

I first present findings organized by construction type for body partitioning, symmetry, and laterality, before reporting on results for handshape and orientation.

Symmetry and laterality

The results for symmetry and laterality are shown in Figure 4.22. As the chart shows, [O1H-OOH] are highly symmetrical (81%) and highly lateral (81%). These findings confirm the profile that we have observed qualitatively thus far in the numerous examples given, in which speakers gesture first to one side (for example, using a PUOH or palm-lateral gesture) and then to the other (cf. Figures 4.5, 4.6, in fact, the majority of the examples given above); however, both [O1H-exp] and [exp-OOH] expressions show a reduction in the degree of laterality and a greater reduction in the degree of symmetry. These findings are also displayed in Figure 4.22. Results of a chi-squared test for significance showed that these differences in distribution for laterality are not significant, while the difference in degree of symmetry is (p<.001). A chi-squared post hoc test showed that the difference in symmetry for [O1H-OOH] as compared to [exp-OOH] is significant.

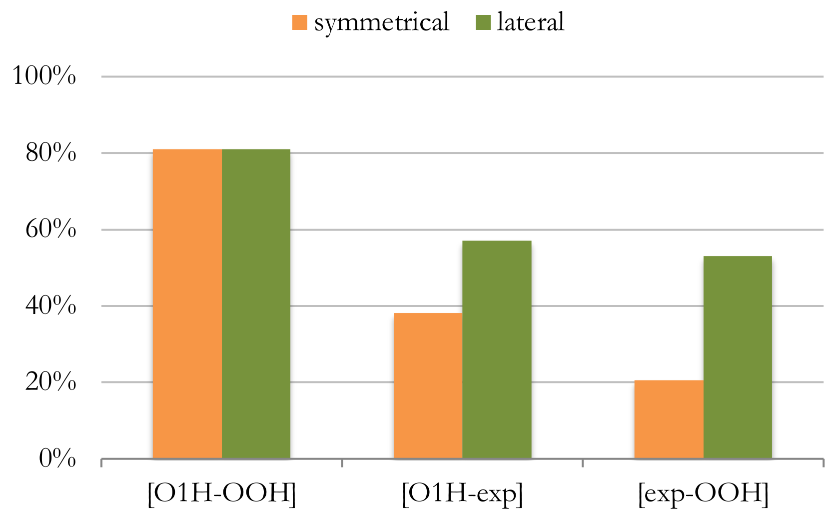


Figure 4.22. Symmetry and laterality by construction type

These findings suggest that laterality is a key to the embodiment of CONTRAST, while symmetry appears to be more closely associated with the specific profile of the linguistic construction. Laterality plays a relatively strong role in each construction type (50%-81%); however, the reader will recall from the description of the data set given above, that the construction types vary widely in their degree of fixedness in the linguistic signal. While [O1H-exp] constructions are resolved by a small subset of utterances such as *but, but then,* and *at the same time,* [exp-OOH] constructions are preceded by a wide range of utterances, none of which particularly mark CONTRAST (other than the affirmative marker *yes…,* as noted). To put it another way, the linguistic signal aligned with [exp-OOH] is far more heterogeneous than the other two related constructions. There is a commensurate reduction in the degree of symmetry with the reduction in homogeneity in the linguistic signal for [exp-OOH]. I return to this observation in the discussion in the summary that ends this section (§4.3.4).

Handshape and handedness

Having established the degree of enactment for [O1H-OOH], which articulators are involved, and the role of symmetry and laterality in the body signal for [O1H-OOH], I now present the findings for the remaining elements of the gesture profile, namely handedness and handshape.

Handshape is known to be a significant contributor to the semantic and pragmatic meaning of manual gestures. As described in the Methods chapter (§2.4.2), handshape consists of two parameters: hand configuration and palm orientation. I begin by presenting the findings for palm orientation, given in Tables 24 and 25. Statistical analysis of palm orientation showed no significance, so I present quantitative summaries here. As the tables show, the three most dominant palm orientations, palm-down, palm-up, and palm-lateral, account for 82% of the gesture forms for [O1H] and 94% for [OOH]. (The remaining 18% of [O1H] and 6% of [OOH] enactments consisted idiosyncratic gestures that are not standard forms and upper body only enactments and are not included in the table).

Table 4.6. Palm orientation for [O1H] in [O1H-exp] constructions

Palm orientation	[O1H] %	[O1H] raw freq.	Detailed annotation
Palm-down	18%	9	9 PDdiTC[87]
Palm-up	28%	14	4 PUOH, 9 PUdiTC[88], 1 PUpurse
Palm-lateral	36%	18	15 PLTC, 3 PLTB
TOTAL	82%	41	

[87] PDdiTC = palm-down diagonal towards-centre. This is a palm-down form in which the thumb side of the hand is angled slightly upwards, rather than the palm facing downwards. I include this level of detail here to demonstrate that all instances of the palm down for [O1H] were in this form, as compared to the palm down forms in [OOH].

[88] PUdiTC= palm-up diagonal towards centre. This is a palm-up form with the thumb side of the hand angled slightly upwards, rather than the palm facing straight up.

Table 4.7. Palm orientation for [OOH] in [exp-OOH] constructions

Palm orientation	[OOH] %	[OOH] raw freq.	Detailed annotation
Palm-down	24%	12	11 PD, 1 PDdiTC
Palm-up	18%	9	3 PUOH, 6 PUdiTC
Palm-lateral	52%	26	20 PLTC, 6 PLTB
TOTAL	94%	47	

The palm-lateral form dominates for both the [O1H] construction (36% of the total) and [OOH] (52%). The majority of the palm-lateral forms were executed with palms oriented towards centre (PLTC), as shown in the two-handed container forms in Figure 4.7 and Figure 4.8. The palm-down orientation, conversely, accounted for 18% of [O1H] and 24% of enactments of [OOH]; however, a closer look at more detailed levels of palm orientation annotations reveals a qualitative difference in the palm-down gestures. This can be seen in the more detailed annotation levels provided in the rightmost column of the tables. The [O1H] construction featured nine palm-down gestures, all of which were bimanual and executed with palms slightly on a diagonal angle (palm-down, diagonally towards-centre, or PDdiTC), which creates a somewhat modified container form (compared to the prototypical PLTC container gesture, with palms oriented laterally). What this means is that the typical palm-down gesture for [O1H] was a container gesture, as pictured in Figure 4.23. [89] This differs greatly for the profile [OOH]. [OOH] palm down gestures featured palms oriented downwards, not on a diagonal. All but one of the 12 palm-down gestures for [OOH] were fully palm-down and executed with just one hand. This yields the gesture profile shown in Figure 4.24.

[89] This instance was coded as a PDdiTC, for palm-down, diagonally, towards centre.

S: *Fareed zakaria, you didn't know what you were going to get by asking that question. We have seen Hillary Clinton <u>on one hand</u> going around, "I apologize, I own it", wanting to move on.*

G: Container form

Figure 4.23.
***on one hand,* palm-down gesture**

S: *If there's water on Mars, that means there might be life on Mars, which would call into question everything mankind believes about its place in the universe. It's unbelievably exciting. <u>On the other hand,</u> now we have to worry about the rovers getting wet.*

G: Palm-down open-hand

Figure 4.24.
***on the other hand,* palm-down gesture**

Upper body movement

To close this presentation of findings for [O1H-OOH], I present results for the movements of the upper body for each part of the construction. As the findings in Table 4.8 show, the majority of upper body for [O1H] is enacted by head movement, with 10/12 instances with upper body movement featuring the head (8 tilts, 2 turns). A further look at the findings showed that 9/12 head movements, or 75%, were aligned with gesture. In comparison with [O1H], the [OOH] enactments featured a much higher degree of upper body involvement (32/50 featured upper body compared to 12/50 for [O1H]). Of the 32 instances, 29 (91%) were co-articulated with gesture.

Table 4.8. Frequency of upper body movement for [O1H] and [OOH]

Upper body	[O1H]	[OOH]
Head	10	27
Shoulder	1	4
Eyebrow raise	0	6
Torso	4	11
Total utterances with UB movement[90]	12	32

Body partitioning

Here, I report on the ways in which the three different construction types ([O1H-OOH], [O1H-exp], and [OOH-exp]) were articulated with regards to the division of labour between the hands and upper body (e.g. by the hands only, the upper body only, or by the hands and upper body). In reporting by construction type, the annotations capture kinesic movements for both sides of the utterance (as compared to the data presented in the previous sections by [O1H] and [OOH]). For example, if one part of an expression was gestured and the second part involved movement in the head only, this would be annotated G&UB, for gesture and upper body. It is worth noting that in the majority of enactments, the articulator remained consistent for both sides of an utterance. That is, if [O1H] was enacted with gesture only, the second part of the construction was generally articulated gesturally as well (though the gesture form, e.g. hand shape, may have differed). In the same way, if the head was not involved in the first part of the expression, it was usually not recruited during the second part of the utterance. For clarity, the raw data for body partitioning that corresponds to Figure 4.25 is given in Table 4.9.

[90] Some utterances featured more than one upper body movement, such as an eyebrow raise and a head nod, or a tilt of the head and a torso shift; thus, the total of the values in the articulator columns exceed the number of total utterances. The total given is the total number of utterances containing upper body movement for each construction.

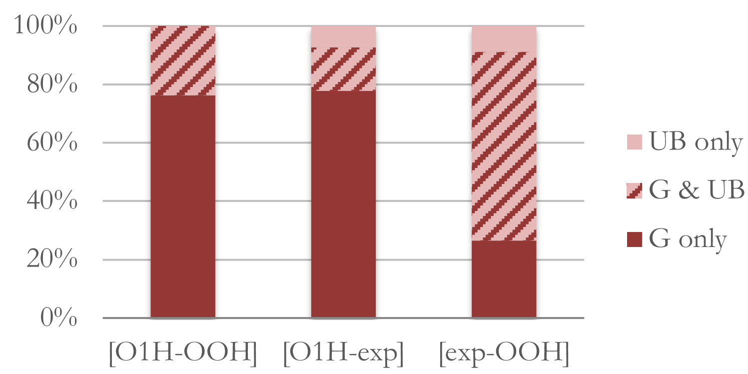


Figure 4.25. Body partitioning profile by construction type

Table 4.9. Body partitioning for [O1H-OOH] by construction type

	[O1H-OOH]	**[O1H-exp]**	**[exp-OOH]**
G	16	21	9
G and UB	5	4	22
UB	0	2	3
Total	21	27	34

As Figure 4.25 shows, the full construction, [O1H-OOH], and the partially fixed [O1H-exp] construction pattern very similarly. These two constructions feature gesture in 76% and 78% of enactments, respectively. For [O1H-OOH], gesture is always present, even in the 24% of enactments in which the upper body is active as well. In comparison, for [O1H-exp], a small percentage are expressed with upper body movements only. This is to say that, when a speaker sets up a CONTRAST using [O1H], regardless of whether she resolves it with [OOH] or an expression such as *at the same time*, the enactment is most frequently expressed only in the hands.

For the semi-fixed expression [exp-OOH], the body partitioning profile is quite different. The upper body is active in 74% of the enactments, with only 26% exclusively gestured in the hands. A chi-squared test showed these findings to be statistically significant. A chi-squared posthoc test revealed that a greater than expected number of utterances for

[O1H-OOH] and [O1H-exp] were enacted with gesture only and fewer than expected involved the upper body. For [OOH-exp] the converse was true, with fewer than expected enactments featuring only gesture as compared to either gesture and upper body and upper body only. [91]

As the reader will recall, [exp-OOH] is the most heterogeneous of the three construction types and lacks an explicit marker to introduce the CONTRAST upstream in the linguistic utterance. I suggest that it may be that the upper body is recruited in these contexts to add more contrastive force, in essence to make up for the lack of marking in the linguistic signal. This subjective force of the upper body would align with other literature on functions of upper body gestures. I return to the correlation between body partitioning and subjectivity in the discussion at the close of this section (§4.3.4).

Semantic prosody

As I introduced in the introduction to this chapter and in Chapter 1, previous research has suggested a lateral bias that is related to semantic prosody or valence. This has been seen to play out in handedness, i.e. a speaker encodes positive sentiment with her dominant hand and negative evaluation with her left (Casasanto & Jasmin 2010). In this study, I coded each element of each contrastive pairing for any discernible semantic prosody in the linguistic utterance. I then compared these annotations with the findings for laterality. For the 14 examples in the data set with notable semantic prosody, eight featured positive value on the right and negative on the left, while six featured negative value on the right and positive on the left. While handedness could not be known for speakers in the data set, it was expected that,

---

[91] For the chi-squared test, to avoid 0-cells, the values for G&UB and UB were collapsed.

on balance, most speakers would be right-handed and thus associate positivity with the right side. Recognizing that, without interviews that could ascertain degree of positive or negative valuation or knowing handedness, both measures used here are coarse and prone to inaccuracy. Having said that, I found no evidence of associations between handedness and semantic prosody. I am not the first researcher to find that Casasanto's results did not hold over their data. In his analysis of Obama's public discourse, Guilbeault (2017) discusses the semantic polarization of space in Obama's gestures. Guilbeault found no correlation with sidedness and semantic prosody. He does not rule out that Obama's placing of positive content to his left is due to his left-handedness, but concludes that, regardless of which side the gesture is placed on, Obama is "carving the gesture space into opposite value poles" (Guilbeault 2017: 427). Indeed, the expression of CONTRAST in my search returns seems to rest more in the placing of gestures in different, visually contrastive spaces than it does in the use of gesture space based on a positive or negative valuing of that option.

Summary: On (the) one hand/on the other hand

Before moving on to the cases study of other phrasal and lexical expressions of CONTRAST, I provide a summary of the most important findings for [O1H-OOH]. The data presented in this quantitative analysis support the tendencies that were clear in the qualitative examples given earlier in the chapter. The prototypical, fully-fixed expression *On (the) one hand/on the other hand* is predominantly gestured in the hands using a PUOH form that is symmetrical and anchored in opposing sides of the gesture space. The semi-fixed expressions, [O1H-exp] and [exp-OOH], have significantly reduced degrees of symmetry in their gestures and also show a reduction in the bilateral use of space compared to the full expression. Notably, these findings for symmetry and laterality reflect the linguistic profile of each construction. As mentioned in the introduction of this section (§4.3.1), [O1H-OOH] is the most homogeneous of the three constructions in its lexico-syntactic form and in its co-speech embodiment, given its status as a fixed expression. The semi-fixed construction *[exp-OOH],* had the most heterogeneous linguistic profile, with no evidence of repeated expressions in my sample of utterances. This suggests an iconic relationship between the degree of fixedness in the linguistic utterance and the degree of fixedness in the kinesic enactment.

In this study, I also reported on differences in handshape for each construction. These differences are found to be most significant for the palm-down gestures, which are the least-preferred for [O1H-OOH] and which have variable hand configurations for the semi-fixed constructions. That is, palm-down gestures associated with [O1H-exp] tend to participate in the container gesture form, whereas for [OOH] they are associated more frequently with the full palm-down gesture, which has been shown to be associated with negation and concession (Kendon 2004; Harrison 2018).

I close this summary with a brief discussion of the findings for body partitioning and how these may relate to the degree of subjectivity inherent in [O1H-OOH]. The body partitioning data give a coarse-grained impression of which articulators are involved in each utterance. The findings show that body partitioning is similar for the full construction and the [O1H-exp] variant. This overlap shows that when a speaker begins an utterance with *on (the) one hand* – whether it is completed by *on the other hand* or another expression, the speaker is most likely to use gesture to mark both sides of the utterance. The last construction type, *[exp-OOH],* seems to function differently. It is used when a speaker is in a train of speaking and has not set up a contrast at the onset of a discourse segment. When the speaker subsequently wants to make a contrastive juncture in the discourse stream, the speaker uses [OOH] paired with *both* gesture and upper body. (Recall that 74% of [exp-OOH] enactments involved the upper body, compared to 24% and 22% for the full expression and [O1H] variant, respectively). This is where I would suggest there is a connection with subjectivity (despite the inconclusiveness nature of the annotations for other subjectivity markers). [OOH] is effectively doing all the contrast marking 'work' by itself. There has been no advance signaling yet in the discourse that there is a contrast coming. Therefore, a higher degree of force is required to create the contrast, which may explain the greater involvement of the upper body as a contrast marker for [OOH]. The handshape analysis would seem to support this interpretation. Recall that [OOH] is associated more frequently with palm-down gestures, which also happen to be associated with concession and negation. This, too, is a marker of the subjective construal associated with the second part of the contrastive utterance.

Having explored the [O1H-OOH] construction, I now conclude this series of case studies with a quantitative investigation of *better than/worse than*.

Better than/worse than

In this section, I describe the findings from the case study of the phrasal construction *better than/worse than* [BT-WT]. Some examples of linguistic utterances which characterize [BT-WT] are given in (58) through (61).

(58) *Things are actually better than Trump supporters had hoped and they are worse than many Trump opponents feared.*[92]

(59) *We don't want our kids, kids of color, to go to school and be treated better than anybody else but not worse than anybody else.* [93]

(60) *We'll be able to answer several questions. One, do cells help compared to placebo and two, are donor cells equivalent to or better than or worse than the child's own cells.* [94]

[92] Red Hen: 2018-01-28_0000_US_CNN_The_Axe_Files_With_David_Axelrod,96

[93] Red Hen: 2018-12-18_1900_US_CNN_CNN_Newsroom_With_Brooke_Baldwin,2886

[94] Red Hen: 2017-07-08_1830_US_CNN_Vital_Signs_With_Dr_Sanjay_Gupta,1540

(61) *Is this version better or worse than the house version?*[95]

As these examples show, [BT-WT] can occur in a variety of lexico-syntactic frames. In (58) and (59), *better than* and *worse than* are followed by clause fragments. In (60) and (61) they are followed by a noun phrase. Among the 25 tokens in the study, there were seven cases in which the [BT-WT] construction was part of a question, as in (61).

Turning now to the kinesic profile of *better than/worse than*, I report on the degree of enactment, the role of symmetry and laterality, the conventionalized handshapes, and, finally, body partitioning. [BT-WT] was embodied in 18 of the 25 search returns to give a degree of enactment of 72%. The co-speech enactments were dominated by symmetrical and lateral enactments. In 14/18 (78%), the enactment featured both symmetrical and lateral movement either in the head or hands. There were two enactments that were neither symmetrical nor lateral (e.g. with a head nod on *worse than* and no head movement or gesture on *better than*) and four that featured either symmetry, or laterality, but not both (e.g. both phrases are indicated with a head tilt to the right, with a return to neutral in between). This was coded as symmetrical, due to the consistent use of the head-tilt, but not lateral, as there was no use of both left and right space.

The profile for the [BT-WT] construction features two-handed gestures in which both hands are moving synchronously, rather than sequentially (i.e. both hands are involved for both *better than* and *worse than,* as compared to a sequence of PUOH on the right and then the left, as was the case in [O1H-OOH]). This accounted for eight of the 13 tokens with gesture. The dominant form of the symmetrical and lateral gesture can be seen Figure 4.20, above.

[95] Red Hen: 2017-06-22_1500_US_CNN_Comey_Senate_Hearing,1553

(Here, it may help to view the video to catch the synchronous movement, which gives a 'topsy-turvy' nature to the gesture. See Appendix A.)

Hand shape for [BT-WT] consisted of the range of forms shown in Table 4.10. Palm-lateral gestures were the most frequent.

Table 4.10. Palm orientation, [BT-WT]

Palm orientation	Raw frequency	Gesture sub-type
Palm-down	1	1 PD
Palm-up	4	2 PUdiTC, 1 PUOH, 1 PU-cupped hand
Palm-lateral	5	5 PLTC
Other	3	1 fist, 1 point, 1 palm-away
TOTAL	13	

Upper body movement for [BT-WT] was characterized by head movement such as the example given above in Figure 4.21. In 11 of the 13 enactments with upper body movement, the head was involved. Table 4.11 shows the distribution for all upper body movement. Again, the sum is not cumulative as some enactments feature movement in multiple articulators.

Table 4.11. Upper body movement for [BT-WT]

Upper body	[BT-WT]
Head	11
Shoulder	0
Eyebrows	4
Torso	1
Total enactments with UB	13/18 72%

The [BT-WT] construction featured a relatively even distribution of involvement by partitioned articulators. Body partitioning findings are represented in Figure 4.26. Of the 18 total enactments, 5 involved gesture only (G); 8 were enacted with gesture and upper body (G&UB); and the remaining 5 involved upper body only (UB).

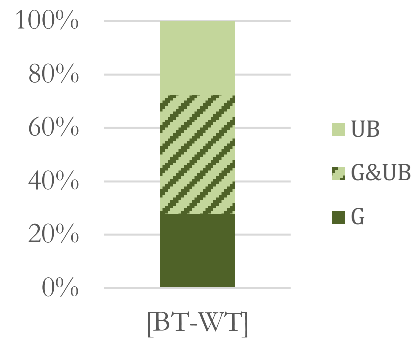


Figure 4.26. Body partitioning profile for *better than/worse than*

In sum, the findings of the case study presented here suggest that [BT-WT] is most likely to involve both gesture and upper body movement. The multimodal profile consists dominantly of two-handed, symmetrical, lateral gestures in the hands, often accompanied by a head tilt.

4.3.4 Summary: Real world CONTRAST

In this section, I have presented findings that demonstrate the ways in which CONTRAST is enacted in English when marked by expressions ranging from *on (the) one hand/on the other hand* to expressions that are set up in a similar symmetrical syntax, such as *in one way… it is also,* and in lexical expressions, including the semi-fixed antonymic adverbial expression *better than/ worse than*. CONTRAST is embodied regularly through bimanual PUOH gestures on each side of the body, by container gestures, by points, and co-aligned head tilts. For the most part, the enactments for these constructions index one side of contrast space and then the other, by means of hands and head tilts. These enactments create highly imageable representations of a comparison or evaluation of two objects or ideas, which happen to occupy different spatial locations in gesture space much as they occupy different positions in cognitive 'space'.

In the next section, I illustrate the embodiment of the CONTRAST frame in the irrealis world. I suggest that these enactments also show iconic representations of the CONTRAST schema. In the irrealis domain, I investigate whether the notion of iconic representation and indexical shifts can be expanded to more schematic representations such as those of the upper body, rather than manual gesture.

4.4 Possible world CONTRAST

In the previous section, I illustrated enactments of CONTRAST in the propositional domain that were expressed through fixed and semi-fixed phrasal and lexical constructions. Most of the examples were predicated in the objective sphere. Given their function is to express CONTRAST, they could be considered in some ways to be subjective because the speaker is pitting one idea against another; however, for the most part, the speakers were reporting on actual events or comparing entities in the real world. In this section, I present evidence from conventionalized expressions of CONTRAST in hypothetical scenarios. I introduce more subjective and internal evaluations, such as conditional constructions, in which the speaker considers alternate mental spaces. In the propositional domain, the findings suggested that depictions of CONTRAST are manifest in specific hand shapes and orientations, in addition to the bilateral use of physical space. As I show in this section, the cognitive domains feature a subtler, more schematic signaling of CONTRAST in the body. The enactments presented here suggest that there is a commensurate use of gesture space to set up contrast, but that these spaces are marked by upper body movement. The body articulators that seem to play a greater role here are the head – through series of tilts, the shoulders, and slight shifts in hand form, rather than the full-blown recurrent gesture forms that characterized propositional CONTRAST.

In §4.4.1, I present an example of a discourse sequence in the irrealis domain that is anchored by the bipartite phrase *should I/shouldn't I* ([SI~SI]). I then present a quantitative case study of [SI~SI] (§4.4.2), before summarizing the findings for the enactment of CONTRAST in possible worlds (§4.4.3).

4.4.1 *Should I or shouldn't I?*

The following excerpt captures an episode in a highly subjective mental space. The speaker-gesturer asks herself what an appropriate response would be to a situation she is presented with. The text features numerous markers of stance – rich in modal, interrogative and conditional markers. I first give the full text of the utterance to provide context. The entire enactment is captured in the series of screenshots presented in Figure 4.27, with the text given in (62) to (65) as well.[96]

(62) *Wow, should I be flattered? Should I be outraged? Should I be insulted?*
(63) *Is this… Should I do it? Should I not do it?*
(64) *Well, if I did it, what would it look like? You know.*
(65) *And I mean, these are new kinds of things for anybody to have to think about.*

Throughout the utterance, the speaker constantly shifts her upper body in alignment with the options that she is considering. Each alternative is enacted with a commensurate shift in one or more articulators. The head tilts, the eyebrows raise and then lower, the shoulders move upwards and then downwards, and gaze shifts strongly to the side, and also upwards. The effect is that the speaker-gesturer creates a constant marking in her body that mirrors the marking of each of these contrasts in her speech. The use of gesture is minimal, with hand movement seen only in the rhetorical question underlying the utterance, *Well if I did, what would it look like?* shown in the final pair of screenshots, Figure 4.27(f) and (g). Even here, the hand

[96] Examples (62)-(65) are from Red Hen: 2016-04-05_0100_US_KOCE_The_PBS_Newshour,3102

shape is not fully formed, as they were in the previous section. Rather, a loosely held palm towards the speaker relaxes into a relaxed gesture, turned slightly upwards.

Looking in more detail, in the first series of expressions in (a) to (c), the speaker enumerates possible emotional responses: *Should I be flattered? Should I be outraged? Should I be insulted?* With each of these, the speaker tilts her head and shifts her gaze upwards in the same direction – first to her right, then to her left, then returning to neutral. In this series, the speaker enacts the difference between the present space and each conditional space. Conditionals are inherently about alternate spaces. In the previous section on real world contrast, I proposed that the gestures first on one and then on the other side of the body create meaning by simultaneously building and indexing a space on each side of the body. The head tilt in this series serves the same function, namely, to create and reference alternate spaces.

In the next series of screenshots, Figure 4.27(d) and (e), the movements that accompany the utterance of *Should I do it? Should I not do it* are also of interest. The modal verb *should* and the interrogative construction both signal the irrealis domain. Three articulators are involved in the utterance, all in the upper body. The first part of the utterance is aligned with a movement of the shoulders and eyebrows upwards and a glance to the left, while the alternative *Should I <u>not</u> do it* is aligned with a downwards movement of the shoulders, a head tilt and eyebrow movement downwards. The contrast in this expression is driven by the negation – indeed from a syntactic perspective, that is the only difference between the two parts of the utterance. In the mental spaces framework, negation has been described as the archetypal example of alternativity: "The negative particle *not* is thus said to set up two alternative spaces, rather than just one: the negative space described in the sentence and its positive alternative" (Dancygier 2012: 69). Here, seeing the alternativity mirrored in the body supports the claim –

formerly made based on mono-modal (i.e. speech/text) data alone – of negation as an inherently contrastive device.

Returning to Figure 4.27, the side-to-side shifts continue as the speaker utters the final conditional, introduced by the '*if*'-statement, *Well, if I did it*. In this case, in addition to the posture and shoulder movement, there is a shift in hand position from facing the speaker to a more relaxed, upwards-oriented hand position. The shift towards a PUOH-like hand form could indicate a presentation of the rhetorical question *what would it look like?* The key point, however, is that every contrast in the speech stream parallels a shift in the enactment.

Now, the reader may be thinking that this is not what we all do in everyday speech. Granted, the example in Figure 4.27 does appear at first glance to be highly exaggerated; however, an investigation of corpus data for other examples of *should I/shouldn't I* yielded similar results, which I describe in the next section.

(a)

S: *Should I be flattered?*

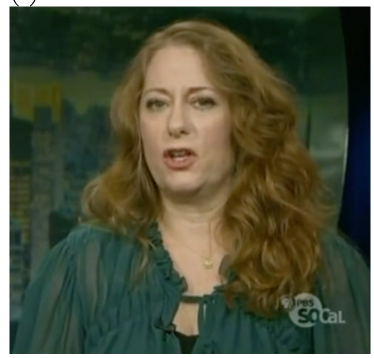
G: gaze upwards, head tilt (R)

(b)

Should I be outraged?

gaze upwards head tilt

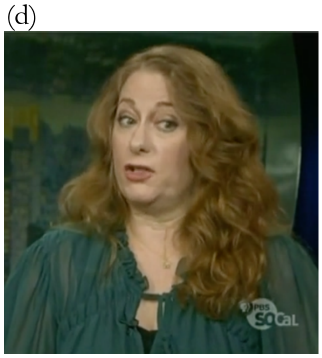
(c)



Should I be insulted?

gaze and head back to neutral

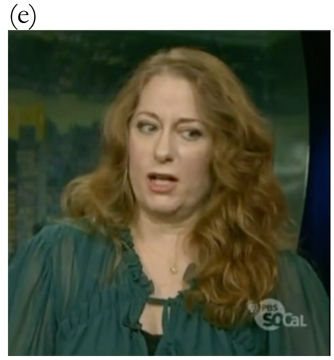
(d)



S: *Should I do it?*

G: shoulders up, eyebrows up, glance left

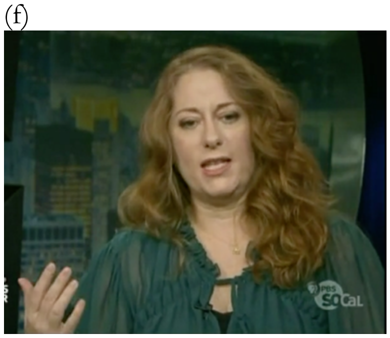
(e)



Should I not do it?

shoulders, head tilt, eyebrows down

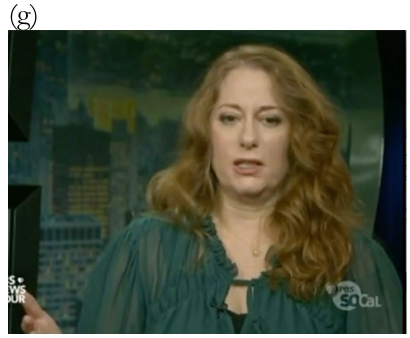
(f)



S: *Well if I did it…*

G: palm facing speaker, head tilt right, right shoulder moves down

(g)



…what would it look like?

palm lowered into PUOH, posture tilt back to centre

Figure 4.27. Contrastive enactments in irrealis marking

4.4.2 Quantitative study: *Should I/shouldn't I*

To explore whether the kind of exaggerated lateral expression seen in the series of utterances shown in Figure 4.27 characterizes the irrealis domain more broadly, I completed a small case study examining the enactments aligned with the utterance pair *should I/shouldn't I* ([SI~SI]). The study was conducted using search returns from Red Hen and it adhered to the general methods outlined in Chapter 2. To search for the target utterance, the search parameters in Red Hen were set to search for *should I* within 10 words of *shouldn't I*. Despite 174 search returns, the query yielded only 69 non-duplicated items. Of these, fewer than 20 met the basic criteria of being in a spontaneous environment. In order to retain the maximum number of data points, I broadened the criteria for acceptable interactional settings to include, for example, a congressman speaking in Congress and political panelists speaking to a host on camera, rather than to a host in the same room. Using these criteria, 13 search returns qualified for consideration, all of which were enacted in the body in some way. Table 4.12 shows the body partitioning data of these 13 enactments. (Given the low number of tokens, body partitioning is given as a table rather than a graph as for the previous case studies). As shown in Table 4.12, 10 enactments featured upper body movement. Table 4.13 provides the distribution of upper body articulations for [SI~SI]. Here, it is clear that the head plays a dominant role. Of the eight head movements, five were tilts. In four additional search returns from Red Hen, the chyron fully or partly obscured the hands. I therefore did not include them in the tables above. It is worth noting, however, that all of these four examples in which the hands were not visible featured a bilateral (e.g. left and then right or right and then left) head tilt. This seems to be a dominant feature of the [SI~SI] construction.

Table 4.12. Body partitioning for *should I/shouldn't I*

Body partitioning	[SI~SI]
Gesture only	3
G&UB	4
UB only	6
Total	13

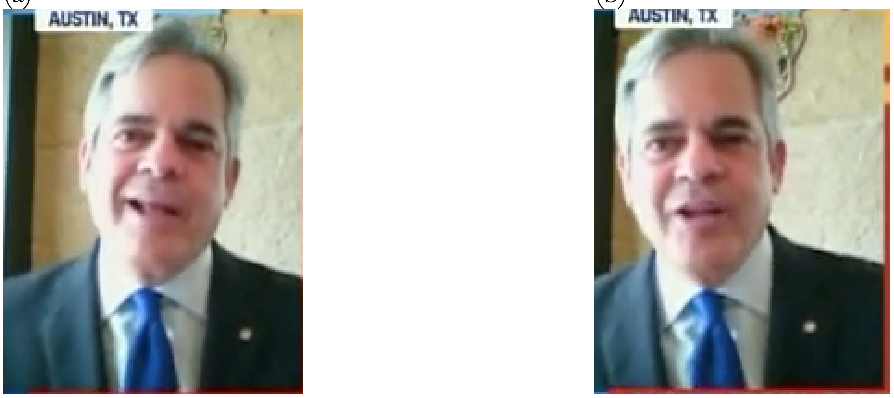
Table 4.13. Upper body movement for *should I/shouldn't I*

Upper body	[SI~SI]
Head	8
Shoulder	0
Eyebrows	0
Torso	2
Total	10

For the seven enactments that featured gesture, five had gesture forms that we have seen already in this chapter, including bilateral PUOH forms and a sequence of PUOH/PDOH that alternates from one side of the body to another. There was also one instance each of pursed hand shape executed from side-to-side with one hand and a sequence of points from in front of the speaker's body to behind him.

Finally, symmetry and laterality also played a role in the enactments of [SI~SI]. In 11 of 13 of them, the gestures or head movements were symmetrical. This entailed a gesture that was either the same or mirrored for each utterance. There was a high degree of laterality (10/13, 77%). This meant that most of the symmetrical enactments were mirrored from side-to-side, e.g. a head tilt to one side and then the other or a PUOH form with the left hand and then the right. Of the three that were not lateral, two featured gestures on the sagittal plane (e.g. a point forward and then back) and one featured an instance in which both phrases, *should I* and *shouldn't I,* were enacted with a head nod (coded as symmetrical but not lateral).

In sum, *should I/shouldn't I* was enacted with a high degree of symmetry and laterality in either the head or hands, with the head playing a dominant role. Below, I provide screenshots of head tilts from the search returns that range from a more subtle tilt shown in Figure 4.28 to the more pronounced tilts shown in the following two figures.[97]



(a)

S: A package shows up, anything at all, we want them to call 911. We want them to pick up the Phone. If there's any doubt in their minds, <u>should I call…</u>

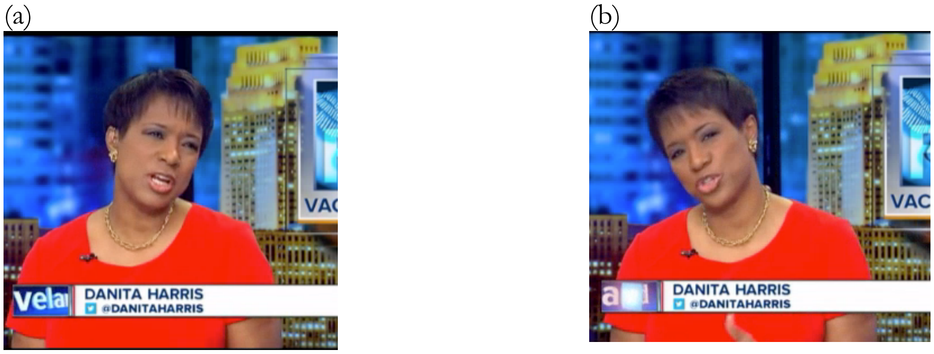
G: Head tilt left

(b)

<u>…shouldn't I call</u>, I want them to make that call to 911.

head tilt right

Figure 4.28. *should I/shouldn't I,* subtle head tilt



(a)

S: Maybe you're watching the newscast right now and you're saying <u>should I</u>

G: Head tilt left

(b)

or <u>shouldn't I</u> get the flu shot. Well there's another benefit that may have people …

head tilt right

Figure 4.29. *should I/shouldn't I,* pronounced head tilt with lateral gestures

[97] The head tilts in Figure 4.28 and Figure 4.30 stem from videos that were not included in the totals for the 13 tokens presented above due to the camera angle or chyron obscuring the hands. The hands are visible in Figure 4.29 when watching the live video, which allowed this token to be included.

(a) (b)

(a) *S: It's just like most stupid things. You don't think about 'I'm risking my life - <u>should I</u>*

G: Head tilt right

(b) *or <u>shouldn't I</u>. I say 'chicken'… poor chicken… freeway, take. Bad idea, children, don't do this at home.*[98]

head tilt left

Figure 4.30. *should I/shouldn't I,* pronounced head tilt

4.4.3 Summary: Possible world CONTRAST

In this section, I have presented a qualitative analysis and a small quantitative case study that suggests that speakers use their bodies to mark CONTRAST when using the bipartite phrasal expression *should I/shouldn't I.* [SI~SI] conveys the construal of two alternative mental spaces in the irrealis domain. The construal of these alternate spaces is manifest in the body's bilateral use of space. In the examples I presented, shifts in head, gaze, shoulders, and hand movements all serve to indicate alternative irrealis spaces. These movements anchor the alternative spaces in real world, physical space in relation to the body. The kinesic profile that emerges seems not to be restricted to [SI~SI]. A similar kinesic profile is evident looking across examples of clausal *if*-statements. Figure 4.31 shows enactments of an utterance that features a pair of contrastive conditional statements. In these utterances, the speaker indexes two alternative spaces through a sequence of head tilts, first to one side, then to the other.

---

[98] As the video clip for this example makes clear, the context of the utterance is that the woman was driving on the freeway and came upon a chicken running across it. She subsequently got out of her car to rescue the chicken. The speaker is defending her action, which put her in danger.

(a) (b)

(a) *S: If Gary Johnson gets to that level, he'll be on the stage*
G: head tilt to right

(b) *If he doesn't get to that level, he won't*
head tilt to the left

Figure 4.31. *if…/if…*, head tilt

4.5 Conclusion

In this chapter, I have presented a series of qualitative examples and three case studies that provide some indication of how the body signals CONTRAST in the realis and irrealis domains. By measuring the degree of enactment, I have demonstrated that the body is regularly recruited to express CONTRAST and the examples illustrated that the body indexes contrastive objects, ideas or mental spaces, and anchors them in physical space around the body. This occurs through the use of the full range of articulators involved in co-speech behavior – from manual gesture, to shoulders, head, brow and gaze movement. Case studies of CONTRAST in the realis domain, such as expressed with *on (the) one hand/on the other hand* and *better than/worse than,* demonstrated that these expressions are expressed more frequently through gesture(with some involvement of the upper body, particularly for *on the other hand*). By comparison, case studies of CONTRAST in the irrealis domain suggested that this domain features more enactment in the upper body. For example, in the enactment of [SI~SI] and conditional constructions, there was a high frequency of enactments featuring only head tilts to index two different mental states.

By examining CONTRAST in both the realis and irrealis domains, I have provided corpus-based evidence that suggests both a high degree of convention and some variation in how contrast is marked across these domains. In the propositional domain, hand forms characterize CONTRAST marking, whereas in the cognitive domain it is movements of the upper body that most often signal CONTRAST. Symmetry and laterality are used conventionally in both domains, while the variability lies in which articulators the body recruits in each domain.

Gesture has been referred to as both a “window onto the mind” (McNeill 1992: 268) and as loci of “partial semiotic portrayal *par excellence*” (Mittelberg 2019: 2). Given the range of ways in which gestures encode meaning, my proposal that the domain of CONTRAST is enacted in the body in iconic ways is not novel; however, in this chapter, I have supported this claim with quantitative studies that show the range of co-speech behaviours that encode CONTRAST. I introduced the idea of degree of enactment as an important measure that should be included in how kinesic gestures are analyzed, for it demonstrates the strength of the kinesic signal as part of an enactment. I also showed how body partitioning can capture meaningful information about the types of language that may be encoded in the hands vs. the upper body. Due to the kinesic affordances of the hand, propositional contrast is iconic of actions and forces in the real world, while the marking of contrast in the upper body that characterized the irrealis domain was shown to be more schematic and often slighter, performatively speaking. This is to be expected given the limited forms available to upper body articulators. Despite this differential recruitment of articulators, the use of space in the marking of CONTRAST was stable across both domains. The data support the need to extend the discussion of iconicity in co-speech behaviour to include more schematic forms of signaling, such as uses of lateral space in gesture form and shifts in upper body movements. When compared to the more imageable representations available given the variability and more depictive or mimetic possibilities that can be signaled with the hands (e.g. hands can change shape, orientation, and direction), these upper body signals are more schematic, but, I suggest, are equally integral as indicators of the cognitive processes in play.

In this chapter, I have focused on investigating the multimodal expression of phrasal expressions with which speakers mark a comparison of two entities or propositions in the real world or two possibilities in the imagined world. In both cases, bodily reference to competing physical spaces indexes a concomitant contrast between mental spaces. In the next chapter, I turn my attention to yet another type of CONTRAST, namely at the discourse level. At the discourse level, CONTRAST is seen at discursive junctures – the moment in a discourse when a speaker leaves the main thread to make a parenthetical aside with expression such as *which is fine* and *which is true*, and again at the moment when he or she returns to the main utterance using expressions such as *however* and *anyways*.

Chapter 5 | *That being said*: Discourse constructions are multimodal

5.1 Introduction

Conversations are an exchange of dialogue between speakers. The ebb and flow of conversation means that in some moments one participant is speaking and the other listening, and vice versa; however, turn-taking is rarely as well-defined as the term suggests. Rather, conversation is messy and speakers frequently interrupt or speak over each other. The dynamic of turn-taking has long been studied as part of conversation analysis. More recently, the phenomenon of interruption has been at the centre of research from both a linguistic perspective and a broader one that examines, for example, the role that that gender plays in who gets interrupted and how often (Chira 2017; Hilton 2018). While speakers frequently interrupt each other, they also frequently interrupt themselves. Take for example the segment of conversation in (66).

(66) *This is crazy. I mean, this election could be like the Wild West, <u>which, by the way</u>, if you're interested in that period, you should read 'How Ronald Reagan Won the Wild West'.*[99]

---

[99] Red Hen: 2015-08-14_0300_US_ComedyCentral_Daily_Show_with_Jon_Stewart,2329

This type of interruption is akin to the convention in theatre known as *breaking the fourth wall*, in which actors temporarily suspend the interaction in the play to speak directly to the audience. In the same way, parentheticals or asides in regular conversation break the flow of the conversation to make a comment that is orthogonal to the direction of discourse. The interruption is a way for the speaker to provide additional information, add an evaluation or take a stance towards the discourse topic, ask a rhetorical question, or combine several of these functions. Such a digression can be short or lengthy. If it becomes too lengthy, the speaker may fail to return to the segment that was interrupted and simply move on to a new topic. The norm, though, is for the speaker to use a discourse connective like *anyways* or *that being said* to navigate back to the topic of the original segment or line of discourse. It is this type of discourse navigation away from and back to the main utterance that I focus on in this chapter.

In the preceding two chapters, I showed how catenative expressions of ASPECT and lexico-phrasal expressions of CONTRAST are enacted in the body. When North American English speakers express these conceptual notions in their speech, the patterns in their co-speech behaviour iconically and conventionally represent the content of the conceptual level. Here, I turn my attention to the investigation of more pragmatic material, namely devices that speakers use to navigate discourse. I focus on a range of expressions used by speakers of English to depart from and return to the main conversation frame. These departures and returns are generally highly stanced, meaning that these utterances have multiple functions: to guide the speaker and interlocutor through the discourse in a coherent manner and to imbue the digression with the speaker's opinions and attitudes regarding the propositional content of the main utterance. Contrast plays a role here in this type of discourse-marking. Speakers often create junctures in their discourse that fall under the broad conceptual category of contrast introduced in the previous chapter. In the discourse analysis literature, these have been described as 'push' and 'pop' markers, respectively, which mark the opening of a new subtopic (i.e. a parenthetical) and the return to the previous topic, respectively (Polanyi & Scha 1983; Grosz & Sidner 1986; Stede & Schmitz 2000). Findings presented in this chapter suggest that many of these markers have unique and recurrent kinesic signatures.

Specifically, I investigate here a set of expressions that carve up the discourse to assist speaker and hearer in navigating the given topic, and in expressing speaker stance (e.g. coming out for or against the previous utterance or discourse thread). Very typically over the course of a conversation, a speaker departs from the main "frame utterance" (Bolinger 1989: 185) to make an aside, or parenthetical comment, introduced by expressions such as *which, by the way*; *but I have to say*; and the like. The speaker then marks the return to the former frame utterance or topic of conversation or even the start of a new discourse segment with a concessive device such as *so anyway(s), but anyway(s)*, *at any rate*, etc. This chapter will thus present two series of case studies. In the first set of case studies, I examine parenthetical expressions with which the speaker adds a stanced comment to their own discourse flow. The second series of case studies will investigate the expressions the speaker uses to return to a previous discourse thread following an aside. While the range of expressions I examine do different functional work and may portray different stances from the speaker (e.g. either an evaluative/contrastive frame or an attitudinal marking), they demonstrate the same phenomenon: these are specific expressions of stance, with a high degree of enactment, that are expressed with a particular repertoire of kinesic articulations.

5.1.1 Background

What I term *excursions* and *returns* in this chapter have previously been discussed in the discourse literature as *push* and *pop* markers as cited above, and more generally are treated under the umbrella of pragmatic markers (Brinton 2017). In discourse analysis circles, these are "discourse structure signaling devices used to mark movement from one discourse unit to another. A *push* to an embedded unit, or a *pop* back to a temporarily displaced controlling unit is often clearly marked in natural interaction" (Polanyi & Scha 1983: 263ff). Polanyi and Scha were well ahead of the curve in terms of modern gesture studies and the multimodal turn in cognitive linguistics in describing these types of devices, stating that they can consist of "intonational, lexic syntactic [sic] and textual mechanisms as well as *kinesic body language gambits"* (ibid.: 264; italics mine).

Push markers, or what I'm calling excursions, are frequently signalled by headless relative clause fragments initiated by *which*. They allow the speaker to add explanation, evaluation, or justification in the course of an evolving conversation (Biber et al. 1999: 223). According to the *Longman Grammar of Spoken and Written English* (Biber et al. 1999), a grammar of usage based on extensive corpus data, common examples include *which is why, which is what, which is where, which explains,* and *which brings me to.* They often start with the relative pronoun *which,* which signals a close connection with the preceding clause. In the microstudies presented in this chapter, examples use the linking relative *which* in an adverbial construction, as in *which is fine, which is great*, and comment clauses beginning with *which, by the way*.

When the speaker desires to pop back to the main discourse unit or to begin a new unit, she has a variety of return strategies to choose from. To express reservation about the preceding clause, she may use a concessive, such as *however* or *but*, or a linking adverbial such as *anyways*. Both of these strategies mark the preceding discourse segment (or aside or digression) as less important when considered in contrast to the main point that the speaker is about to make in the following utterance.

Discourse particles such as those discussed here have a particular role in spoken language. Generally speaking, spoken language is "rich in particles that do not contribute to the propositional content of utterances, but play important roles in steering the flow of the dialogue and in conveying various attitudes and expectations of the speaker" (Stede & Schmitz 2000: 125). While much of the literature has used textual data (i.e. transcribed speech) to identify patterns in discourse navigation, some studies use speech data and include characteristics of the speech signal, such as intonational contour, shifts in speed, pitch, or volume in their investigations (Horne et al. 2001; Redeker 2006).

Given the important role of discourse navigation in conversation, it is surprising that manual gestures marking discourse functions have garnered relatively less attention than iconic gestures in the field of gesture studies. As described in the introduction, *interactional gestures*, or *pragmatic gestures*, refer to gestures that "serve the special conversational demands of talking in dialogue" (Bavelas et al. 1995: 394). The Away family of gestures (brushing away, sweeping away, and holding away) are principally interactional or pragmatic in nature and do much of the work in signaling discourse returns, for example. The uniting theme underlying the Away family is the clearing away of unwanted objects (Bressem & Müller 2014). It could, therefore, be expected that gestures in the Away family are used to mark concession, as the findings below suggest, given that concessive markers express speaker reservation about the preceding discourse (Biber et al. 1999).

As described in Chapter 1, a growing body of literature is emerging on co-speech behaviour that extends the previous focus on manual gesture to include the shoulder, head, gaze, and facial movements that accompany face-to-face interaction. In Chapter 1, I introduced studies that have shown that movements of the upper body have been shown to be used particularly to mark stance, for example the shrug (Debras & Cienki 2012; Debras 2017). Similarly, other channels, such as eye-gaze, are conventionally used to mark phenomena such as alignment between speakers, turn-taking in conversation, and viewpoint (Sweetser & Stec 2013; Oben & Brône 2015). Finally, what Chovil (1989) identified as *conversational facial gestures* and Kendon (2004) called simply *facial gestures* have been defined as "configurations of the face, eyes, and/or head that are synchronized with words and other co-speech gestures" (Bavelas & Chovil 2018: 91).[100] In the microstudies presented here, head, shoulder, brow, torso, and gaze movements are shown to be conventionally as well as idiosyncratically aligned with specific discourse navigating expressions.

5.1.2 STANCE and constructional specificity

The constructional approach to language taken in this dissertation claims that language structure emerges from conventionalized units. A construction captures regularities in usage both at a schematic level (e.g. transitive verb constructions) and in a specific verb's preferred patterning. Studies have shown, for example, that verb senses and transitivity patterns are often tied to specific collocation patterns (Rice 1987; Thompson & Hopper 2001; Newman & Rice 2006). For example, while *eat* and *drink* are typically ascribed the status of transitive verbs,

---

[100] Kendon's term *facial gestures* expressly differentiates time-aligned and meaning-conveying facial movements from emotional facial expressions that have dominated the study of facial movement (Ekman 1993).

which require a direct object, usage favours utterances such as *it's time to eat* and the verb phrase *to go drinking* (which is specific to the frame of drinking alcohol). That is, "specific collocations involving specific verb senses develop lives of their own" (Thompson & Hopper 2001: 44).

Collocational strength has been demonstrated to hold not only for transitivity schemas, but also for collocations involving specific verbs and specific inflections and TAM marking (tense, aspect, and mood) (Newman & Rice 2006). For example, the verbs *mean* and *know* prefer first- and second-person inflection, respectively. The resulting collocations, *I mean* and *you know*, have grammaticalized to acquire the status of discourse markers. To give another example of inflectional verb patterns, in a study of inflectional patterns of the verb *think* in the British National Corpus*,* first-person forms of *think* accounted for an average of 80% of all uses across all inflected forms (Rice & Newman 2005). Similarly, Rice and Newman (2018) looked at the epistemic verbs *think* and *know* in COCA and showed how they skewed to first-person singular (1SG) and second person singular (2SG), respectively. Broadly speaking, corpus studies like these have shown that semantic properties tend to inhere in specific morphosyntactic inflections of a lexical item and may not extend across the lemma as a whole. That is, we should expect idiosyncrasies of meaning, form, collocation, genre, and distribution in the verbal signal as well as in the kinesic signal (Hinnell & Rice 2014).

These text-based corpus studies demonstrate the degree of constructional specificity inherent in language use and the quasi-formulaic nature of (spoken) language. Speakers use and re-use patterns that are familiar and entrenched collocations. In this chapter, I extend this line of research to explore whether specific collocational patterns differentially affect the use of gesture and other co-speech behaviours. Are some collocations more strongly signaled in the body than others? Comparing the degree of enactment between expressions captures this behaviour. Moreover, do fine lexical and inflectional distinctions in the speech signal lead to different patterns in kinesic behaviour? Finally, the utterances investigated here serve a dual purpose as discourse navigation markers and as stance markers. What role does the stance conveyed by a specific collocation play in the conventionalized form of that utterance? To illustrate what I mean by this last question, I briefly compare two near-neighbour interrogative collocations, *what can I say?* and *what can I do?* (hereafter [WCIS] and [WCID], respectively).

[WCIS] and [WCID] are both fixed expressions that vary lexico-syntactically only in the verb slot. The sentences function quite differently. The [WCIS] construction is most frequently expressed rhetorically and functions as a parenthetical aside. By contrast, [WCID] functions largely as a true interrogative in which the speaker asks whether she can assist with anything. To observe the behaviours in the speech and kinesic patterns for these two utterances, I performed a corpus study using 100 examples of each expression collected from COCA_{sp} and 20 valid search returns of each expression in the Red Hen archive. The search returns in COCA_{sp} indicate that the utterance patterns for these two phrases differ widely. Of 100 [WCIS] search returns in COCA_{sp}, 87% were not modified in any way. Instead, the question stood as a complete sentence and was generally followed by highly stanced, collocating clauses which expanded on or responded to the rhetorical [WCIS], as in the underlined segments in (67)-(69).

(67) *What can I say? I don't know*.[101]

(68) *What can I say? I think he should have known*.[102]

(69) *What can I say? I love her*[103]

---

[101] COCA: 2005-01-15,CNN_NEXT,Extreme Nature

[102] Red Hen: 2019-05-30_0630_US_KOCE_Amanpour_and_Company,677

[103] COCA: 1996-11-20,Ind_Springer,My brother stole my wife

The remaining 13% featured sentences with collocation patterns such as those in (70)-(72), which feature prepositional phrases and dependent clauses.

(70) *What can I say about the Superbowl that hasn't already been said?* [104]

(71) *What can I say to change his mind?* [105]

(72) *What can I say that's not going to get me taken out?* [106]

By contrast, [WCID] was followed by a modifying phrase in over 65% of cases, as in (73)-(75).

(73) *What can I do to make money?* [107]

(74) *What can I do but cover it up with a beard?* [108]

(75) *A lot of times when I travel people are saying, what can I do as a concerned citizen?* [109]

A study of 20 enactments each for these two lead-in phrases, [WCIS] and [WCID], showed that, while equally frequently enacted (both 17/20, or 85%), the kinesic profiles themselves vary widely. [WCIS] features a high degree of PUOH forms (70%) and shoulder shrugs (53%), and a high degree of head movement (83%), mostly head shakes. By contrast, the form for [WCID] showed no discernible patterns. While [WCID] featured gestures in 88% of the cases and head movement in 40% of the cases, the forms were too varied to capture any meaningful patterns in the enactments. The prosodic profile in these two expressions also impacted the nature of the kinesic enactment. In one instance of [WCID], the prosody was more typical of the typical intonational contour of [WCIS]. In this instance, the gesture form also mirrored the typical [WCIS] gesture, i.e. the PUOH form. When [WCID] is paired with

---

[104] Red Hen: 2019-02-05_0837_US_KNBC_Late_Night_With_Seth_Meyers,45

[105] COCA: 2018-10-26,ABC_GMA,How to go after that raise

[106] Red Hen: 2019-02-04_1500_US_KABC_Good_Morning_America,6207

[107] Red Hen: 2019-01-23_2300_US_KNBC_The_Ellen_DeGeneres_Show,548

[108] Red Hen: 2019-05-31_1700_US_KABC_The_View,2641

[109] Red Hen: 2019-05-03_0300_US_CNN_CNN_Tonight_With_Don_Lemon,2586

the prosodic profile of [WCIS], the outcome as a gesture coerced into the [WCIS] kinesic form. Several enactments each for [WCIS] and [WCID] are shown in Figure 5.1 and Figure 5.2.

In sum, [WCIS] is a much more fixed construction from a syntactic perspective and displays a high degree of conventionalization in the kinesic signal. [WCID], by contrast, has more varied usage patterns linguistically and this wide-ranging quality is reflected in the wide-ranging, ad hoc nature of the kinesic signal. While this comparison between [WCIS] and [WCID] has been preliminary and deserves far more exploration in and of itself, my intention is for the pair to serve as a brief example of the degree of specificity that inheres to a construction, both in its usage context and in its kinesic enactment, and as an example of the impact that a high degree of stance can make in the conventionalization of form.

(a) (b) (c) (d)

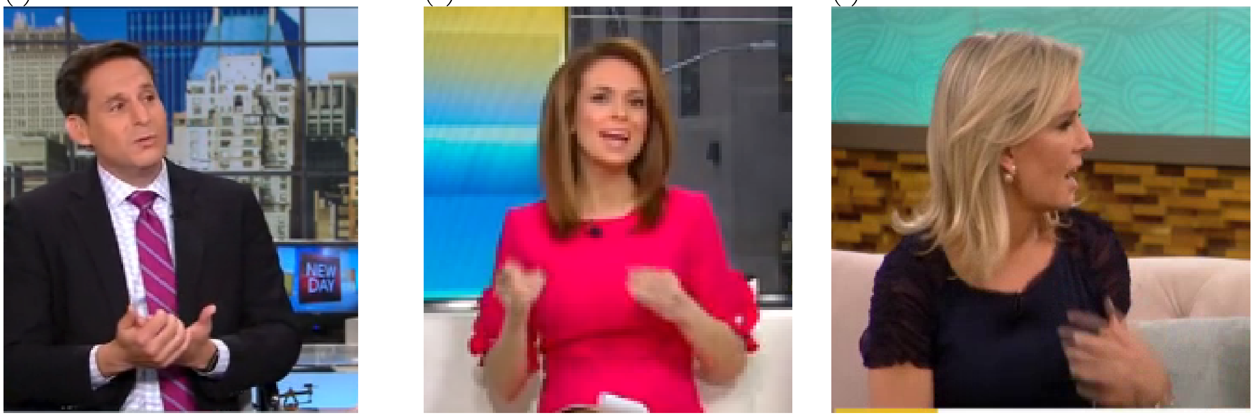
(a) S: *Yes. What can I say?*
G: PUOH, head shake

(b) S: *What can I say? It's very dark, yeah. I swear that's a true story.*
G: PUOH, head shake

(c) S: *What can I say? It's an emotional attachment to it.*
PUOH, shoulder shrug

(d) S: *Well, okay. What can I say? If the democratic party has moved so far to the left that Bernie Sanders…*
G: PUOH, head tilt

Figure 5.1. *what can I say?,* PUOH gesture



(a) S: *What can I do? I need to raise my game because they raised their game.*

G: palms up, head tilt

(b) S: *What can I do here? We cannot allow this problem to get bigger…*

G: pursed hands

(c) S: *What can I do to help you?*

G: indexical gesture to self (on *I*) and then interlocutor (on *you*)

Figure 5.2. *what can I do?,* representative enactments

5.1.3 Summary

In this chapter, I focus specifically on the use of utterances that introduce parenthetical asides or other types of subjective and seemingly extraneous comments and on devices that are used to return from these excursions. Excursion constructions that I examine include *how should I put it?; which is true; which is fine; which, by the way;* and *although, you know.* As 'toss-off' comments, these (inter)subjective qualifiers are disruptive to the flow of discourse and are frequently matched by co-speech behaviors that depart from bodily actions that precede or follow the aside. I also investigate the kinesic articulations that accompany spoken return utterances such as *anyways* and *however.* Speakers use these devices to return from their parenthetical comment to the main flow of discourse. Both excursions and return expressions perform dual functions in discourse: they sign post the flow of discourse in addition to marking speaker stance. The findings suggest that the specific bodily articulators used (hands, head, eyes, eyebrows, torso) and the form these bodily articulators take (e.g. open palm in the gesture, head tilt/nod/shake, eyebrow raise) are reliably associated with particular expressions. The fixed and semi-fixed expressions in the spoken utterance are accompanied by conventionalized kinesic enactments of these discourse segmentation devices.

In order to focus on a wide range of utterances types, I present a series of microstudies of five excursion expressions and seven return expressions and establish their multimodal profile. Each study consists of 20-25 tokens of the target utterance (20 each for excursions, 25 each for returns).

5.2 Methods and data overview

The data presented in this chapter are sourced from the Red Hen archive. The general methods adhere to the methods detailed in the Methods chapter (Chapter 2). Here, I outline methods specific to this chapter. As a brief reminder, all corpus returns stem from interactions which are spontaneous and unscripted and which show enough of the hands and upper body to annotate.

The data range of data collection for the excursion and return studies presented in this chapter varied slightly depending on the expression. For the excursion studies, data were collected from programs that were broadcast between November 1, 2006 and November 1, 2015. For most of the return expressions, the date constraint was set to search from January 1, 2012, to January 1, 2018; however, for three return expressions (*that being said*; *however, I*; and *nevertheless*), data were collected from November 1, 2012 to November 1, 2015. These differences reflect when I started to collect the data.

While most of the data were collected from the television broadcast networks listed in Appendix B, the search parameter for *at any rate, but anyway(s),* and *so anyway(s)*, did not place a restriction on the networks or programs searched so as to maximize the number of returns. Since the criteria for data collection demanded a spontaneous, unscripted speech context, the returns nonetheless stemmed largely from the list of programs used elsewhere in this dissertation. For the target utterances *at any rate, but anyway(s),* and *so anyway(s)*, extending the range of programs did not appear to add noise to the search returns (as one would expect of such highly fixed expressions). Screenshots are given as single shots or as sequences as appropriate.

In what follows, I motivate the target expressions that were selected, including providing a comparison of frequencies of these utterances in the Corpus of Contemporary American English COCA (§5.2.1). I then explain methods of data annotation (§5.2.2) and outline how I will present findings (§5.2.3).

5.2.1 Target utterances

In this section, I motivate the selection of the five excursion and seven return markers for which I present case studies, beginning with excursions, linguistic devices that introduce parenthetical asides or other types of subjective, seemingly superfluous comments. The excursion expressions I investigate are *how shall/should I put it?; which is true; which is fine; which, by the way;* and *although, you know.*

Excursions

The selection process for the excursion markers consisted first of gathering data on a wide range of target utterances to determine which were the most commonly used. I used COCA_{sp} as a proxy for North American spoken English. I consulted discourse analysis literature as well as corpus-based studies of discourse markers to develop a fulsome list of devices that speakers use to make parenthetical asides. Many of these utterances take the form of a small, evaluative relative clause headed by *which.* They are not connected to the main clause propositionally, only discursively. Examples include *which is true, which is fine, which is great,* etc. A search of COCA yielded the most frequent collocates in these [which is ADJ] constructions. The top 20 collocating predicate adjectives are shown in Table 5.1.

Table 5.1. Top 20 collocates for [which is ADJ] in $COCA_{sp}$

	[which is ADJ]	Freq. in $COCA_{sp}$		[which is ADJ]	Freq. in $COCA_{sp}$
1	*which is great*	263	11	*which is amazing*	49
2	*which is good*	194	12	*which is hard*	44
3	*which is important*	105	13	*which is better*	41
4	*which is interesting*	99	14	*which is unusual*	40
5	*which is supposed*	90	15	*which is wonderful*	37
6	*which is fine*	85	16	*which is worse*	35
7	*which is right*	79	17	*which is available*	29
8	*which is different*	78	18	*which is understandable*	29
9	*which is true*	73	19	*which is ridiculous*	28
10	*which is nice*	64	20	*which is responsible*	27

The search extended beyond these relativized [which is ADJ] constructions to include idiomatic expressions such as *let's face it, sad to say,* and others. The selection of target expressions was thus an iterative process which involved seeking out expressions in the literature and finding near neighbours in corpus searches. Preliminary investigations also included a search of Red Hen to determine if there were enough search returns to support a case study. The results of these preliminary searches are shown in Table 5.2. Parenthetical aside markers are listed in descending order of frequency in Red Hen, with COCA frequency also provided.

Table 5.2. Frequency of excursion expressions in Red Hen and COCA$_{sp}$

Expression	Red Hen*	COCA$_{sp}$
which is great	1575	264
let's face it	1385	5
which is good	1198	194
which I love	883	72
which, by the way	**767**	**373**
but I have to say	571	261
which is nice	455	64
which is weird	307	18
which is fine	**300**	**85**
which is fantastic	251	18
which is interesting	248	99
which is important	237	105
truth be told	194	57
sad to say	184	83
which is true	**174**	**73**
but I want to say	146	90
although you know	**95**	**93**
how should I put it	**23**	**13**

* Red Hen search: Nov. 1, 2006-Nov. 1, 2019

The five expressions chosen as final target utterances are in bold in Table 5.2. I performed corpus searches for an additional 20 utterances that are asides in and of themselves or are considered items that can introduce other parentheticals. These collocations included *which seems to be the case, which I completely, which I fully, now let me see*. None of these were associated with enough search returns in Red Hen to include as case studies.

To close this section on selection of excursion utterances, I provide examples of the contexts in which they occur. In the case studies of the [which is true] and [which is fine] constructions, I included examples in which the clause was at the end of the sentence as well as examples in which the *which*-clause was modified (e.g. by a prepositional phrase). The distinction between these two scenarios is exemplified in the pairs from $COCA_{sp}$ shown in (76)-(77) and (78)-(79), respectively.

(76) *He calls himself a business guy, which is true. But I have to say I know plenty of businessmen and women who have achieved remarkable success without leaving a trail of lawsuits and unpaid workers and people feeling like they got cheated.*[110]

(77) *I use a line that humor is pain seen from a safe distance. In the case of the – which is true of my book generally -- but in the case of the Holocaust there was no humor and I played it very stark and straight.* [111]

(78) *They shut down the straw poll on Friday. I spoke Saturday. And so I was in the position of Palin and Huckabee. I didn't -- for the purposes of the straw poll, I didn't get to speak which is fine. I mean, they got to have rules and that's fine with me.* [112]

(79) *They're real good about paying me back, which is fine with me, I don't mind.* [113]

Which, by the way requires a clause to complete it, whereas *how should I put it* (hereafter HSIPI) is a parenthesis in and of itself and is wholly independent of the surrounding clause structure. A pair of each of these utterances is given in the following examples:

(80) *On the other side there's post-traumatic stress disorder, which a lot of soldiers have served in Iraq and Afghanistan have and don't necessarily impair their judgment or their ability, and it means they have nightmares and intrusive thoughts, which by the way a lot of journalists do too because they go through the same kind of traumas.* [114]

(81) *And Nick wanted to go out on top so the critically acclaims sketch comedy series will say farewell after its third and final season, which by the way is already underway.* [115]

(82) *Well, I think that with the Supreme Court nominee, just as with the Attorney General - how shall I put it? - the president got a lot better than his messy process deserved to get, right?*[116]

---

110 COCA: 2016-07-28,CNN Newsroom 2am EST

111 COCA: 1998-04-09,NPR TalkNation

112 COCA: 2011-02-13,Fox Sunday

113 COCA: 2014-07-12,ABC Nightline 10pm EST

114 COCA: 2017-11-13,Tucker Carlson Tonight 8pm EST

115 COCA: 2015-01-20,NBC Today Show 7am EST

116 COCA: 1993-06-19,NPR Weekend

(83) *The Iraqi- Israeli non-conventional arms race is something terrifying because it's, <u>how should I put it</u>, it's an arms race in which there are no rules.* [117]

In this section, I have introduced the target utterances that form the investigation for parenthetical asides in this chapter. I present case studies for each of them in §5.3, below. In the remainder of this section, I review the selection process for the set of utterances that I examined that speakers use to return to the main discourse after a parenthetical comment.

Returns

The original list of return expressions I compiled was based on the literature on concessive devices and discourse markers, as these are the types of linguistic material that function as return expressions. As outlined in the introduction, lists of concessive markers are given in the *Longman Grammar of Spoken and Written English* (Biber et al. 1999). Other sources included Lenk (1998) and Precht (2000).

I encountered several challenges in operationalizing the data collection. Firstly, some of the expressions are highly polysemous. That is, they have functions other than to signal a return to prior discourse. For example, *however* can function as an adverb of degree or manner, as in (84) and (85), or as a concessive marker to mark narrow-scope contrast, as in (86).

(84) *<u>However alert</u> we are, <u>however much</u> we think we know what will happen...*[118]

(85) *<u>However you may feel</u> about them, they are America's royalty...* [119]

(86) *Others, <u>however</u>, see and hear something quite different.* [120]

[117] COCA: 1990-06-20,PBS Newshour

[118] COCA: 201-06-29,NPR FreshAir

[119] COCA: 1994-10-18,Ind Geraldo

[120] COCA: 2006-03-02,CNN AM

There was no way in Red Hen to isolate the search listings of *however* to include only the ones that occurred as discourse navigation markers. To give an impression of the magnitude of returns, a search for *however* in all programs over the full Red Hen archiver results in over 100,000 hits, too many to manually review. In order to delimit the search for the targeted function, I included the first-person singular subject pronoun and searched for the text string [however I]. It was not possible to search for *however* at the left periphery by delimiting the search by punctuation as I did for the COCA search, as in [. however ,], as punctuation is not read by the Red Hen search engine. The search for [however I] yielded 2,297 search returns and reduced the noise in the search returns dramatically. Search returns for other collocations with pronouns in Red Hen, such as *however it* and *however he/she,* lacked the more subjective discourse navigation function that is strongly suggested by the presence of 1SG. In (87), from COCA_{sp}, the speaker even underscores the return function of *however* by following it with *I go back to my first point.*

(87) *Yes, I think it's undeniable it's part of a strategy. <u>However, I</u> go back to my first point which was that on the other side of this that is what you should be looking forward to.*[121]

The polysemous nature of *anyway* generated a similar challenge in data collection. While there were 10,674 search returns for *anyway* in COCA_{sp}, the majority of these were medial or at the right periphery of a sentence. Occurrences at the right periphery accounted for approximately 30% of those search returns, as in (88). Medial occurrences accounted for approximately half, as in (89) and left periphery occurrences amounted to the remaining 20%. While the use of *anyway* in the medial and right periphery positions does add speaker stance to the utterance,

[121] Red Hen: 2019-10-29_1400_US_FOX-News_Americas_Newsroom,605

they lack the discourse navigation function that inheres to *anyway* when it appears at the left periphery, as it does in (90).

(88) *You might remember that the memo from the call that was released from the White House specifically said it was not a verbatim transcript of the discussion but that didn't stop Trump from insisting it was anyway.*[122]

(89) *And it seems like this is going to, in the short run anyway, make it a little bit harder for them to make the case.*[123]

(90) *Earlier this year, there were -- oh, maybe last year, I've lost track Anyway, there was an ad in China for laundry detergent and -- last year -- they, you know, have a look at this ad.*[124]

Due to the inability of Red Hen to take punctuation into account (which could assist one greatly in distinguishing between peripheral vs. medial uses), I chose instead the more phrasal collocations, *but anyway* and *so anyway,* which only ever occur at the left periphery. In addition to *anyway*, speakers of North American English at times use *anyways* (with an 's') in the same context. I checked the frequency of both forms in left periphery position in COCA_{sp} and found the form without an 's', namely, *anyway*, to dominate. In my Red Hen searches, I included both forms in my searches, effectively collapsing *but anyway* and *but anyways* and the same pairing with *so*. The target utterances with *anyway/anyways* are therefore represented hereafter as *but anyway(s)* and *so anyway(s).*

The chart in Figure 5.3 indicates the relative frequency in COCA_{sp} of several discourse-marking utterances. Here, relative frequency is the frequency of each target utterance relative to the total number of search returns for all of these utterances. Because COCA supports punctuation in the search query, I use the punctuated forms, e.g. [. However] and [. Anyway],

---

122 Red Hen: 2019-11-09_0837_US_KNBC_Late_Night_With_Seth_Meyers,532

123 COCA: 2017-03-13,PBS Newshour

124 COCA: 2017-10-10,CNN Newsroom 2am EST

to limit the search to left periphery occurrences. The set consists of 8,060 total returns. Of these, over 51% were *however,* with sentence-initial *anyway* and its variants, *but anyway(s)* and *so anyway(s),* making up another 33%. *In any case* and *nevertheless* constituted another 5% each of this set of returns, with the remainder making up the remaining 6%. A scan of the search returns for the phrasal target utterances (i.e. those not delimited by punctuation, such as *but anyway(s)* and *at any rate*) confirmed that in the vast majority of search returns, these expressions occurred in the discourse navigation function.

I also looked across genres in COCA to explore which discourse markers characterize speech as compared to other genres. The frequency of selected return expressions by genre is show in Figure 5.4. Discourse markers are listed in the legend from most frequent to least frequent in the spoken portion of COCA. As the figure shows, *however* and *nevertheless* are most likely to be used in academic prose, but despite this seeming aversion for spoken usage compared to other genres, these two markers are still among the most frequent spoken return markers (see Figure 5.3).

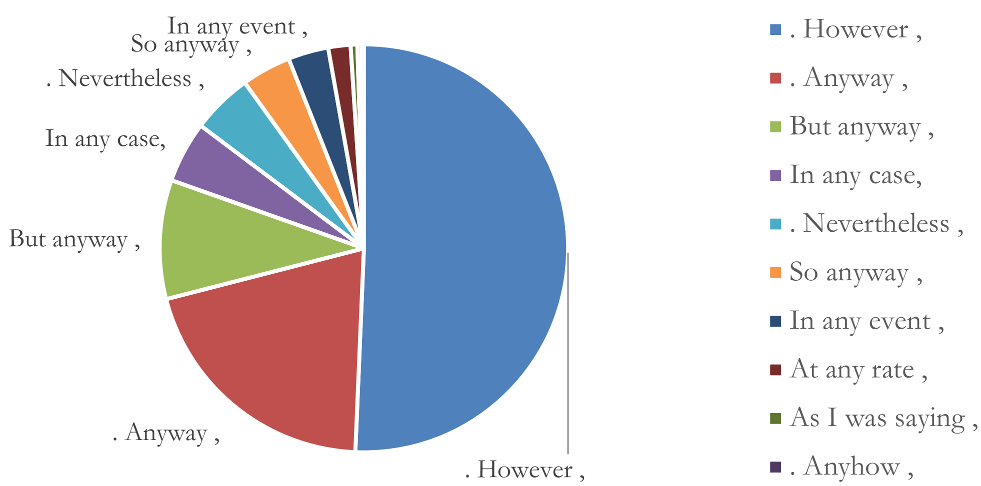


Figure 5.3. Relative frequency of selected return expressions in COCA_{sp}

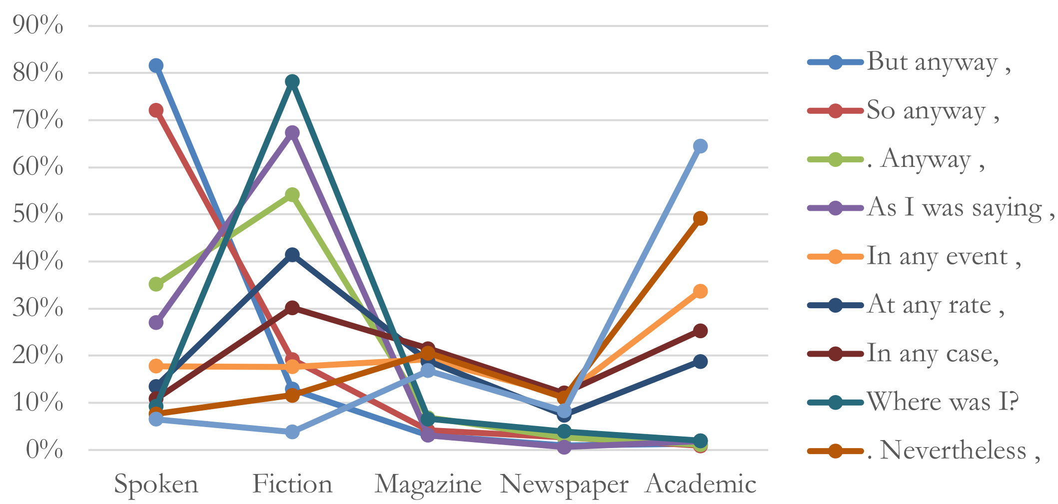


Figure 5.4. Genre preference for return markers in COCA

The final decision about which discourse markers to include in the case studies was made based on availability of data in Red Hen. In some cases, there were not enough search returns that met the criteria for spontaneous conversation and an unrestricted view of the body. The markers that did not have enough viable hits in Red Hen were *in any case, as I was saying* and *where was I* and were not included in the case studies presented here. In Table 5.3, I provide the final list of return expressions that I investigate in this chapter.

Table 5.3. Return utterance case studies

Discourse return constructions
anyhow
at any rate
but anyway(s)
so anyway(s)
however, I
that being said
nevertheless

5.2.2 Data annotation

Data annotation proceeded as described in the previous two chapters. I viewed all the search returns in reverse chronological order and noted whether each return met the criteria of occurring in a spontaneous utterance in which I could see the speaker's hands and upper body. Once I reached a total of 20 (in the case of excursion markers) or 25 (in the case of return markers) valid search returns for each utterance type, I proceeded with annotations. The annotation variables and levels are summarized in Table 5.4. The annotation schema is consistent with the ones used in Chapters 3 and 4, with a few exceptions, which I outline here. For gesture, I did not capture the axis or direction of the movement, focusing rather on palm orientation to capture gesture form (i.e. palm-up, palm-down, etc.). Similarly, for head movement I recorded only the presence of head movement and the form (e.g. shake, nod, tilt), but not the direction of the movement. The camera distance and camera angles in many cases made the determination of gaze shift very difficult and so I did not annotate gaze behaviour with these utterances. Likewise, shoulder movement was exceedingly rare for these expressions and I therefore did not include annotation for these articulators.

Table 5.4. Kinesic annotation variables

Modality	Variable	Levels
Gesture	Gesture space	core, left of body, right of body
	Hands	left, right, both
	Recurrent form and/or palm orientation	palm-down, palm-away, PLTC, PUOH, container, etc.
Head	Movement type	nod, tilt, shake, other
Shoulders	Direction	up, down
Brows	Direction	raise
Torso	Direction	turn, lean (left, right, forward, back, other)
All	Partitioning	G (gesture only), G&UB (gesture and upper body), UB only

To summarize, I have laid out the utterance types that I explore in this chapter and have detailed how data collection and annotation proceeded for the case studies of discourse navigation devices. These are devices that are generally highly stanced. They aid the speaker in sign-posting shifts in the direction of discourse and also imbue their discourse with attitudinal, dismissive, or contrastive force. I begin by presenting the microstudies of five excursion markers (§5.3), followed by microstudies of seven return markers (§5.4).

5.3 Excursion markers

Excursions are a natural part of spontaneous interaction. As anyone who has listened to a person tell a story or impart information will know, the journey is rarely linear. Rather, spontaneous narrative (as well as casual conversation) is rife with expressions that mark the speaker's stance toward the happenings that are being reported on. These incidental remarks or asides go by the name of *obiter dicta* in common law or court proceedings. This term comes from the Latin, meaning 'something said by the way/in passing'.

In natural discourse, stance markers are often tacked on at the beginning or end of an utterance, or, as Beeching and Detges (2014) and Traugott (2014) describe these positions, at the left or right periphery. The constructed examples below show left-periphery stance markers in (91) and right-periphery markers in (92).

(91) a. *Obviously, he has no idea what he's doing.*
b. *As one might expect, everyone's a bit disappointed in him.*
(92) a. *Everyone wants him to leave, to tell you the truth.*
b. *We're going to need a new director, I suppose.*

Stance markers can also be wedged into the middle of an utterance. It is this medial position that I focus on in this section of the chapter. In examples (93) to (95) from COCA_{sp}, the underlined expressions do more than simply mark speaker stance. They mark an excursion away from the main thread of discourse. I use the term *parenthetical aside* or *excursion* to reflect this discourse navigating function.

(93) *It's just an example, Mr. Green, of how nasty this has gotten, you have to admit, on both sides.*[125]

(94) *I mean, if there was nothing in the DNA report – and, keep in mind, this is the defense lawyers characterizing the DNA report – if there was nothing, he'd say, 'There's nothing.'*[126]

(95) *Well, I want to mention a film that's going to loom very large at the Academy Awards, which by the way I hate, but I feel like I have to acknowledge.*[127]

Parenthetical asides vary widely in their structure. They may be introduced by relativizing or subordinating particles such as *which* or *although,* but they may also be made up wholly by fixed expressions such as *you have to admit* as in (93), as well as *how shall I put it* and the others introduced in the Methods section (§5.2.1). In Table 5.5, I repeat the list of phrases that mark discourse excursions given in Table 5.2 with their frequencies in Red Hen and COCA_{sp}. Here, rather than focus on their frequency in corpora, I include the degree of enactment. The table is sorted by decreasing degree of enactment of each expression; those in bold are the ones I provide case studies for in this section.

---

125 COCA: 2001-11-06,NBC Today

126 COCA: 2006-05-12,Fox HC

127 COCA: 2014-12-24,PBS FreshAir

Table 5.5. Degree of enactment for excursion expressions in Red Hen

Expression	Red Hen	Degree of Enactment
which by the way	***767***	***90%***
which is fine	***300***	***90%***
which is true	***174***	***90%***
how shall/should I put it	***23***	***85%***
which I love	883	80%
which is great	1575	70%
which is interesting	248	70%
which is nice	455	60%
which is weird	307	60%
although you know	***95***	***55%***
let's face it	1385	55%
sad to say	184	55%
but I want to say	146	55%
but I have to say	571	50%
which is fantastic	251	50%
which is good	1198	45%
truth be told	194	43%
which is important	237	n/a

NB: Excursion expressions in bold are those included in this chapter's case studies

Regardless of their position in the sentence and their specific structure, what these expressions share is their dual function as excursion markers and stance markers. As described in §1.2.5, stance uses a bodily metaphor to refer to a positioning or the footing on which someone speaks. The same bodily metaphor is at play in our language around *making an aside,* which refers to a comment made to the side of our bodies in physical space as in the case of an aside in theatre when actors break the fourth wall and turn to address the audience directly. Both stance and parenthetical asides signpost a speaker's position with regard to the utterance and its context in the larger discourse. The material introduced in this chapter – both excursions and returns – are markers of highly subjective content in discourse. Asides are parenthetical and often phrasal, whereas the returns presented in the next section (§5.4) are shorter collocations, mainly consisting of just a single word (e.g. *nevertheless, anyhow).* I suggest that the aside expressions I present in this section signal the highest degree of subjectivity of all the case studies presented in this dissertation.

In sum, excursions are a locus for significant stance-marking in natural discourse. In this chapter, I examine the specific kinesic profile of several of these excursion markers. Given their highly stanced content, it was expected that they would be highly marked in the body, but excursion markers are also remarks made *in passing*. As such, they are often said with relative *soto voce* verbal quality. This conflict may be one reason for the relatively low degree of enactment shown for some markers, as described above in §5.2.1 (Table 5.5).

5.3.1 Excursion microstudies

In this section, I present case studies of five parenthetical asides: *which is true* [WIT]; *which is fine* [WIF]; *which, by the way* [WBTW]; *how shall/should I put it* [HSIPI]; and *although, you know* [AYK] (20 tokens of each). The presentation of the multimodal profiles of these expressions will consist of a summary of characteristics of the utterance and include examples with screenshots from Red Hen. I will begin with a report on the degree of enactment and body partitioning for all the expressions and follow that with a report of the kinesic annotations for each utterance.

I begin with the rate at which the excursion expressions were enacted in the body, or the degree of enactment. The degree of enactment ranged from 55-90%, as presented in Table 5.6; four of the five excursion markers were enacted 85% or more of the time. The excursion expressions headed by *which* ([WIT, [WIF], and [WBTW]) were the most highly enacted, at 90% each.

Table 5.6. Degree of enactment by excursion marker

	[WIT]	[WIF]	[WBTW]	[HSIPI]	[AYK]
Degree of enactment (raw)	18/20	18/20	18/20	17/20	11/20
Degree of enactment (%)	90%	90%	90%	85%	55%

After obtaining the degree of enactment, I grouped the annotation variables by articulator in order to show the relative contribution of each set of articulators, namely, the hands vs. the upper body (which included head, shoulder, eyebrow, and torso). Thus, the findings for body partitioning reveal whether the enactments of these utterances were dominated by gesture or by upper body movement, or both occurring simultaneously as a composite signal. The findings for body partitioning for the excursion expressions are presented in Figure 5.5, with the number of tokens on the vertical axis (i.e. data sets were all n=20) and the body partitioning presented as a proportion of the enacted forms.[128]

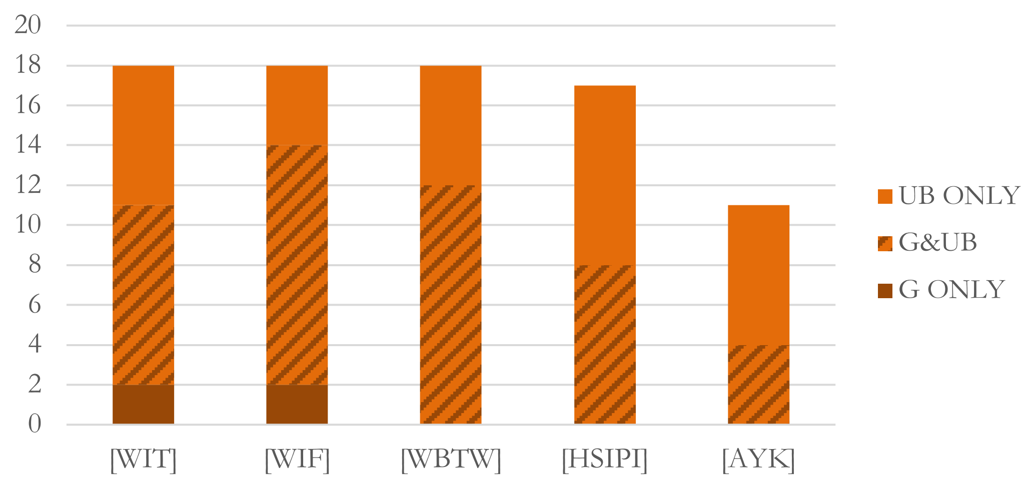


Figure 5.5. Body partitioning for excursion markers

[128] Body partitioning data for *how should I put it* includes gaze shifts. As I explain below, gaze featured very prominently for HSIPI, while for other asides it was inconsequential both with regard to execution (i.e. very difficult to observe) and pronounced gaze shifts occurred at only a low frequency or not at all.

The markers with *which* (i.e. [WIT], [WIF], and [WBTW]) all featured higher degrees of gesture and upper body use compared to [HSIPI] and [AYK]. The latter two constructions featured a higher percentage of enactments with upper body only.

Having presented the degree of enactment and body partitioning profiles for each utterance, I now report on the findings for each utterance individually. Excursion expressions are reported on beginning with the headless relative clauses *which is true; which is fine;* and *which, by the way*; and finishing with *how shall/should I put it,* and *although, you know.*

Which is true

Which is true ([WIT]), is an excursion marker that contains an overt epistemic marker, *true*, making it one of the most highly stanced of the five parenthetical asides profiled here. In my data set of 20, 14 occurred at the right periphery, as shown in (96), while the other six were followed by conjunctions such as *but* as in (97)*,* by *if* as in (98), or by the rhetorical particle *right* (as in *which is true, right?)*

(96) *And the other thing about Andy Cohen, I notice he talked about, that I'm the only person who goes to Brazil and gets paler, <u>which is true</u>.*[129]

(97) *Yeah, but there's never been solid studies about it. ... It's been more than 13 years <u>which is true but</u> relatively brand-new - but in the previous wars, you didn't have any of your kids until after you came back from the war.* [130]

(98) *They say that spring is when love is in the air, <u>which is true if</u> you love pollen and sneezing.* [131]

Recall from the body partitioning profile given earlier in this section that [WIT] was divided almost evenly between gesture and upper body (50% or 9/18 enactments) and upper

[129] Red Hen: 2013-04-06_0737_US_KNBC_Late_Night_with_Jimmy_Fallon,1338

[130] Red Hen: 2015-05-25_1400_US_KABC_Good_Morning_America,1020

[131] Red Hen: 2014-05-15_2300_US_KNBC_The_Ellen_DeGeneres_Show,145

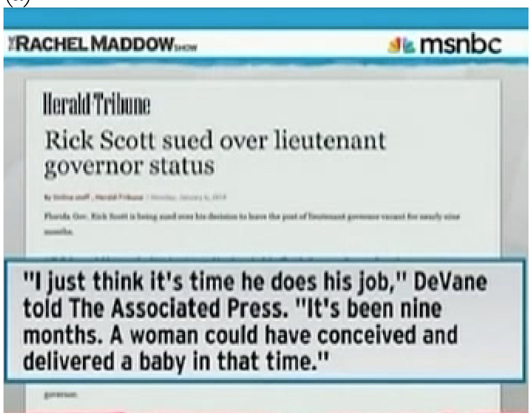
body only (40%, or 7/18 enactments), with the remaining two enactments being gesture only. A closer look at the distribution of articulator forms is given in Table 5.7. Of the 16 enactments that featured upper body movement, 14 featured head movement (78% of the utterances). Head movement is, therefore, the most frequent marker of [WIT] and is divided between turns and nods. One third of the utterances were also being accompanied by brow raises. Turning to the gesture form for [WIT], the PUOH gesture form dominated, playing a role in one third of the 18 enactments of [WIT].

Table 5.7. Distribution of articulator forms: *which is true*

Articulator	Tokens (n=18)	Articulator form
Gesture	10	6 PUOH, 1 purse, 1 PLTC, 1 palm forward, 1 point
Brow	5	5 brow raises
Head	14	6 turns, 5 nods, 3 tilts
Shoulders	1	1 shrug
Torso	1	1 shift

The [WIT] example pictured in Figure 5.6 shows a PUOH gesture with a head turn and brow raise, the dominant profile that emerged from the annotations summarized in Table 5.7. The profile corroborates what we know about these specific gestural and head forms. The PUOH gesture is frequently used to present information or as a marker of obviousness. The use of the PUOH in the [WIT] context thus enacts the epistemic strength of the construction; i.e., when aligned with [WIT], the PUOH gesture is used to underscore the speaker's commitment to the truth of the utterance.

(a)



S: *(Rachel Maddow narrates text above)*

(b)

which is true

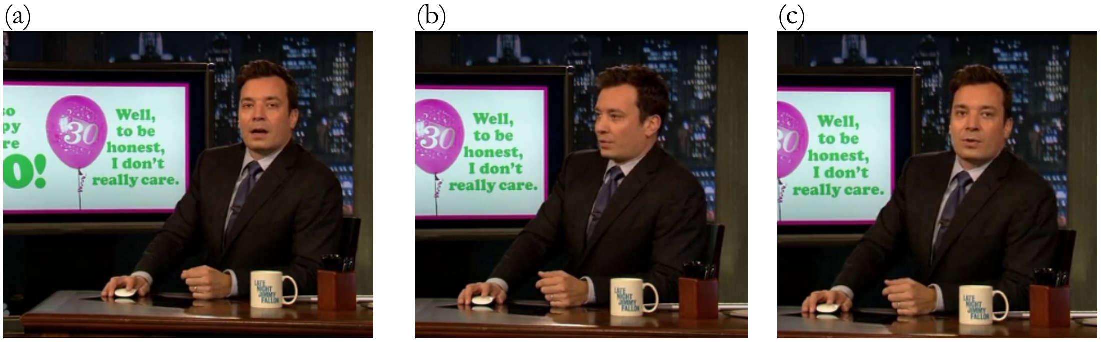
G: palm-up open-hand gesture, head turn, brow raise

Figure 5.6. *which is true*, PUOH and head nod

Finally, the use of physical space to quite literally create an aside with one's body is seen in a few gesture forms as well as in the head turns. With regard to gesture space, given its function of creating an aside in discourse, it is no surprise that two of the instances enacted this bodily metaphor in the gesture form by displacing their gesture form to the side of their body. This was done in one instance with a point hand shape with each hand and in another instance with a palm-lateral container gesture. The latter is shown in Figure 5.7. The relatively equal representation of two head movements, namely nods and turns, shown in Table 5.7, also bears some scrutiny. The five nods had the quality of an emphasis marker that were enacting the epistemic strength of the assertion. In the same way as the lateral gestures pictured in Figure 5.7 move a discourse object to the side and back, the six head turns seemed to highlight the salient characteristic of the utterance of marking the literal 'aside'. That is, the head turns mark the discourse function through the speaker's look to the side and breaking the gaze with the interlocutor to make the aside. An example of this is shown in Figure 5.8.

(a)	(b)	(c)
S: … *and you have this nice corporate structure that is legally going to focus on its own self-interest,*	*which is true*.	*But you have this influence of political power, financial influence that think about issues beyond their short-term needs…*
G: palm-lateral container gesture in center of gesture space	palm-lateral container gesture moves to the right	hands return to center, head returns to neutral position

Figure 5.7. *which is true*, palm lateral gestures to each side



(a)	(b)	(c)
S: *It's a balloon, it says "I'm so happy you're 30," and inside it says, "Well to be honest, I don't really care."*	*ugh, which is true*.	*We're down to our last card here.*
G: brow raise and head shake on *don't really care*	head turn and nod downward	head returns to neutral

Figure 5.8. *which is true*, head turn

In sum, the [WIT] construction is characterized by the PUOH gesture and head turns and nods. The nods appear to profile a positive assessment in this context – affirming the *truth* of the utterance in a similar epistemic assertion as the PUOH gesture in this context. The frequency of head turns in this context, and the quality of the movement, suggest that they mark the parenthetical nature of the utterance, i.e. the speaker turns away to make an aside, and then returns to neutral following the aside.

I now turn to another embedded relative clause, which differs only in the adjectival slot from *which is true,* namely *which is fine.*

Which is fine

In current usage, *fine* is frequently used to mean satisfactory or acceptable. Moreover, it sometimes implies that "the speaker does not in fact agree with the preceding statement, or is not happy with the situation in question."[132] The examples of the [WIF] construction in (99) to (101) show the range of uses and placements with regard to clause structure. In the data set examined here, 12/20 utterances followed the usage pattern in (99), in which it occurs at the end of a clause. The latter two examples show [WIF] followed by *because* (featured in five of 12 utterances) and *but*, which occurred just once.

(99) A: *Do people know who you are?*
B: *No. Nobody knows who I am, <u>which is fine.</u> On the subway, it's great not to be recognized.*[133]

(100) *For me personally, it's probably one of the worst decisions because I have to wear a pregnancy suit, <u>which is fine</u> because I wear Spanx and girdles, so that was fine…*[134]

(101) *So a lot of people don't think that you can wear a snow boot to work. They think you have to change <u>which is fine</u>, but we're going to show you can bring it to work and take it with you the whole day.*[135]

The kinesic annotations for [WIF] are shown in Table 5.8. There are several findings here that make [WIF]stand out from the other excursion expressions. Firstly, of all the parenthetical expressions, [WIF] has the highest number of gestures as well as the most uniform gesture profile. By this I mean that these three forms account for 15 gestures. There

---

132 OED Online. Accessed 15 November 2019. www.oed.com/view/Entry/70361

133 Red Hen: 2015-08-03_2300_US_KNBC_The_Ellen_DeGeneres_Show,1612

134 Red Hen: 2015-03-11_2300_US_KNBC_The_Ellen_DeGeneres_Show,1683

135 Red Hen: 2015-01-23_1600_US_KABC_Good_Morning_America,1992

are no gesture forms with only one instance, for example, a point, or someone reaching out, which we see in the gesture forms for the other asides. Secondly, as the table shows, [WIF] is dominated by palm-down or palm-away gestures. Together, these account for over 73% of the gestures. (As we will see in its profile below, *which, by the way* is the only other expression to feature more than one palm-down or holding away gesture, with 3 of each).

Table 5.8. Distribution of articulator forms: *which is fine*

Articulator	Tokens (n=18)	Articulator form
Gesture	15	6 PDOH, 5 holding away, 4 PUOH
Brow	6	6 brow raises
Head	13	8 turns, 5 shakes
Shoulders	0	n/a
Torso	3	2 leans back, 1 lean forward

As shown in Table 5.8, the two prominent hand forms were palm-down and holding away gestures. Three examples of holding away gestures are shown in Figure 5.9. Palm-down gestures are very similar to the holding away gestures pictured. In fact, the palm-down and holding away gestures for *which is fine* shared more similarities than differences in their execution. When the angle of the forearm was closer to horizontal, the resulting gesture is palm-down, whereas a more steeply angled forearm results in the holding away gestures shown in Figure 5.9. Figure 5.10 shows somewhat of an anomaly in the palm-down gesture in that it is one-handed and clearly very forceful. It also features a lateral gesture stroke prior to the final hold pictured. In this way it resembles the forceful outward moving *quit* gesture shown in Chapter 3.

(a) (b) (c)

(a) S: *By the way, what most people say they're planning on getting their dads for father's day? Ties or cards - which is fine. To be honest, it's fine.*

G: bimanual holding away gesture with head turn

(b) S: *I think -- I think I'm going to end up working on my birthday, which is fine. To be working on your birthday is a gift.*

G: bimanual holding away gesture with head shake

(c) S: A. *So basically, every time he comes on your show, he's lying to you.*

B. *Apparently so. Which is fine, You're not under oath.*

G: bimanual holding away gesture with head shake and lean back

Figure 5.9. *which is fine,* three examples with holding away gesture

S: *It's interesting, those guys up there were standing up. The people down here were not standing up. Which is fine. Any way, any way of chanting my name is still wonderful.*

G: palm-down gesture (originates with lateral gesture stroke across body from left to right), with brow raise and lean back

Figure 5.10. *which is fine,* palm-down gesture

The PUOH form also occurs with *which is fine*, although not as frequently. I provide an example of the PUOH in Figure 5.11. Here, the PUOH form, *which is fine* evokes a stance that is more accepting and lacks the concessive or negative element.

S: A. *Ellen: do people know who you are?*

B: *No. Nobody knows who I am, which is fine. On the subway, it's great not to be recognized.*

G: bimanual PUOH

Figure 5.11. *which is fine,* PUOH gesture

Finally, upper body movement for *which is fine* consisted of head turns and shakes, as well as several torso leans backwards. This is the only expression with more than one head shake. Again, the head shake has many functions, but is often used in situations involving hedging or negation. Although the low number of tokens prevents any strong conclusions, it is interesting to note that four of the five head shakes occur with palm-down or holding away gestures, which implies the same function as the [WIF] enactments with palm-down and holding away gestures that mark concession or negative stance. (The other instance of the head shake occurs with no gesture).

In sum, the profile of [WIF] is dominated by palm-down and holding away gestures, as well as head turns. The holding away and palm-down gesture forms are both known to be associated with negation and concession, and it would seem that this is the case here. These forms seem to enact the weakened stance that the OED refers to, of being acceptable, but not optimal. This interpretation is supported by the co-occurrence of head shakes aligned with four of the palm-down gestures. Head shakes are not seen when [WIF] is gestured with a

PUOH form; in those cases, [WIF] lacks the negative assessment that the palm-down and holding gestures and head shakes convey.

Which, by the way

In contrast to the previous two excursion markers, *which, by the way* ([WBTW]) cannot stand alone. Rather, it introduces a longer excursion in which the speaker provides more information or elaborates on her stance, as in (102) and (103).

(102) *But whenever I'm sick I always Google what it is. One time I was convinced I had rabies which by the way you can't go back. If you have rabies you're dead but I did a google search…*[136]

(103) *But yeah, because of that picture, the hairstylist, which by the way, that hair, I can't even believe that hairstylist got work again. [light laughter] but he recommended me to 'elite'.*[137]

As shown above, [WBTW] is highly enacted (90%) and, like the other excursion expressions, recruits the upper body in particular (100% feature upper body movement). The summary of kinesic annotations for [WBTW] is given in Table 5.9.

Table 5.9. Distribution of articulator forms: *which, by the way*

Articulator	Tokens (n=18)	Articulator form
Gesture	12	3 PDOH, 3 holding away, 2 points, 2 PLTC, 1 PUOH, 1 reach out
Brow	4	1 brow raise
Head	14	7 turns, 5 tilts, 2 nods
Shoulders	0	n/a
Torso	4	3 leans forward, 1 side step

In addition to [WIF], [WBTW] is the only other excursion marker to feature palm-down and holding away gestures, as shown in Figure 5.12.

136 Red Hen: 2015-10-13_1700_US_KABC_The_View,1623

137 Red Hen: 2015-09-29_0737_US_KNBC_Late_Night_with_Seth_Myers,1740

(a) (b) (c)

S: and then one guy, which was --	*by the way, I had the greatest time... uh...*	*making this movie.*
G: (speaker on left) neutral	holding away gesture	neutral

Figure 5.12. *which, by the way,* holding away gesture

(a) (b) (c)

S: *Rumor is he's going to make an announcement by Thursday*	*which,*	*by the way, is also the day that front-runner Hillary Clinton is testifying before a committee about Benghazi.*
G: slight held head tilt to right	point gesture and further head tilt	point gesture moves to left and is held, nod

Figure 5.13. *which, by the way,* point gesture and head movement

There were also two points, one of which is shown in Figure 5.13. Note the use of lateral space to mark the aside, with the point gesture moving right-to-left (R-to-L) in front of the speaker. This R-to-L movement to mark the parenthetical comment suggests contra-flow to the normal 'direction' of discourse, which is often construed as moving from L-to-R. The head movements consisted of turns and tilts, which did not differ substantially from other expressions.

How shall/should I put it?

The [HSIPI] construction differs in two ways from the parenthetical asides profiled thus far: it cannot be modified or complemented and it asks a rhetorical question. I provide some linguistic examples and then present the main findings of the kinesic components that I annotated for these expressions.

As the examples below demonstrate, [HSIPI] usually interrupts a clause at a phrasal boundary and thus is truly a disruptive digression from the main utterance. It gets interjected before predicative modifiers in (104)-(106) and between the subject and the verb phrase in the example illustrated in Figure 5.14. As such, it functions as a signal that a very pointed opinion follows. The stance taken is usually harsh, but it may be ironic. For example, in (104), the speaker is evaluating the book in question as *a pack of lies.* In (105), [HSIPI] signals the highly stanced delivery of *memorable* in *very memorable lunch.* [HSIPI] is a direct signal to an interlocutor that what follows is an evaluation or estimation by the speaker.

(104) *He's pretending not to be a candidate for the White House. The media is going along with it. The book is, <u>how should I put it,</u> a pack of lies.* [138]

(105) *We had a Christmas break after our first week's rehearsal and there was a very…<u>how shall I put it,</u>…very, uh, uh, memorable lunch.*[139]

(106) *We all know that kings and presidents have, <u>how should I put it</u>, a bumpy or active love life.* [140]

The kinesic annotations for the [HSIPI] construction are shown in Table 5.10. While this expression was highly enacted (90%), it featured far less expression in the hands than the excursion markers with *which*. With a gesture rate of only 47%, only *although, you know* was less

[138] Red Hen: 2011-02-02_0400_US_MSNBC_The_Last_Word_With_Lawrence_Odonnell,2719

[139] Red Hen: 2007-10-23_0635_US_KCBS_Late_Show_with_Dave_Letterman,164

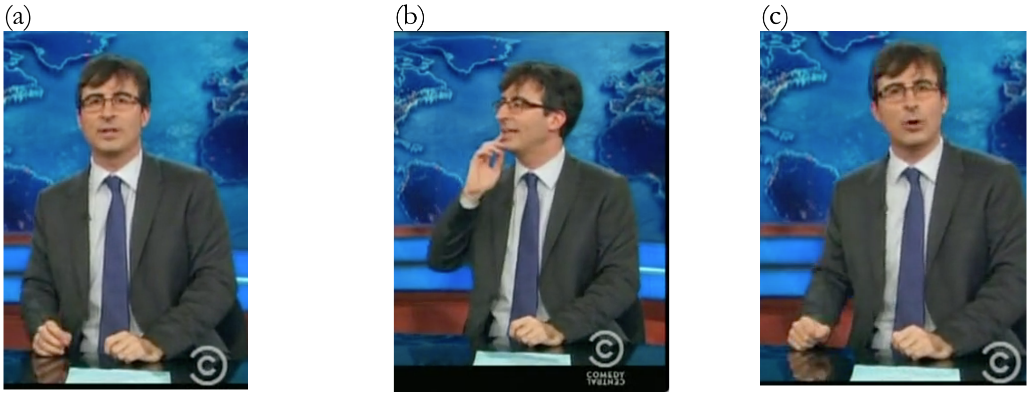
[140] Red Hen: 2008-01-14_1500_US_KABC_Good_Morning_America,1646

frequently gestured. Of the eight that were gestured, three were self-touch gestures, a form which did not occur elsewhere in my Red Hen video clips. An example of a self-touch is shown in Figure 5.14. While this enactment does seem particularly performative in nature, the other two enactments with self-touch did not feature this performative quality.

Table 5.10. Distribution of articulator forms: *how shall I put it*

Articulator	Tokens (n=17)	Articulator form
Gesture	8	3 self-touch, 2 PLTC, 2 PUOH, 1 beat
Brow	0	n/a
Head	11	6 turns, 5 tilts
Shoulders	0	n/a
Torso	0	n/a
Gaze[141]	11	8 up, 3 down

(a) (b) (c)



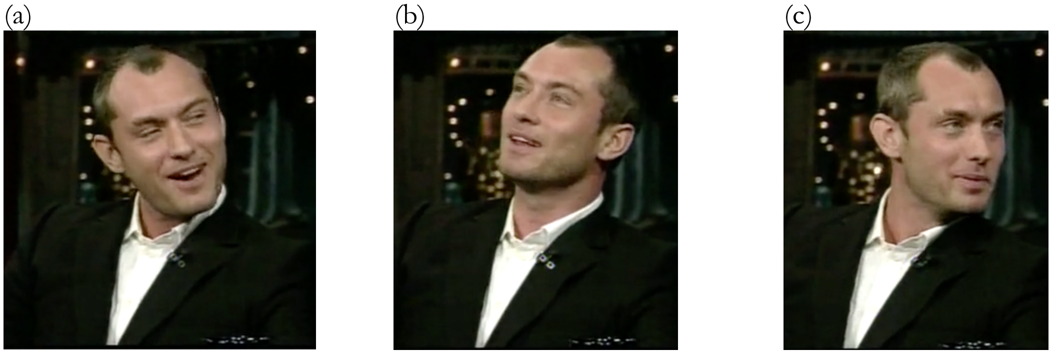
(a)	(b)	(c)
S: *So this reporter…*	*…how should I put it,…*	*he tossed another question on the barbie.*
G: hands resting on table	self-touch gesture to chin, head turn	return to neutral

Figure 5.14. *how should I put it,* self-touch gesture and head turn

[141] As explained in §5.2 and in further detail below, gaze featured so prominently for [HSIPI] that I captured gaze movement in the annotations.

Compared to other excursion expressions, [HSIPI] had the lowest occurrence of head movement, at 65%. What differs for [HSIPI] is that the other upper body articulators were not involved, i.e. there were no brow, shoulder, and torso movements. (The movement in Figure 5.14(b) was coded as head turn rather than torso shift, since the shoulders stay square to the viewer). What this means is that, when there was movement of the upper body, it was all head movement, as in the example above and in Figure 5.15.

(a) (b) (c)



(a)	(b)	(c)
S: *We had a Christmas break after our first week's rehearsal and there was a very…*	*…how shall I put it?…*	*…very, uh, uh, memorable lunch.*
G: hands clasped on lap, head turned to interlocutor	head turn and tilt, gaze upwards	head turn back to interlocutor

Figure 5.15. *how shall I put it,* head movement only

Furthermore, while I did not code gaze as a part of the regular annotations, for [HSIPI] it was immediately clear that gaze plays a key role in its kinesic form. As discussed in the Methods in Chapter 2, gaze proved too difficult to assess by observation; however, for [HSIPI] the gaze movements were prominently marked. I thus added it to the annotations for this profile only. Of the 17 enacted utterances, 11 featured gaze shift (8 upward and 3 downward) and 4 of these were enacted *only* through gaze shift.

In essence, [HSIPI] shows a profile that is restricted in its recruitment of articulators. The upper body is always involved, but only through gaze and head movement. The rhetorical semantics of the question in [HSIPI] suggests that the speaker is thinking or in a lexical search mode. At the same time, the speakers enact the thinking process by tilting their head or shifting their gaze upwards, with the occasional self-touch gesture as pictured in Figure 5.14. This physical manifestation of the process of thinking seems to be the iconic underpinning of the [HSIPI] profile. In the next section, I introduce the final excursion marker, *although, you know*.

Although, you know

Although, you know ([AYK]) is a compound excursion marker consisting of the concessive *although* and the discourse collocation *you know*. Of the 18 excursion markers for which frequencies were given in the methods section earlier in this chapter (Table 5.2), [AYK] was among the six most frequent in $COCA_{sp}$, with 95 items. The most frequent items directly following [AYK] were first-person singular and first-person plural subject pronouns, which together accounted for 22 of the 95 utterances. Sample sentences from the Red Hen dataset are given in (107)-(109).

(107) *Although, you know, I know that she gets a lot of attention for, like when she did 'Monster' she gained all this weight, you know for 'Monster'.* [142]

(108) *Although, you know. Well, I'm from New Orleans, so I used to talk a little more like that, but…*[143]

(109) *Although, you know, we got in fights -- arguments and they would have us put on boxing gloves and go out in the backyard and fight.*[144]

The kinesic form for [AYK] featured palm-up gesture forms and head tilts as the dominant movements. A reminder here that the degree of enactment was lower for [AYK], at 55%,

---

142 Red Hen: 2011-12-07_0837_US_KNBC_Late_Night_with_Jimmy_Fallon,2003

143 Red Hen: 2012-08-20_2300_US_KNBC_The_Ellen_DeGeneres_Show,2059

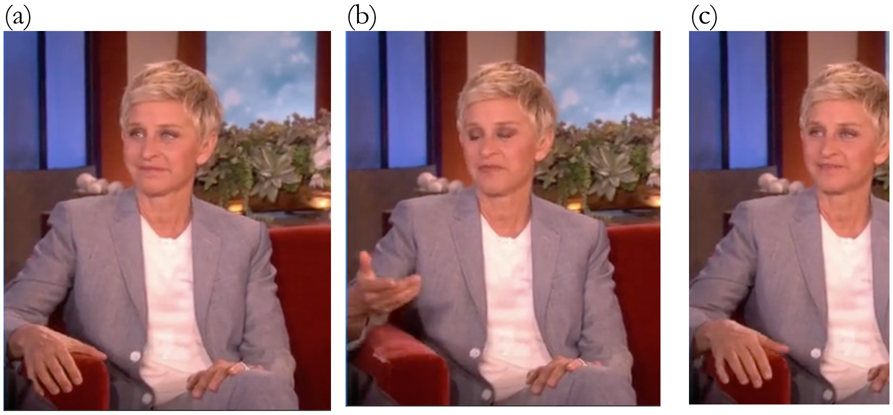
144 Red Hen: 2010-03-31_0735_US_KNBC_Late_Night_with_Jimmy_Fallon,2421

which meant that only 11 utterances were enacted. The results of the annotations are shown in Table 5.11.

Table 5.11. Distribution of articulator forms: *although, you know*

Articulator	Tokens (n=11)	Articulator form
Gesture	5	4 PUOH, 1 PDdiTC
Brow	3	3 brow raises
Head	11	7 tilts, 2 turns, 1 shake, 1 nod up
Shoulders	0	n/a
Torso	2	1 back and forth, 1 turn

(a) (b) (c)



S: *Yes.* | *Although, you know.* | *(pause)*

G: hands at rest | PUOH right hand, head tilt, mouth shrug | return to neutral

Figure 5.16. *although, you know,* PUOH and head tilt

An example of [AYK] with PUOH and head tilt is given in Figure 5.16. While I was not annotating mouth movement, note the mouth shrug here in Figure 5.16(b) as well.

When compared with the other excursion markers, two elements of the profile for [AYK] stand out. It has both the lowest co-occurrence of gesture forms (5/11 enactments, or 45% featured gesture) and the highest degree of head movement, with 11/11 featuring head movement. By comparison, the next highest excursion markers with head movement were [WIT] and [WBTW], both with 78%. Furthermore, seven of the 11 head movements were tilts. [AYK] is also the only excursion marker to be dominated by head tilts.

Lastly, there was a quality to the gesture and head movements that was not captured in the annotation schema, namely the presence of equivocation or indicating 'so-so' or 'comme ci, comme ça' movements. Gestures of this nature were, for example, palm-down forms with splayed fingers that featured a wobble-type movement. In the head, a quick back-and-forth head tilt seemed to convey the same equivocating pragmatic force. An example of one of these 'comme ci, comme ça' sequences aligned with *although, you know* is given in Figure 5.17, in which the gesture and head are both involved in expressing the equivocation or hedging stance. While this pattern occurred in enactments with both head movement and gestures and head movement only, the head movement alternation in cases without gesture was often more subtle and poorly captured in screenshots. Qualitatively, however, both cases give the impression of an equivocation.

(a) (b) (c) (d)

(a)	(b)	(c)	(d)
S: *It was absolutely*	*great.*	*Although*	*you know we saw Mach Pichu…*
G: bimanual PUOH, posture lean, head tilt	all articulators return to center	bimanual PUOH, posture lean (R),head tilt (R), and shoulder shrug (exaggerated version of (a))	hands in reversal of PUOH gesture form in (c), posture lean (L), head tilt (L)

Figure 5.17. ***although, you know,*** **equivocation gesture**

In sum, while it had the lowest degree of enactment, [AYK] featured the highest regularity in head movement, it was one of only two expressions in which the PUOH was the dominant gesture form, and there was an equivocation, or hedging, element to several of the gestures. This concludes the case studies for the excursion markers. In the next section, I provide a summary and comparison across all excursion expressions.

5.3.2 Summary: Excursions

Up to this point, I have outlined the principal findings for the excursion-marking expressions *which is true; which is fine; which, by the way; how shall I put it;* and *although, you know*. By annotating 20 of each of these utterances, I established the degree of enactment, which ranged from 55% for [AYK] to between 85% and 90% for the other expressions. I also presented data for body partitioning, which captures whether an utterance type prefers to be marked only with gesture, only with the upper body articulators, or both gesture and upper body. Findings showed that the markers with *which* all featured higher degrees of gesture and upper body use compared to upper body alone, which typified [HSIPI] and [AYK].

For each utterance, I also presented a kinesic profile of each of the articulators. The distribution of involvement of each articulator for each expression is shown in Figure 5.18. On first glance, this comparative view highlights some of the findings described for each utterance, for example, the strong role played by gaze for [HSIPI] and the dominance of head movement for [AYK].

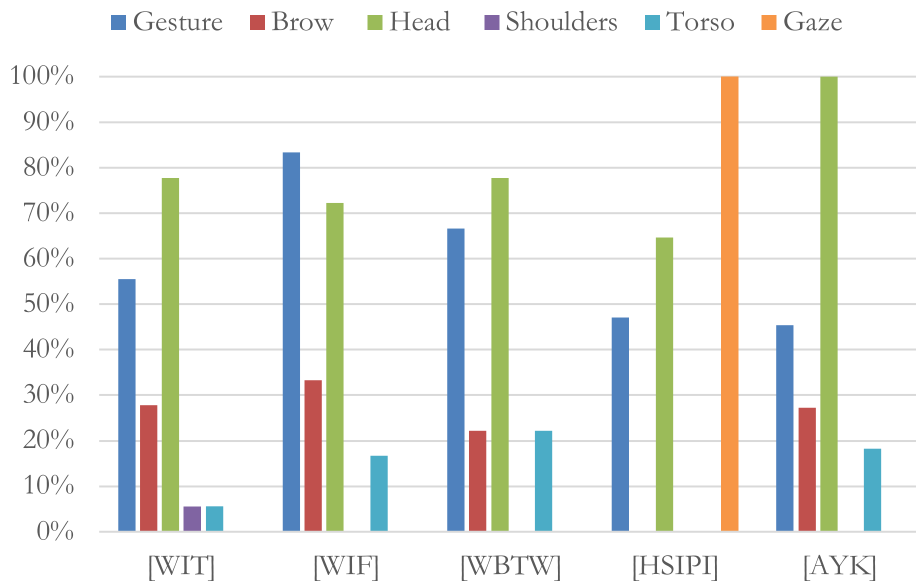


Figure 5.18. Distribution by articulator for all excursion expressions

The dominant forms for gesture and head movement are summarized in Table 5.12 and Table 5.13, respectively. The dominant hand forms for which is fine and which, by the way were the palm-down or holding away gesture (these are grouped together in this table) while the epistemic excursion marker which is true is marked by the PUOH. The equivocating aside *although, you know* is also marked by a PUOH. In addition to the form, I include the percentage of the enactments that this form accounts for. This provides an indication of how variable the gesture forms for that utterance are. As Table 5.13 shows, head movement was characterized predominantly by head turns for each aside marker other than [AYK]. To foreshadow the study of return markers in the next section, these results for head movement are in stark contrast to the dominant head movements for return markers, which feature a range of tilts, turns, and nods. With this in mind, I would suggest that the prevalence of head turns for the parenthetical asides can be viewed as multimodal marker of the discourse parenthetical function of these utterances. As stated above, the exception is [AYK], for which head tilts dominate. For this utterance type, it would seem that the pragmatics of cognizing and equivocating (marking the contrast between being more and less sure) outweigh the marking of the utterance as an aside.

Table 5.12. Dominant gesture form by excursion expression

	[WIT]	[WIF]	[WBTW]	[HSIPI]	[AYK]
Gestured*	56%	83%	67%	47%	45%
Dominant form	PUOH	PDOH/ holding away	PDOH/ holding away	self-touch	PUOH
Dominant form frequency**	60% (6/10)	73% (11/15)	50% (6/12)	38% (3/8)	80% (4/5)

* Gesture = gestured/total enacted
** Dominant form freq. = frequency of dominant gesture/total gesture

Table 5.13. Dominant head movement by excursion expression

	[WIT]	[WIF]	[WBTW]	[HSIPI]	[AYK]
Head*	78%	72%	78%	65%	100%
Dominant form	turn	turn	turn	turn	tilt
Dominant form frequency**	43% 6/14	62% 8/13	50% 7/14	55% 6/11	64% 7/11

* Head = head movement/total enacted
** Dominant form freq. = frequency of dominant head movement/total head movement

The data and findings presented here allow me to begin to build a multimodal profile for five expressions that mark parenthetical asides in discourse. In the next section, I apply the same methods to a set of return markers. Return expressions frequently follow a parenthetical comment and bring the speaker back to the main thread of the discourse or signal that she is moving on to a new topic.

5.4 Return markers

Speakers mark junctures as they progress along a discourse path. They add comments, articulate stances, and create arguments by establishing discourse relations between statements. In the previous section, I presented microstudies of highly stanced parenthetical statements that create asides, or slight detours, in the flow of discourse. In this section, I examine the devices speakers use when returning from such discourse excursions to the main utterance frame. Devices used to signal junctures include common concessive markers such as *however,* and *anyway*. Concessives create distance or express reservation about the comment in the previous utterance (Biber et al. 1999: 878). Other phrases that mark that a speaker is returning to a previous utterance include *as I was saying* and *where was I?*

I use three examples here to illustrate the range of embodied expression that co-occur with returns. As shown in the next three figures, when *anyhow, anyway,* and *at any rate* are used as return expressions, they are enacted in the body. Furthermore, in these examples, the articulators that are recruited are dominantly those of the upper body. In Figure 5.19, the speaker creates an irrealis scenario introduced by the utterance *It would be great if*…. There is then a distinct head tilt when she utters *anyhow*. The concessive device returns the speaker (and interlocutors) to the present realis space, in which she goes on to introduce another guest. Her utterance *anyhow* creates a contrast between two worlds: the irrealis world in which Mitch McConnell says what she desires him to say and the realis world in which he does not. The *anyhow* both signals the return to the realis world and simultaneously displays her stance towards the fact that her desired irrealis world is not achieved. The mouth shrug here indicates her stance, namely the futility of the wish she expresses in Figure 5.19(a). In this way, the composite of the head tilt and mouth shrug index the irrealis space and simultaneously express an evaluation of it (i.e., expressing the futility of reconciling the real world with the desired world).

(a)	(b)	(c)
S: *It would be great if Mitch McConnell would just say the tweets are no good, we don't agree with them, they're inappropriate, and they're bullying, and we don't stand by that.*	*Anyhow*.	*Joining us now is the…*
G: neutral head position	head tilt to speaker's right	head returns to neutral

Figure 5.19. *anyhow*, head tilt and mouth shrug

The indexing of space is equally present in Figure 5.20, a sequence of co-speech behaviours that are aligned with [AAR] Here, the enactment features a clear holding away gesture at the same time as an eyebrow raise and a postural shift. The holding away gesture distances the speaker from the preceding discourse segment, in which the show's host and the speaker in Figure 5.20 are speaking over each other after watching a video clip of a movie that the actor had been in recently overseas. The speaker says *at any rate* and performs a holding away gesture, which functions to signal a turn away from the previous discourse segment to a new comment and stance, in which he comments on how he feels about being back in the United States, where he lives. The raised eyebrows layered with a postural shift to the side and back function create physical distance from his previous posture and also a metaphorical distance from the previous utterance.

(a)	(b)	(c)	(d)
S: ..*a crusher.*	*At any rate, I'm very very happy to be back*	*in the States, I'm happy*	*to be back here.*
G: holds lean to right	holding away gesture, brow raise, head raise	torso shift up and to the side	rest

Figure 5.20. *at any rate*, holding away gesture, brow raise, head movement, and torso lean

In Figure 5.21, the speaker gives a shrug as she says *anyway*. In this scenario, there has been a great deal of audience laughter in the preceding seconds. To redirect the discourse the speaker begins with *Well, um, anyway*, accompanied by a shoulder shrug. Shrugs have been described as markers of obviousness (Jehoul et al. 2017), in addition to heavily stanced gestural enactments of disaffiliation (Debras 2017). Certainly, the speaker could be disaffiliating herself with the content of the previous discourse sequence; however, this shrug movement could also be considered an instance of self-relocation. It is highly pronounced in nature – with a brief hold at the top of the shrug, and is not aligned with any other markers of shrugs (mouth shrug, for example). With regard to the speaker's body position, the neutral posture in the centre of her body space is the one in which the speaker-gesturer is situated in the 'here and now'. Thus, the shrug performed here could be a re-centering of this speaker in her own space, i.e. it could be considered an ego-centric indexical shift that mirrors the centering of the discourse back to neutral from the preceding chatter.

(a) (b) (c)

(a) *S: …Well…* G: crossed arms

(b) *…um, any(way)…* shoulders rise

(c) *I did this awful thing on the subway* shoulders drop (at *way* of *anyway*)

Figure 5.21. *anyway*, shoulder shrug

In the three examples presented here, I have shown how the body marks the returns in natural discourse. Body articulations included head tilt, holding away gesture, and shoulder and mouth shrugs. The kinesic forms used show that these returns simultaneously mark stance. In the next section, I provide evidence that the type of patterns exemplified in the return utterances examined here are conventionalized. In studies of seven return expressions, a unique kinesic profile emerges for each expression.

5.4.1 Return microstudies

The return expressions for which I present a multimodal profile are: *anyhow, but anyway(s), so anyway(s), at any rate, however, nevertheless, that being said,* and *in any case.* Before presenting the individual case studies, I will provide an overview in which I summarize the rates of enactment and the body partitioning profile across all expressions. In individual subsections that follow, I then outline the specific profile that emerges from the quantitative analysis of each articulator. For each utterance, I provide a summary of the linguistic profile of the utterance, followed by a presentation of the quantitative analysis of the co-speech enactments that accompany the utterance. Lastly, I give a screenshot example from the corpus to illustrate the profile that emerges from the quantitative results.

To begin, I provide the degree of enactment and body portioning findings for all utterances together. In Table 5.14, I present the degree of enactment for each of the return constructions in the order of presentation in this chapter. As Table 5.14 shows, the body was recruited with the spoken return expressions in 72% to 96% of the cases. The most highly subjective return marker, given its explicit first-person subject, is [however, I], which is also the most enacted.

Table 5.14. Degree of enactment by return marker

	[anyhow]	**[but anyway(s)]**	**[so anyway(s)]**	**[AAR]**	**[however, I]**	**[nevertheless]**	**[TBS]**
Degree of enactment (raw)	18/25	19/25	21/25	23/25	24/25	23/25	20/25
Degree of enactment (%)	72%	76%	84%	92%	96%	92%	80%

The body partitioning profile, that is, the division of labour by each 'set' of articulators (gesture, upper body, or both gesture and upper body), is shown for each return expression in Figure 5.22. Here, the vertical axis represents the total number of search returns in each case study (25) and the relative frequency of the body partitioning zones for each expression.

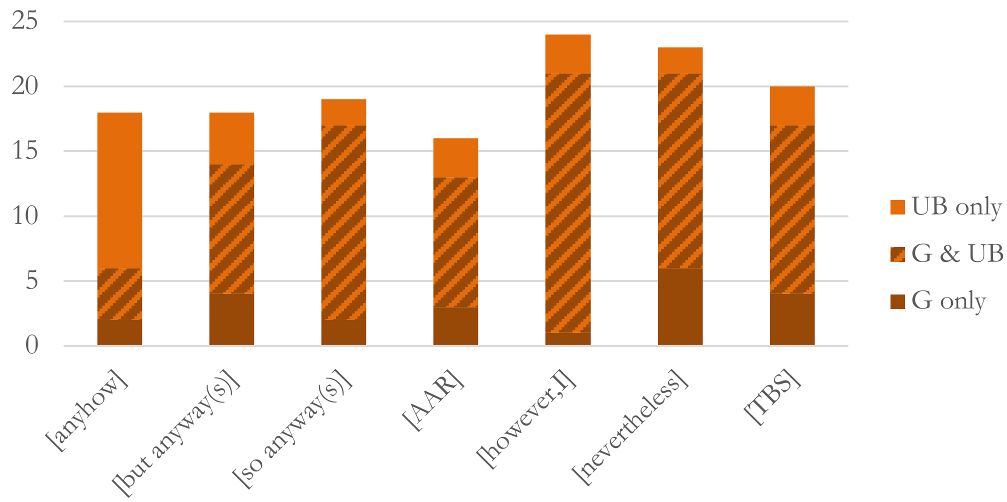


Figure 5.22. Body partitioning profile for return markers

While the trend is for most return expressions to feature a combination of gesture and upper body movements, the notable findings here are the tendency of [anyhow] to be expressed dominantly in the upper body, and the tendency of [however, I] to be expressed almost exclusively by synchronous gesture and upper body movement.

Anyhow

Anyhow is a common discourse marker that functions both as a stance adverbial and as a concessive linking adverbial (Biber et al. 1999: 879). As described in §5.2.1, the dominant use of *anyhow* is as a stance adverbial at the left periphery. Of 301 search returns in COCA$_{sp}$, only 27 occurred in sentence-initial position as a linking adverb, as in the examples in (110)-(113), which demonstrate the usage focused on in this study. Note that *anyhow* can precede an overt acknowledgement of a digression, as in (110). Despite this low frequency in COCA$_{sp}$, Red Hen had enough viable search returns between 2012 and 2018 to form a case study.

(110) *I've never had an empty nest. I wouldn't know what that's like. Anyhow, it's lovely to have the girls home, actually.*[145]

(111) *No, but, why can't they just answer -- anyhow, White House Secretary Sarah Huckaby Sanders also insisted the President felt vindicated by yesterday's developments.*[146]

(112) *That's just – it's really bad. Anyhow, I mean, Mark at some point don't you think the National Press Association ought to do something about this?*[147]

(113) *Bad day for Mika. Anyhow, how's the forecast, Bill?*[148]

Of the 25 search returns in Red Hen for *anyhow*, 7 of 25 featured *anyhow* at the beginning of an utterance after an interjection by another speaker. As shown in (114) and (115), these utterances generally featured an interjected acknowledgement from the interlocutor (e.g. *it was* and *yeah)*. I saw no reason to exclude these utterances from the case study.

(114) A: *Because, like, she was yelling at me, and then I said, "Well, why did you make fun of me for being fat?" It was the wrong time.*
B: *Yeah.*
A: *Anyhow, TMX had me with a wrench busting windows and then leaping on her car and saying, "You better call the coppers"*[149]

(115) A: *It was incredible.*
B: *It was.*
A: *Anyhow, thank you to everybody. My team, awesome.*[150]

The enactment of [anyhow] when it is used as a linking adverbial is characterized by upper body movements, particularly head movements; over half of the enactments featured head movement. The distribution of movements by articulator is shown in Figure 5.23.

145 Red Hen: 2017-09-05_1100_US_MSNBC_Morning_Joe,102

146 Red Hen: 2017-10-31_1000_US_MSNBC_Morning_Joe,1756

147 Red Hen: 2017-06-29_1100_US_MSNBC_Morning_Joe,6845

148 Red Hen: 2017-12-12_1200_US_MSNBC_Morning_Joe,250

149 Red Hen: 2017-10-28_0737_US_KNBC_Late_Night_With_Seth_Meyers,2044

150 Red Hen: 2017-10-31_1000_US_MSNBC_Morning_Joe,3410

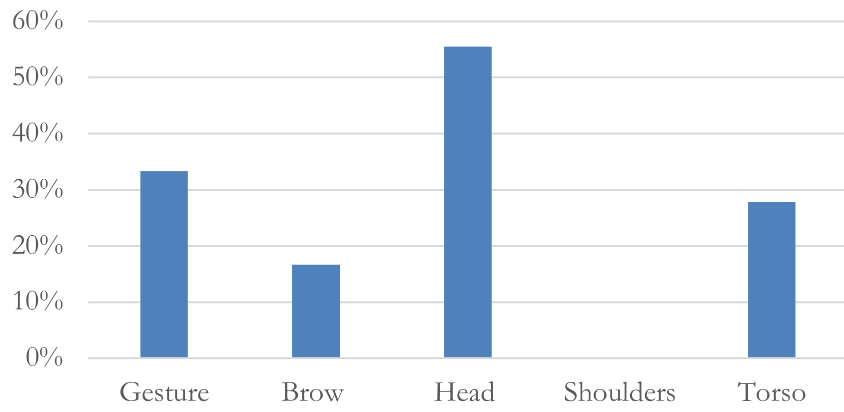


Figure 5.23. Distribution by articulator: *anyhow*

The type of movement associated with each articulator is given in Table 5.15. There were only six enactments with manual gesture. When speakers gestured with [anyhow], it was with a palm-down or palm-up, open palm gesture in the centre of their gesture space. The 10 head movements were evenly distributed among tilts to the side, tilts upward or downward, 'waggles' (a non-technical term for wobbling the head back and forth, usually from side-to-side), and two turns to the side. Torso movement was generally a backward lean, a postural shift that has been shown to mark disaffiliation (and, in argumentative discourse, may signal the ceding of a turn or argument space that had been taken through an earlier lean forward (Wehling 2017)). Of the five torso leans, three were aligned with a PUOH or PDOH gesture.

Table 5.15. Distribution of articulator forms: *anyhow*

Articulator	Tokens (n=18)[151]	Articulator form
Gesture	6	2 PDOH, 2 PUOH, 1 point, 1 palm-forward
Brow	3	n/a
Head	10	3 tilts, 3 nods, 2 'waggles', 2 turns
Shoulders	0	n/a
Torso	5	4 leans back, 1 lean forward

In sum, when speakers use *anyhow,* they most often recruit their upper body, most consistently head movements, and they do this without recruiting their hands to gesture. This prototype is exemplified in Figure 5.19, given in the previous section.

At any rate

At any rate ([AAR]) is characterized by utterances such as those in (116)-(119). Of the 25 tokens, 5 featured *but* or *though* in conjunction with [AAR]*,* as in *at any rate though,* as in (116), and *but at any rate*, as in (118) and (119). I included utterances following an interjection by the interlocutor, such as in (119), as I did for [anyhow] above.

(116) *On the other hand, they know that the south is going to be pretty tough for them. At any rate, though, they are still raising millions of dollars very quickly.* [152]

(117) *Like I said, if there was a chamber of commerce for ladies of the evening, he would be the head of it. At any rate, I included a story about that in the book but then realized, oh, my god, my children are going to read this.* [153]

[151] As per the presentation of excursion findings, totals for the number of tokens exceed the number of tokens, as many of the instances had several articulators in play at the same time. The same is true for the remainder of the *Distribution of articulator forms* tables in this section.

[152] Red Hen: 2016-03-01_2000_US_MSNBC_The_Place_for_Politics_2016,2290

[153] Red Hen: 2016-07-05_1700_US_KABC_The_View,2988

(118) *I think that polling runs between debates, too. If he gets to it, I think he would be on the stage for the second debate. But at any rate, what is – what impresses me is that both candidates now, from major parties, seem committed to this debate schedule.*[154]

(119) A: *There have been a lot of things written about the beach boys but I've never written anything –*
B: *Why are they laughing?*
A: *But at any rate, I just wanted to do it for my children, friends, fans.*[155]

[AAR] is one of only two return expressions examined here that feature more head movement than gesture (the other is [anyhow]). Gesture features in 57% of the enactments and the head marks the return expression in 65% of the cases, as shown in Figure 5.24. The full distribution of articulator forms is provided in Table 5.16. As displayed in the table, [AAR] is dominated by the palm-down form and head nods and tilts.

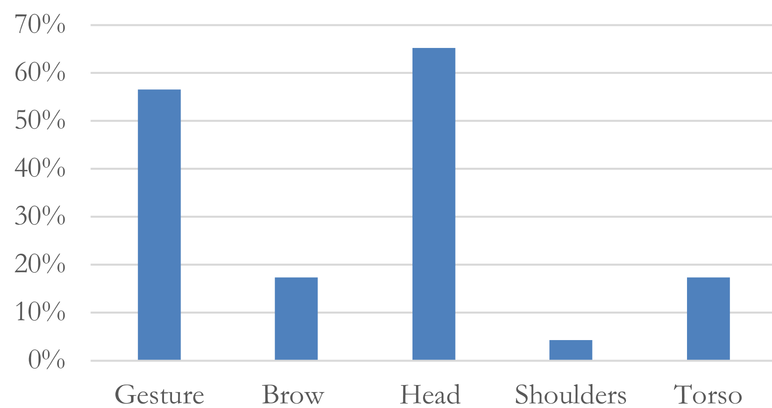


Figure 5.24. Distribution by articulator: *at any rate*

[154] Red Hen: 2016-09-08_1400_US_MSNBC_The_Place_for_Politics_2016,2944

[155] Red Hen: 2016-09-15_1700_US_KABC_The_View,2801

Table 5.16. Distribution of articulator forms: *at any rate*

Articulator	Tokens (n=23)	Articulator form
Gesture	13	6 PDOH, 2 PUOH, 2 PLTC, 1 palm forward, 1 pursed
Brow	4	4 brow raises
Head	15	6 nods, 5 tilts, 3 shakes, 1 turn
Shoulders	1	1 shrug
Torso	4	3 leans, 1 turn

An example of *at any rate* was given in Figure 5.20.

So anyway(s) and *but anyway(s)*

Anyway(s), and its frequent collocations, *so anyway(s), but anyway(s),* are among the most frequent mechanisms speakers use to end a digression and return to a main utterance in $COCA_{sp}$. The [but anyway(s)] construction gains concessive force form the *but* concessive marker, while the continuation marker, *so,* contributes a continuative quality to [so anyway(s)] (Schiffrin 1987). What these two phrasal discourse markers share is that the elements, *but* or *so* plus *anyway(s)*, can either combine to form a collocational discourse marker or, in cases with a discernible pause between them, they can be taken as a sequence of individual discourse markers (Lenk 1998: 85). I discuss this factor for each expression in turn. I begin by providing the quantitative results for the case study, firstly, of [so anyway(s)], and then for [but anyway(s)].

So anyway(s)

Typical examples of [so anyway(s)] from the dataset include those in (120) and (121).

(120) *It started with off the record. Now, this is Wikileaks. The word* leaks *is right in the name. There is no off the record with them. So anyway, Junior Donald tried to minimize his relationship with Wikileaks on his twitter account but they did have a relationship.*[156]

---

[156] Red Hen: 2017-11-15_0735_US_KABC_Jimmy_Kimmel_Live,375

(121) A: *And I would sit with her for 45 minutes and hear about the day's events.*
B: *Because you're a good son.*
A: *Because I'm a good boy. So, anyway, for two years, my wife and brother and baby lived on the other side in 91.*[157]

So anyway(s) was generally aligned with movement across several body articulators. As shown in the body partitioning summary above (Figure 5.22), 71% of the enactments were articulated both in the hands and in the upper body. A more detailed look at the involvement of articulators summarized in Figure 5.25 shows that the upper body involvement is largely due to head movement and that the hands and head were involved to the same degree, in over 80% of the enactments.

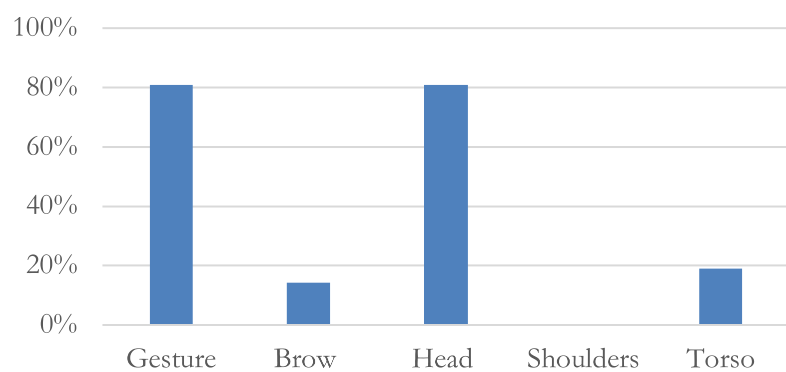


Figure 5.25. Distribution by articulator: *so anyway(s)*

The detailed annotations by articulator for *so anyway(s)* are shown in Table 5.17. Here, the counts show that the dominant hand form is the PUOH, followed by palm-lateral and palm-down gestures, which are featured in Figure 5.26. The head movements were fairly evenly distributed across all forms.

[157] Red Hen: 2017-10-21_0635_US_KCBS_The_Late_Show_With_Stephen_Colbert,2865

Table 5.17. Distribution of articulator forms: *so anyway(s)*

Articulator	Tokens (n=21)	Articulator form
Gesture	17	8 PUOH, 5 PLTC, 3 PDOH, 1 other
Brow	3	3 brow raises
Head	17	5 turns, 5 nods, 4 shakes, 3 tilts
Shoulders	2	2 shrugs
Torso	4	2 lean back, 1 lean forward, 1 turn

(a)

S: *…the voice she was able to determine was not you.*

G: left shoulder lean on desk

(b)

So anyway, you brought something along here that I think is going to be very helpful.

PUOH (begins on *anyway* and continues over full segment)

Figure 5.26. *so anyway(s),* PUOH

(a) (b) (c)

S*: um*…<u>*so*</u>…	… <u>*anyway*</u>.	*he comes to find out that his first commander…*
G: neutral, hands clasped on kne	palm-lateral towards centre container gesture (PLTC) gesture, head tilt right	hands clasped together on knee (out of view), head tilt towards interlocutor

Figure 5.27. *so anyway(s)*, PLTC and head tilt

I now return to the issue of whether *so anyway(s)* in my search returns occurred as one intonation ‘chunk’ or two. To determine this, I checked for a planning pause between the two discourse markers in each token in the dataset. There were three utterance tokens that had discernible pauses (> 300msec) between *so* and *anyway(s),* measured in ELAN video annotation software, and 18 that were delivered fluently (i.e. with no discernible break between *so* and *anyway*). Here, though, it also became clear that the length of the pause was not the only determiner of whether the two segments were perceived as one intonational chunk; rather, the co-speech gesture could also be used to observe this. I therefore also observed the alignment of the onset of the gesture stroke with each element in the construction (e.g. with *so* and *anyway(s).* I coded for whether the onset of the stroke occurred on or just prior to *so* or on or just prior to *anyway(s).* In 38% (8/21) of the enactments, the gesture stroke onset aligned with the speech onset of *anyway(s),* while in the remaining 62% (13/21), the stroke onset was aligned with *so.* As might be expected, in the three utterances that featured discernible pauses (>300 msec), the stroke onset always aligned with *anyway(s).* A more comprehensive examination of the role of prosody on whether *so anyway(s)* constitutes one or two intonational chunks is beyond the scope of this case study. In arguing for the conventionalization of the collocation *so anyway(s),* though, it was important to understand, at the very least, how many of the utterances in the data set used here might actually inhere to the expression of *anyway(s)* rather than *so anyway(s).* I return to this discussion for *but anyway(s),* which I treat next.

In sum, *so anyway(s)* was characterized by the PUOH and palm-lateral gestures and was strongly enacted in the head as well. More data and further analysis is required to investigate patterns of the fixed collocation *so anyway(s)* as compared to the continuation marker *so* and the discourse marker *anyway(s)* independently.

But anyway(s)

In discourse analysis literature, [but anyway(s)] has been described as both a marker of contrast (Quirk et al 1985: 674 and Halliday and Hasan 1976: 270) and "as a signal for cutting short a digression or topic" (Lenk 1998: 88ff). Rather than seeing these as independent functions, I view these as intertwined. For moving away from the previous digression or topic creates an inherently contrastive position with the next segment. Examples from the Red Hen dataset are given in (122) to (124).

(122) *Against another individual such as her date and then, of course, it begs the question if this was the first date, you wonder what the second date would look like. <u>But, anyway</u>, in terms of the criminal mischief charge, yes.*[158]

(123) *I think Easter is the high Christian holiday. <u>But anyway</u>, I think -- so here are two things that happen. One is ornaments on the tree that, you know, my parents and my wife's parents both handed down ornaments.* [159]

[158] Red Hen: 2017-12-28_1800_US_HLN_On_the_Story_With_Erica_Hill,1116

[159] Red Hen: 2017-12-24_2200_US_FOX-News_The_Five_Christmas_Special,498

(124) A: *It's -- I thought it was a great movie.*
B: *Where is Joey on the Porgs?*
A: *Joey is fine with the Porgs. It's not like it's Jaja Binx. That's not a spoiler. <u>But anyway</u>, it's great movie making and Mark Hamill, holy moly, incredible.*[160]

The findings for the analysis of the enactment of the [*but anyway(s)*] construction are given in Figure 5.28 and Table 5.18. As the bar chart in Figure 5.28 shows, like *so anyway(s), but anyway(s)* is dominantly expressed in the hands and head; however, there is marginally less manual gesture (74% for *but anyway(s)* rather than 81% for *so anyway(s)*). The greatest difference lies in the proportion of enactments that feature head movement. Although head is the second-most involved articulator for both phrases, there's an incidence of 47% involvement of the head for *but anyway(s)*, while a full 81% incidence for *so anyway(s)*.

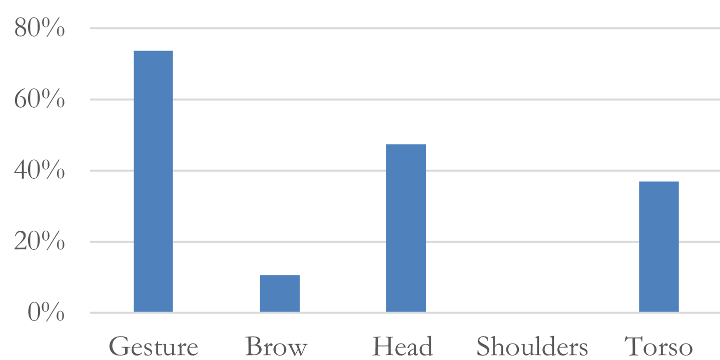


Figure 5.28. Distribution by articulator: *but anyway(s)*

Looking more closely at the form of the gesture (Table 5.18), we see that the dominant hand form for *but anyway(s)*is the palm-down open-hand (PUOH) gesture. Recall that, by contrast, *so anyway(s)*featured dominantly palm-up open-hand (PDOH) gestures, as featured in Figure 5.26.

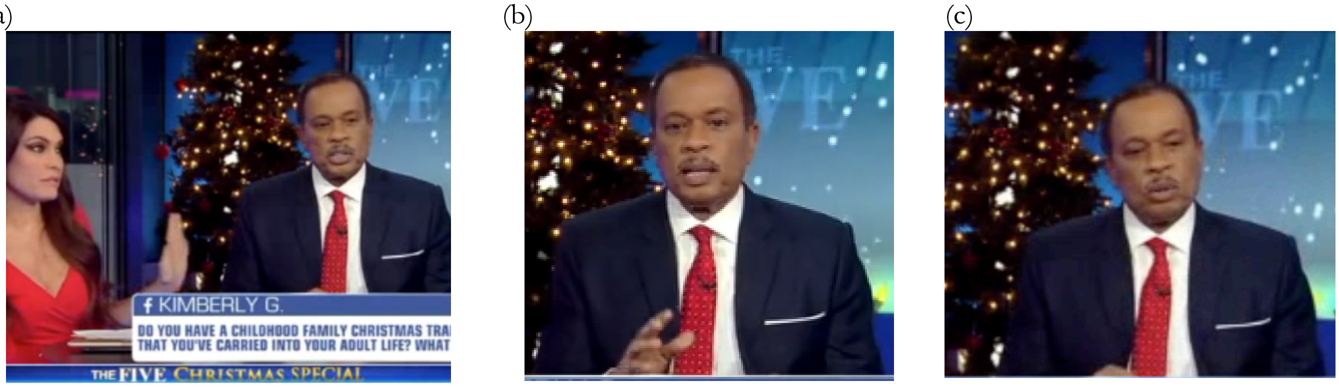
[160] Red Hen: 2017-12-18_1100_US_MSNBC_Morning_Joe,167

Table 5.18. Distribution of articulator forms: *but anyway(s)*

Articulator	Tokens (n=19)	Articulator form
Gesture	14	7 PDOH, 3 PUOH, 2 PLTC, 1 palm-away, 1 PL away-body
Brow	2	2 brow raises
Head	10	6 nods, 4 turns
Shoulders	0	n/a
Torso	7	5 shifts back/up, 1 turn, 1 lean

A closer analysis of the execution of the seven PDOH and the singular palm-away gesture reveals a type of gradation in the gestural form. That is, the angle of the wrist and forearm is rarely absolutely horizontal. For example, in Figure 5.29 and Figure 5.30, the difference between the gesture forms in the central (b) image is in the angle of the forearm and/or wrist.

(a) (b) (c)



(a)	(b)	(c)
S: *I think Easter is the high Christian holiday.*	*<u>But anyway</u>,*	*I think -- so there are two things that happen*
G: hands together playing with pen (obscured by chyron)	palm-down	right hand lowers to position in (a)

Figure 5.29. *but anyway,* palm-down open-hand

S: *There's enough violence in the world without me flashing.* | *but anyway* | *I was standing…*

G: bimanual PUOH | palm-away, body shift left and back | posture and hands return to neutral

Figure 5.30. *but anyway,* one-handed holding away gesture (palm-vertical)

Thus, for *but anyway(s)*, I suggest that the PDOH and palm-away forms are schematically related. This seems to hold across both one-handed and two-handed gestures. While the previous two figures featured one-handed gestures for *but anyway(s),* three enactments of two-handed gestures are given in Figure 5.31.[161] In these examples, the palm-away gesture occurs in the prototypical form of the holding away recurrent form described in Chapter 2. Figure 5.31(a) and (b) feature the palm-vertical variant that is typical of the holding away gesture, while in (c) the holding away gesture is executed in a more relaxed execution, which more closely resembles a palm-down gesture. I discuss this subtlety in execution as a grammaticalization of gesture forms in Chapter 6.

[161] Note: these three examples stem from videos outside of the date range of the study. I include them here to illustrate the variations of the holding away gesture with *but anyway*.

(a) (b) (c)

S: *but anyway*
G: palm-vertical holding away gesture

S: *but anyways*
G: palm-vertical holding away gesture

S: *but anyway*
G: palm-down holding away gesture

Figure 5.31. *but anyway,* bimanual vertical palm-away and palm-down gestures

In addition to these holding away gestures, it is worth noting that two of the PDOH gestures and one instance of the palm-lateral gesture for *but anyway(s)* featured an inflection reminiscent of the *quit* gesture profile presented in Chapter 3, namely a one- or two-handed PDOH gesture moving in a lateral, outwards direction. An example of *but anyway(s)* with this gesture is given in Figure 5.32.

S: *I'm not sure.*

G: very brief palm down form with back and forth motion, head shake, nod on sure

but anyway (extends over the 'quit' like gesture stroke in b, c, and d)

G: *PDOH, lateral, moving across body from right to left*

Figure 5.32. *but anyway,* outward-moving PDOH

Lastly, the head movements for *but anyway(s)* were present in over half of the cases (10/19) and were dominated by a singular nod downwards and return to neutral. These occurred both in utterances with gesture and in those without. The distribution of head movement differed from *so anyway(s)* in that *but anyway(s)* lacks head shakes.

Regarding the issue of whether *but anyway(s)* is uttered as one intonational chunk or two, none of the instances with co-speech behaviour featured a discernible, measurable pause (measured using ELAN time stamps) between *but* and *anyway*. In the full data set collected (i.e. including utterances with no enactment) there was one instance that featured a pause greater than 300 msec between *but* and *anyway*.

To summarize, *so anyway(s)* and *but anyway(s)* both featured high degrees of gestural enactment, with the *so anyway(s)* collocation featuring more head movement than *but anyway*. The gestures for these two closely related return expressions differed in their dominant forms. *So anyway(s)* featured PUOH and palm-lateral towards centre (often container forms), while *but anyway(s)* was dominated by variations of the holding away gesture. The head movements also varied in that *but anyway(s)* did not feature any shakes.

The findings suggest that the seemingly synonymous *so anyway(s)* and *but anyway(s)* actually have different profiles. The gesture forms that are conventionalized for each utterance would suggest that there is a difference in concessive force between the two collocations. The conjunction *but* is inherently contrastive. On the other hand, *so* expresses continuation or consequence (e.g. *therefore*). With this in mind, I propose that the gestural and head movement profiles for *but anyway(s)* iconically represent the much greater degree of contrastive force implied by *but* as opposed to *so*. This would explain the dominance of holding away gestures and singular head nods for *but anyway*. *So anyway(s)* seems to signal a return to a previous topic without necessarily creating an overt contrast between the previous and upcoming discourse segments. It marks continuation of the main thread of the discourse as a whole (i.e. *so anyway(s)* can also begin a new thread), rather than as an explicit marker of contrast.

The previous four case studies have featured discourse markers that contain the quantifier *any*, as in the adverbs *anyway(s)* and *anyhow,* and the phrasal connective *at any rate*. By contrast, the next three discourse markers are concessive markers or phrases without *any*: *however, nevertheless,* and *that being said.*

However, I

Like *but anyway(s), however* has been considered as an indicator of CONTRAST as well as a discourse marker. Nevertheless, the literature also acknowledges the challenges in determining "whether *however* is used in either one or the other of the two functions" (Lenk 1998: 108). I suggest that when *however* is used as a return expression at the left periphery, it fulfills both functions, i.e. it marks contrast and simultaneously serves a discourse navigation function. As explained in §5.2.1, I limit my exploration of *however* to its concessive function. Preliminary Red Hen searches made very clear the exceptionally heterogeneous nature of *however*. In order to reduce the data to contexts in which *however* predominantly functioned as a return expression, I searched only for the collocation of *however* with the first person singular subject pronoun *I,* as in [however, I], given that I is the most frequent collocating pronoun for *however*. A search of COCA_{sp} returned 2,028 search returns for *however* followed by a subject pronoun, or [however, PRO]. The [however, PRO] construction occurs most frequently with 1SG compared to any other person pronouns. The distribution of subject pronouns in [however, PRO] is given in Table 5.19. Together, *I* and *we* account for just under 50% of the search returns.

Table 5.19. Pronoun frequency in [however, PRO] in COCA$_{sp}$

Pronoun	Frequency in COCA$_{sp}$
I	641
it	371
we	359
they	215
he	192
you	185
she	65
Total	2028

An examination of the verbs that follow the pronouns shows that [however, PRO VERB] displays a preference for the inflected collocations *I think* and *you know*. In Table 5.20, I show the eight most frequent collocate verbs in COCA for [however, PRO VERB] with the specific pronouns *I, we, you,* and *they* (there were too few search returns with 3SG to merit inclusion). Note that the verb *know* features in the top three verbs for all pronoun forms and *believe* in three of the four; however, these represent only a small portion of the data, since *however, I think* and *however, you know* account for the vast majority of the utterances. As mentioned above, *you know* is an inflected verb collocation that has grammaticalized into a stanced discourse marker and *think* is a verb that overwhelmingly prefers first-person subjects, whether singular or plural. The point here is simply to illuminate the behavioural properties of the [however, I] with regard to other pronoun and pronoun-verb collocates (e.g. [however PRO] and [however PRO VERB].

Table 5.20. Five most common verb collocates for [however , PRO]

however , I		*however, we*		*however , you*		*however , they*	
think	**150**	*know*	8	***know***	**59**	*say*	8
believe	15	*believe*	8	*see*	4	*know*	3
know	8	*think*	7	*want*	4	*believe*	3
want	7	*need*	7	*say*	3	*want*	2
understand	5	*want*	5	*choose*	3	*seem*	2
feel	4	*feel*	5	*look*	2	*worry*	1
agree	4	*owe*	2	*need*	2	*work*	1
Total	193		42		77		20

Returning to Red Hen examples, frequent collocations for [however, I] from Red Hen are shown in Table 5.21. This adjustment to the target utterance forced the return of more subjective examples, which is evident in the large number of collocating stance expressions in the surrounding context.

Table 5.21. Collocating stanced expressions with *however, I*

Collocate function	Tokens (n=25)	Examples
Discourse marker	1	*I mean*
Evaluation	4	*I like/don't like, I would like to, I am in favour of*
Verbs of cognition	5	*I think, I understand, I think it's important to, I know*
Verbs of communication	6	*I am happy to say, I can tell you, I should say, I don't say this very often, I have to say this*

Full utterance contexts for [however, I] are given in (125) and (126).

(125) *So wonderful. However, I can tell you – I can tell you that Josh is not the only person on our show tonight with musical theater chops.*[162]

(126) *And just that awkwardness. And I enjoy that. However, I don't like watching it. So, my wife makes fun of me.* [163]

162 Red Hen: 2015-04-09_0737_US_KCBS_Late_Late_Show_with_Craig_Ferguson,1486

163 Red Hen: 2015-09-05_0737_US_KNBC_Late_Night_with_Seth_Myers,2689

The kinesic profile for this highly subjective utterance featured both gesture and upper body movement in 83% of the enactments, with only 1 instance of gesture only and 3 instances of upper body only. (See the body partitioning results shown at the beginning of this section in (§5.4.1.) As Figure 5.33 shows, gestures were aligned with the utterance 88% of the time, the second highest rate among the seven return expressions. The percentage involvement of head movement was 58%, in the middle range of head movement percentages across all target utterances.

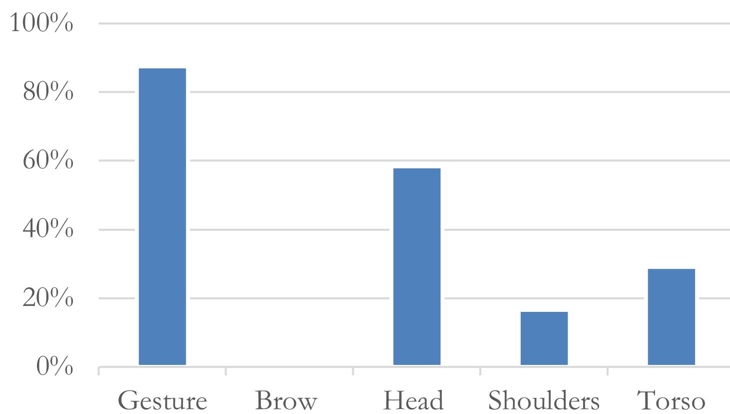


Figure 5.33. Distribution by articulator: *however, I*

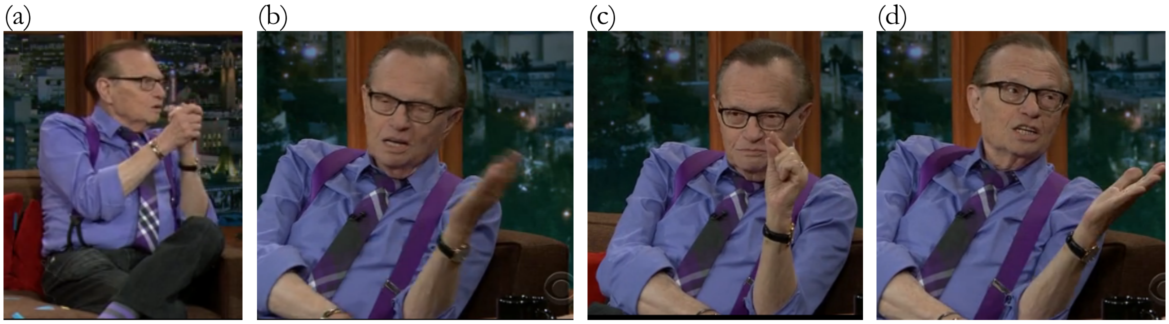
Table 5.22 shows the distribution of articulator for [however, I]. For manual gesture, the dominant form was the point, while the dominant upper body movement consisted of head tilts and turns. With nine points in the 24 total utterances, the point certainly dominated. An example is given in Figure 5.34. The second most common gesture form was the pursed hand, given in Figure 5.35. Similar to the point, the pursed hand is a precision hand form created by precise shaping of one's fingers. These forms are markedly different than the majority of hand forms that occur for return expressions, which are open-hand forms such as the palm-down open-hand, PUOH, and palm-lateral towards centre, all of which feature flat hand shapes.

Table 5.22. Distribution of articulator forms: *however, I*

Articulator	Tokens (n=24)	Articulator form
Gesture	21	9 point, 4 pursed hand, 3 PD, 2 palm-away, 2 PUOH, 1 other
Brow	0	n/a
Head	14	6 tilts, 5 turns, 2 nods, 1 shake
Shoulders	4	4 dips
Torso	7	5 leans, 2 turns

(a)	(b)	(c)
S: *(pause)*	*<u>However, I</u> have to say Anthony Hopkins*	*was and is the sexiest guy on this set.*
G: n/a	torso turn, head turn, raised hand in point	pursed finger

Figure 5.34. *however, I,* pointing gesture



(a)	(b)	(c)	(d)
S: I never was into alcohol.	*Pot, I loved.*	*<u>However,</u>*	*I got a heart attack*
G: clasped hands	PUOH	pursed hand	PUOH

Figure 5.35. *however, I,* pursed hand

Nevertheless

Like *however, nevertheless* is a linking adverbial that is more common in written genres than in speech. Of the ~17,000 search returns for *nevertheless* in all genres of COCA, only 8% were in the spoken genre, with 50% in the academic section of the corpus. Despite this low relative frequency in COCA, with 184 tokens within the restricted program range and required time frame in Red Hen (2012-2015), there were enough to extract 25 tokens that matched the criteria for usage and visibility. Representative uses are given in (127)-(129).

(127) *Only Taylor can create static that's magic on iTunes. <u>Nevertheless</u>, they fixed the problem.*[164]

(128) *She realizes it was a mistake right away and she crossed the line and feels terrible, but he was giving her that attention that she needed at that moment, <u>but nevertheless</u>, not the right thing to do, and he flips out.* [165]

(129) *Last night their opponent had no way to compete monetarily with that. <u>And nevertheless</u>, those little Davids fighting that Goliath. They whomped Chevron in every single race.*[166]

In seven of the 25 utterances, the contrastive conjunction *but* preceded *nevertheless,* as in (128). *And* preceded it in 2 more utterances, as in (129). These variations did not appear to be strongly correlated with differences in the enacted forms.

A preliminary look at the involvement of articulators shows that it mirrors that of [however, I] in the role of each articulator. That is, over 90% of the enactments are gestured, while 60% are enacted by head movement, similar frequencies as for [however, I]. Also similar was the fact that torso and shoulders play a role in only a few instances.

---

164 Red Hen: 2014-10-22_1400_US_KNBC_Today_Show,3125

165 Red Hen: 2015-01-21_0000_US_KNBC_The_Ellen_DeGeneres_Show,2083

166 Red Hen: 2014-11-06_0200_US_MSNBC_The_Rachel_Maddow_Show,258

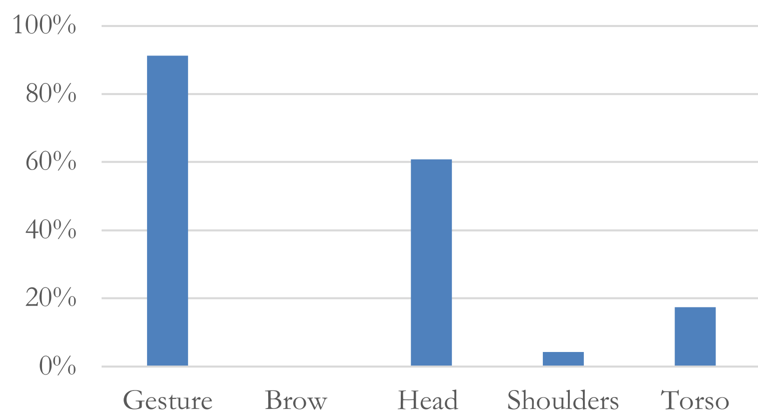


Figure 5.36. Distribution by articulator: *nevertheless*

Table 5.23. Distribution of articulator forms: *nevertheless*

Articulator	Tokens (n=23)	Articulator form
Gesture	21	7 PUOH, 5 points, 9 other
Brow	0	n/a
Head	14	5 nods, 4 turns, 2 tilts, 2 shakes, 1 bob
Shoulders	1	1 dip
Torso	4	2 turns, 1 tilt, 1 lean

The distribution of the articulator forms given in Table 5.23 shows that PUOH was the most frequent form (7/21) for [nevertheless]. In 6 of the 7 PUOH gestures for [nevertheless], the gesture was performed with both hands, as in the prototypical shrug position. Points also featured in 24% of the cases (5/21). Not only was this fewer points, proportionally, than for [however, I], but the nature of the points differed as well. While [however, I] featured points in the vertical plane (i.e. raised in front of the speaker's face) as depicted in Figure 5.34, above, the points for [nevertheless] were more varied. Three examples are given in Figure 5.37. In Figure 5.37(a), I show a bimanual gesture with weakly articulated finger points in both hands. Figure 5.37(b) shows a flat hand directed towards the audience, and (c) shows the speaker stretching out an arm towards the audience on *nevertheless*. The latter two gestures have a deictic element to them as well, as they point to the interlocutor/viewer.

The reference to the audience in (b) ends the preceding digression by pointing to the audience as if to bring them back into the conversation, while in (c) the palm down gesture that points to the audience seems to express the speaker's assurance that the problem has been fixed (i.e. that the viewers should not worry since the problem with iTunes has been fixed). Regardless of the specifics of each utterance, the points are notably different in form and function from the [however, I] points.

(a) (b) (c)

(a) S: *Either she didn't believe it was the Today Show or she was a teenager awaken at 6:00 am.* *Nevertheless, we got Mariah on the phone and by popular request…*

G: bimanual point gesture with brow raise

(b) S: *People are likening it in some bases to the Watergate investigation, the depth of it.* *Nevertheless what, it was it a total of 11 hours?*

G: palm-lateral towards centre, deictic gesture towards listener

(c) S: *Only Taylor can create static that's magic on iTunes.* *Nevertheless, they fixed the problem.*

G: arm stretched towards audience, flat palm, step backwards

Figure 5.37. *nevertheless,* range of deictic gesture forms

The remainder of the gestures for *nevertheless* were one-off executions that defied any efforts to categorize them. There were four gestures with various hand forms and orientations that moved in a downward stroke once or twice, reminiscent of beat gestures[167]. These featured a flat palm facing downwards or the fingertips of a PLTC container gesture beating

[167] Beat gestures keep the rhythm of language and can have a range of hand forms (McNeill 1992).

once or twice on the table in front of the speaker. These gestures were aligned with *nevertheless* and featured on average only one or two beats. They, therefore, seem to underscore the emphatic nature of *nevertheless*. Speakers did not maintain their prosodic rhythm over a longer segment, as beat gestures often do.

To summarize, [nevertheless] has a distinct profile, especially when considered in comparison to [however, I]. While [however, I] featured points and pursed hand shapes, [nevertheless] was dominated by PUOH forms and a more diffuse representation of point gestures. Similarly, while [however, I] was dominated by tilt and turn head gestures, [nevertheless] was dominated by turn and nod gestures. These two return markers are listed as synonyms in the Oxford English Dictionary, both having the meaning *despite that* or *notwithstanding,* which qualify the preceding sentence or clause as a whole. The *Longman Grammar of Spoken and Written English* states that *nevertheless* focuses on concession, whereas *however* can combine concessive and contrast functions (Biber et al. 1999: 878-879). While the data presented here are not robust enough and the linguistic contexts too varied to make grand claims regarding this functional difference, the clear distinctions in the multimodal profile of [however, I] and [nevertheless] suggest they are independent constructions, rather than synonyms.

I present one more return marker, *that being said,* before summarizing the findings in §5.4.2.

That being said

That being said ([TBS]) is a return marker that is equated with *having said that, even so,* and the return construction presented in the previous section, [nevertheless]. Examples from Red Hen are provided in (130)-(132).

(130) *This is not 'you have to give us 95% of your money'. This is, I think, a reasonable way of looking at it. That being said the fear is we return to a different time in New York when it was less orderly.*[168]

(131) *It was hard for me to say goodbye. But that being said, the show, it was time for it to come to its natural end.*[169]

(132) *It is time to say it: the NFL is a morally, indefensible organization. Of course, that being said, man am I glad training camps are back.*[170]

[TBS] is most frequently followed by abstract nominals (rather than pronouns), as in *the fear, the show,* and *the training camp* in these examples.

To introduce the kinesic profile for [TBS], I remind the reader that, as introduced in §5.4.1, the degree of enactment for this return marker was 80% (20/25) (compared to [however, I]: 96% and [nevertheless]: 92%). The body partitioning was 60% (12/20) gesture and upper body, with five featuring gesture only and three featuring upper body only. The distribution of involvement by articulator is shown in Figure 5.38. The relative involvement of gesture is less than for [however, I] and [nevertheless], but head involvement (60%) is very similar. In movements of the shoulders, [TBS] more closely resembles [however, I] while the involvement of the torso resembles [nevertheless]. Finally, it is the only one of these three return markers to feature brow movement.

[168] Red Hen: 2014-02-14_0400_US_ComedyCentral_Daily_Show,1520

[169] Red Hen: 2015-02-20_0837_US_KCBS_Late_Late_Show_with_Craig_Ferguson,2030

[170] Red Hen: 2014-08-14_0300_US_ComedyCentral_Daily_Show,915

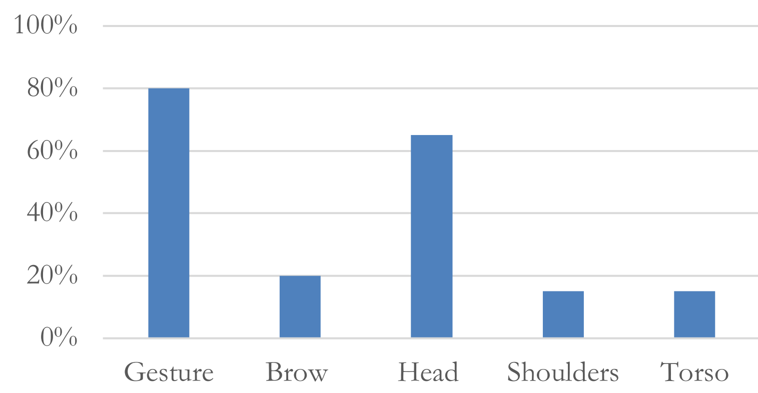


Figure 5.38. Distribution by articulator: *that being said*

Table 5.24 presents the profiles within each articulator for [TBS]. Of the 16 enactments with gesture, over half featured a palm-down or holding away gesture. The other gestures were a mix of PUOH, points and pursed forms. The frequent head forms were nods and tilts.

Table 5.24. Distribution of articulator forms: *that being said*

Articulator	Tokens (n=20)	Articulator form
Gesture	16	6 holding away, 3 PUOH, 2 PD, 3 points, 1 pursed, 2 other
Brow	3	3 brow raises
Head	13	5 tilts, 5 nods, 3 turns
Shoulders	3	3 shrugs
Torso	3	2 leans, 1 turn

In Figure 5.39, I present the prototypical gesture form for [TBS], the palm-down or holding away gesture. This one includes a head turn and brow raise.

(a) (b) (c)

(a)	(b)	(c)
S: … *just makes it seem kinda like a cheap shot, so that's not my style.*	*That being said…*	*(pause)*
G: adjusts tie, some torso movement, head turn left, gaze down	G: bimanual holding away gesture, turns head right, brows up	G: returns hands to lap, adjusts pants

Figure 5.39. *that being said*, holding away gesture with brow raise and head turn

In sum, *that being said* presented with a dominant enacted profile that resembles that of *but anyway(s)* (see Figure 5.31), namely with a holding away gesture. In comparison with the previous two return markers, *however* and *nevertheless*, it was not as frequently gestured, but was more frequently accompanied by head movement and brow raises. Indeed, it is in comparing the results across several return expressions that the profiles stand out. In the final section in this analysis of the enactment of return expressions, I provide summary tables and discuss differences in the profiles that have emerged in these microstudies.

5.4.2 Summary and discussion: Returns

In the series of microstudies presented in this section, I have profiled the linguistic characteristics and kinesic enactments of seven return expressions in North American English. By annotating the dominant gesture forms, head movements, shoulder, brow, and torso movements, I have attempted to capture the most common ways that *anyhow, at any rate, so* and *but anyway(s), however, I, nevertheless,* and *that being said* are expressed in the body at the time of utterance. By including characteristics of the spoken usage patterns, I have endeavored to provide a picture of both linguistic and kinesic patterns that emerge when these expressions are used in interactional discourse to mark a juncture in the narrative flow.

To summarize the results here, I present findings for the distribution by articulator for all expressions. Figure 5.40 presents the *any* expressions, while Figure 5.41 shows those without *any*. The results show a high degree of variation across expressions. For example, *anyhow* and *at any rate* are the only expressions with more head movement than gesture, while *so anyway(s)* shows equal use of both articulators. These differ from *but anyway(s)* and the remaining three expressions shown in Figure 5.41, which feature gesture use over 70% of the time and head movement in 45-65% of the cases.

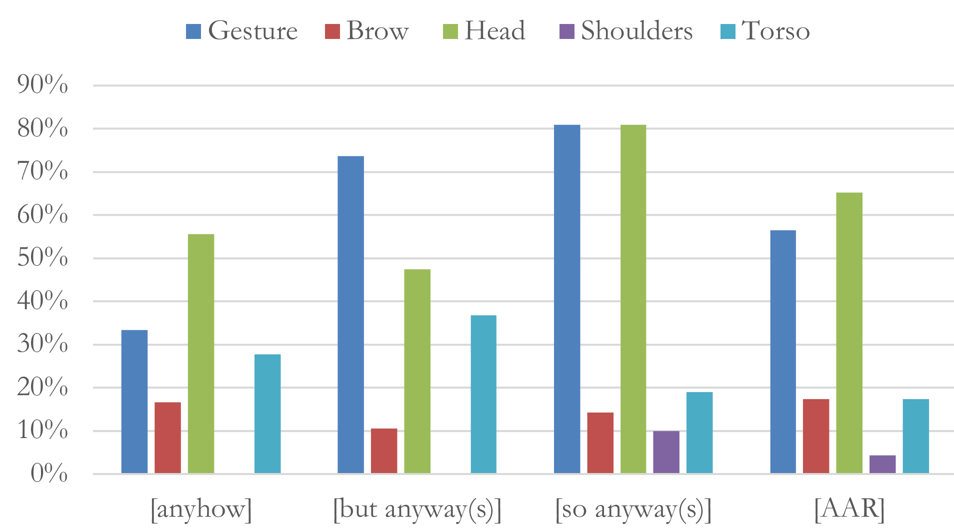


Figure 5.40. Distribution by articulator for all return expressions with *any*

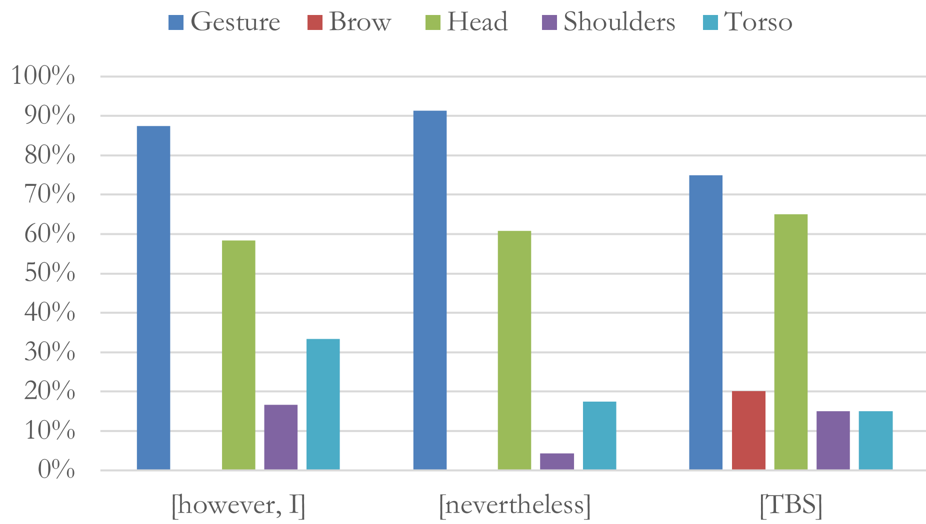


Figure 5.41. Distribution by articulator for [however, I], [nevertheless], and [TBS]

These figures show only which articulators were involved. In the microstudies presented throughout this section, I also investigated the patterns within the form for each articulator. To summarize, in Table 5.25 and Table 5.26, I provide a summary of the dominant forms for gesture and head movement, as well what percentage of the tokens this form accounts for.

Table 5.25. Dominant gesture form by return expression

	[anyhow]	[but anyway(s)]	[so anyway(s)]	[AAR]	[however, I]	[nevertheless]	[TBS]
Gestured*	33%	74%	81%	57%	88%	91%	75%
Dominant form	n/a	PDOH	PUOH	PDOH	POINT	PUOH	holding away
Dominant form freq.**	n/a	50%	47%	46%	43%	29%	53%

* Gestured = gestured/total enacted
** Dominant form freq. = frequency of dominant gesture/total gesture

Table 5.26. Dominant head movement by return expression

	[anyhow]	[but anyway(s)]	[so anyway(s)]	[AAR]	[however, I]	[nevertheless]	[TBS]
Head involvement*	56%	47%	81%	65%	58%	61%	65%
Dominant form	tilt	turn	nod/turn	nod	tilt	nod	tilt/nod
Dominant form freq.**	30% 3/10	44% 4/9	29% 5/17	40% 6/15	43% 6/14	36% 5/14	38% 5/13

* Head = head movement/total enacted
** Dominant form freq. = frequency of dominant head movement/total head movement

Given the number of tokens involved, all the findings here need to be taken as preliminary. What is needed is a much larger case study of each return expression that would allow for statistical analysis across the expressions to determine if the trends presented here withstand statistical significance testing. That being said, I believe the findings presented in this section identify how these expressions differ from one another and offer important avenues for clarifying the relationship between contrast and concession. The gesture forms for *but anyway(s), at any rate,* and *that being said,* all have a concessive, negating hand form associated with them: the palm-down open-hand. It is executed either with a palm facing fully downwards or a palm facing outwards away from the speaker in a holding away gesture. Their gesture form distinguishes these expressions from *anyhow*, which has no dominant gesture form, and *so anyway(s)* and *nevertheless,* both dominated by the PUOH form. The trends for the near-neighbours *but anyway(s)* and *so anyway(s)* suggest that concessive force is enacted in gesture. The more forceful of the pair – *but anyway(s)* – features more palm-down gestures, while the more continuative *so anyway(s)* is dominated by PUOH. Finally, the dominance of points for *however, I* suggests a unique profile for this highly subjective collocation.

5.5 Excursions and returns as stance-marking, multimodal constructions

In this chapter, I have examined the linguistic and multimodal means with which speakers navigate discourse in spoken interaction. I have examined parenthetical asides, in which speakers make excursions to express a stance about something they have just said and the mechanisms by which they return to the main discourse thread. By collecting and analyzing data from the Red Hen archive for five excursion expressions and seven return expressions, I developed a multimodal profile for each expression. Each profile consisted of the degree of enactment (i.e., how frequently it co-occurred with meaningful articulator movement), the body partitioning profile (i.e., the involvement of groups of articulators from the hands to upper body), and the specific forms of each articulator movement. The findings raise a number of important issues, which I will address in the remaining sections of this chapter.

Given their functions as navigating devices, the excursion and return expressions treated in this chapter are, by definition, highly stanced. In addition to the subjectivity involved in simply directing the flow of discourse, the excursions explicitly express a stance about the immediate topic of the discourse. Similarly, in addition to simply shifting away from a previous topic, the return markers inherently mark a stance of contrast or concession towards the previous discourse. The findings presented in this chapter offer some insight into the nature of stance, the nature of iconicity in co-speech behaviour, and the nature of multimodal constructions. I address each of these in turn.

5.5.1 The nature of stance: Embodied, stacked, and idiomatic

Seminal research on stance in the previous decade was based in corpus-based approaches exploring linguistic devices that mark stance (Precht 2000; Du Bois 2007). More recently, manual gesture and other meaningful body movements have been included in investigations of how speakers mark stance in face-to-face interaction. Stance is now studied from a range of perspectives and includes multiple modalities ranging from spoken language to multimodal and embodied communication (including both co-speech behaviour and stance-marking in signed languages). Whether studied from a monomodal (i.e. speech only) or multimodal perspective, it is clear that stance is frequently expressed in composite utterances in which multiple expressions of stance coincide. That is, stance tends to 'stack'. "Stance stacking" is a term that Dancygier (2012) introduced to capture how multiple expressions of stance can create a constructional cohort. For example, stance verbs are frequently combined with negation to yield new 'stacked' fixed units and discourse markers often occur together (e.g. *I mean, you know* uttered together). In the data presented in this chapter, several of the excursion and return markers are themselves stacked, e.g. *but anyway(s), so anyway(s),* and *although, you know.* In multimodal enactments, stance stacking is readily seen in the shrug, a composite utterance that frequently involves a fixed combination of PUOH gesture, shoulder shrug, mouth shrug, and head tilt (Gibbs 2005; Debras 2017). The stacking is layered across modalities when an utterance accompanies a shrug form, for example a shrug with the utterance *I dunno.*

Throughout this chapter, I have investigated a set of fixed phrases that function as stanced discourse navigation markers in spoken English and investigated their multimodal enactments. I propose that the profiles that emerge are examples of stance-stacking in both senses described here, that is, the co-articulation across multiple modalities and across multiple articulators. Take, for example, the compound excursion utterances *so anyway(s)* and *but anyway(s).* I originally chose to use these compound utterances as place holders for *anyway(s)* due to the operational challenges of searching in Red Hen for *anyway(s)* on its own. Recall that *anyway(s)* is heterogeneous to the point that it made the search returns very noisy. I therefore delimited the search to *but anyway(s)* and *so anyway(s)* to force the return function of *anyway(s).*

This search strategy provided near-neighbour constructions that nevertheless differ in their concessive force. The unique profiles for each utterance suggest that fine semantic differences such as those which hold between these two utterances are encoded in the co-speech behaviour.

The highly stanced fixed expressions discussed in this chapter are each associated with discernible embodied profiles. The individual case studies have allowed for the development of individual profiles for each excursion and return marker. Within each set of examples for each marker, there are some that seem to more strongly stanced. One can consider stance as a continuum that ranges from weak to strong, in which case each expression can be ascribed positions as more or less stanced than others. In this chapter, I used the body partitioning variable to operationalize the involvement of the upper body versus manual gesture. That is, I captured whether an utterance was expressed only in the hands, only in the upper body, or as a composite utterance with synchronous expression in both the hands and upper body. To test the hypothesis that more objective or neutral material is expressed in the hands, whereas more subjective material is expressed in the upper body, I compared the relative degree of subjectivity of the expressions with their body partitioning profiles. While for some utterances the findings for body partitioning seem to align with the inherent subjectivity (or lack thereof) of the utterance, for others, the findings were not as clear cut. For example, in §5.4.1, I showed that in the embodied profile for [however, I], there were no enactments featuring only the hands; rather, [however, I] is expressed through both manual gestures and upper body simultaneously (85%), with only 15% expressed exclusively in the upper body. By comparison, *but anyway(s)* features 21% gesture only, 53% gesture and upper body, and 26% upper body only. The question remains, though, as to what this tells us about the difference in subjective force between the highly subjective [however, I] and similar expressions. For example, [however, I] and [nevertheless] vary by their pronoun subject, which was limited to first-person subject pronoun for [however, I] but was frequently third person nominals for [nevertheless]. Their kinesic profiles were similar in their degrees of enactment and the involvement of unique articulators, but the specific articulated forms varied greatly, with point gestures dominating for [however, I] as compared to PUOH forms and a range of low-frequency forms featuring for [nevertheless], to focus only on gesture for a moment.

Whether a speaker recruits the hands or the upper body as they express themselves depends on a wide range of linguistic, conceptual, and interactional factors, to mention only a few broad areas. Furthermore, body partitioning conflates regions, i.e. the upper body collapses shoulder and head movement together, whereas perhaps these distinct bodily articulators contribute differentially to stance-marking. In addition to conflating articulators, the body portioning profile does not capture the form that the articulations take or how conventionalized these forms are across tokens of the utterance.

To close this discussion of the role of the body in the marking of stance, I return to the measure I call the degree of enactment. In this chapter, I have presented the kinesic profile for the non-restrictive relative clauses *which is true* and *which is fine*. For comparison, I also captured the degree of enactment for 20 examples each of other frequently used and similarly semi-evaluative non-restrictive clauses, namely, *which is good, which is great,* and *which is nice*. These are plotted in Figure 5.42 in order of estimated degree of subjectivity, with more neutral expressions on the left and more subjective expressions on the right. *Which is true* is highly subjective, given its strong epistemic force from the adjective *true*. Of the remaining adjectives, *good* and *nice* are more neutral, *great* has an increased evaluative force (Zyp 2018), and *fine* a higher subjective force in its use in the non-restrictive relative clause *which is fine,* as discussed in its profile earlier in the chapter. The degree of enactment increases along this cline, which suggests that the recruitment of the body overall may be an important factor in marking stance as well.

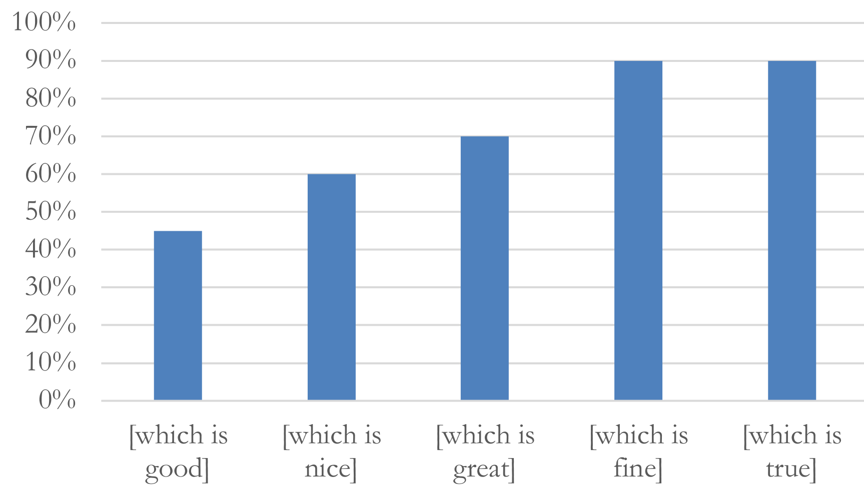


Figure 5.42. Degree of enactment for non-restrictive relative clauses

For the multimodal profiles presented in this chapter, I included the degree of enactment, the body portioning profile, and an analysis of the conventionalized form that emerged from the annotations. I believe that each of these provides important information regarding the kinesic profile of these utterances and provides fruitful grounds for comparison across them. What remains to be seen is the degree to which each of these elements of the kinesic profile specifically mark stance in what are all strongly stanced constructions.

What the findings presented in this chapter do show are that these stance-marking constructions are heavily layered across the spoken and embodied signals and that they present complex construals through conventionalized compound enactments. I have attempted to explore the specific importance of sets of articulators to the expression of stance, but I believe many questions remain and the relevance of body partitioning to these constructions requires a great deal more study. The constructions treated here recruit different articulators to organize and express stance in the body, thus I maintain that the body partitioning profile remains an important coarse-grained tool that captures some critical elements of a multimodal profile.

5.5.2 Concluding remarks

Excursion and return markers can be thought of as sign posts in a discourse. They mark points of departure from a current path and points of return to that path or the embarkment onto a new path. The multimodal profiles that emerged for the utterances investigated in this chapter demonstrate that the body is an active participant in marking these turning points in the discourse. This brings us back to a theme that recurs throughout this dissertation, that of iconicity in co-speech behaviour. The discourse navigating utterances investigated here featured degrees of enactment from 55% (*although, you know*) to 96% (*however, I)*. This means that on average, 84% of the discourse navigation markers that were part of these microstudies were marked in the body in some way. The profiles show exactly how these were marked and include a high degree of activity through shifts in the head, shoulders, and torso. In comparison to the rich affordances of the hands to aptly express propositional CONTRAST through both form and placement of a manual gesture, for kinesic reasons, this range is not available to upper body articulators. Rather, in the upper body, articulators generally move in a smaller range of ways (e.g. the head can nod, tilt or shake, shoulders can move up or down). As the data presented in this chapter suggests, speakers are adept at enacting discourse navigation. We must simply expect that it looks different given the kinesic movements available to the upper body. Shifts in direction can enact – and iconically represent – the shifts in discourse that are marked by the spoken utterance.

Of course, in many expressions, manual gesture was highly conventionalized as well. In these cases, the gesture tended to convey the semantic force of the target utterance. For example, in the case of *which is fine*, the high degree of palm-down and holding away gestures imbued this utterance with a high degree of negative stance. By comparison, the more neutral *which, by the way,* which often introduces further detail rather than simply making a (negative) evaluation about the previous utterance, had a more varied gesture profile. It is equally enacted, but less conventionalized.

In the case studies presented here, I have attempted to bring order to some of the messiness inherent in taking a wide-angle, multimodal, and spontaneous discourse lens to language phenomena. I have included in my analysis both linguistic and multimodal variables, and considered the potential divisions of labour taken on by different kinesic articulators. At the very least, the multimodal profiles that emerge from the excursions and returns investigated in this chapter put an end to the notion that these utterances are synonyms, or even near synonyms. Comparing the embodied profiles for [but anyway(s)], [however, I], and [nevertheless], as well as [WIT], [WIF], and other such expressions, make clear that each of these makes specific distinctions that are motivated by discourse, pragmatic, and/or semantic forces.

Taken together, these studies provide further evidence that, as argued throughout this dissertation, the inclusion of multimodal data in linguistic analysis directs us towards an understanding of language as highly idiomatic and conventionalized at the same time. They also raise important issues around conventionalization and variation in multimodal expressions. What does a conventionalized form look like? Is the simple involvement of articulators, rather than their precise forms, adequate for describing a multimodal construction? While answers to these larger questions will require more research from a variety of approaches, I believe the data and findings presented in this chapter provide motivation for considering the role that multimodal constructions should play in language description, pedagogy, and even theoretical grammars. In the final chapter, I turn to some of these larger issues.

Chapter 6 | Discussion and conclusion

6.1 Introduction

Cognitive and functional approaches to language champion meaning and usage as the true starting points for linguistic analysis. Specific constructional pairings of form and meaning are construed as the basic unit of linguistic analysis. In this view, language itself is seen as a structured inventory of conventional form-meaning pairings (constructions) (Langacker 1991b: 15) that can vary in size and nature, extending from single morphemes and lexemes to grammatical structures and fixed and idiomatic phrasal bundles. Constructions can also include forms and functions that have typically been left out in the cold, theoretically speaking, and treated as lying beyond theoretically interesting linguistic phenomena. However, despite the theoretical openness of functional and cognitive approaches to widening the linguistic lens, the progressive inclusion of a broader swath of linguistic material in the description of language structures and patterns has been an uphill climb. Indeed, the past few years have seen several pleas for increased quantitative study of multimodal patterns in order to address the paucity of data-driven investigations of natural discourse that would bring to the linguistic analytical table such factors as prosodic, kinesic, and interactional profiles of utterances (Feyaerts et al. 2017; Zima & Bergs 2017b).

In this dissertation, the primary data that served as my starting point were numerous examples of video clips of different speakers verbalizing the same utterances taken from actual, unscripted usage situations. I was as interested in the bodily enactment as the verbal signal when these utterances were spoken. In this way, I have repositioned the multimodal context of spontaneous language use at the centre of linguistic analysis. I have investigated a variety of spoken utterances across a range of conceptual domains that are co-produced with meaningful bodily behaviour. I have developed multimodal profiles for these utterances and included such factors as the rate at which target utterances are accompanied by kinesic movement, the differential recruitment of articulators across utterances, and specific patterns of manual gesture, head, brow, shoulder, and postural movements that occur with each utterance type. I have proposed that this way of capturing embodied profiles should inform our knowledge of these multimodal expressions, which I argue are fundamentally multimodal when uttered in face-to-face interactional situations. Moreover, I suggest that the degree of conventionalization that emerges in these multimodal profiles provides compelling evidence that these linguistic structures are inherently embodied and that these findings should expand both our notion of what constitutes a linguistic construction and how we expect language to behave more broadly.

Specifically, in a series of case studies, I have captured the multimodal enactment of a range of expressions that mark ASPECT, CONTRAST, and DISCOURSE NAVIGATION in North American English. English has an impoverished system of grammatical aspect, which means that aspect is frequently marked by periphrastic auxiliary constructions such as those investigated in Chapter 3. These constructions adopt the lexical aspect of the verb root to mark open aspect in expressions such as *continue to VERB* and *keep (on) VERB-ing*, while the lexical aspect of *stop* and *quit* is used to mark terminative aspect and *start* marks inceptive aspect. Compared to these relatively wide-open verbal constructions, the expression of CONTRAST in English is characterized by a narrower range of fixed and semi-fixed idiomatic expressions. In Chapter 4, I presented analyses of some of the most common expressions of CONTRAST, including fixed phrases (*on (the) one hand/on the other hand*), comparative adjectives (*better than/worse than)*, and antonymic noun phrases. While these types of expressions tend to express CONTRAST in the real world, expressions that mark CONTRAST in the hypothetical world include *should I/shouldn't I* and conditional *if*-statements. Finally, speakers of English use a range of fixed expressions to weave their way through spoken discourse and to imbue their discourse with their own subjective take on the discourse objects. These can be thought of as contrastive in that they mark a series of junctures throughout the discourse. In Chapter 5, I presented microstudies of the common resources that are recruited for discourse navigation, including a range of fixed expressions to mark parenthetical asides (*which is true; which is fine; which, by the way; how shall I put it;* and *although, you know)* and concessive markers to both move ahead in the discourse and to position the speaker towards what they have just said (*anyhow, at any rate, but anyway(s), so anyway(s), however, nevertheless,* and *that being said)*.

For all of the studies presented here, I used the Red Hen archive, an archive that consists of broadcast television programs from the United States and around the world. Red Hen contains over 400,000 hours of audiovisual data, which represents approximately five billion words of closed-captioned text. Using the Red Hen provided access to spontaneous natural discourse in interactional settings.

The findings in each chapter provide preliminary embodied profiles for utterances in the domains of ASPECT, CONTRAST, and DISCOURSE NAVIGATION. Taken together, these findings highlight the close level of interaction between speech and kinesic enactments. For example, in the study on ASPECT, the profiles for *keep* and *continue* differed from each other, as did the profiles for *stop* and *quit.* When comparing across types of aspect, the results were even stronger, with open aspect verbal expressions, with auxiliaries *continue* and *keep*, differing significantly from phase aspect verb expressions with auxiliaries *start, stop* and *quit*. The significant differences were found in kinesic properties of manual gesture that have largely not been identified in studies of gesture, such as the difference in onset time between a gesture and its related utterance, and stroke segmentation, to name just two factors. In the case studies of CONTRAST, meaningful differences were found to be associated with varying degrees of fixedness in a linguistic utterance. That is, for highly fixed linguistic utterances, the kinesic enactment was largely highly fixed. By contrast, in cases in which the linguistic utterance featured a high degree of variability, so, too, did the kinesic enactment. Findings also showed that some of the conventionalized features that marked CONTRAST across real world and hypothetical contexts were similar. Characteristics of the enactments such as symmetry and laterality – which have not been considered in kinesic annotations to date – play an important role in both domains. Finally, by extending the study of CONTRAST to discourse, I examined how speakers create junctures in their discourse with asides and then returns to the main discourse. This provided a broader perspective on CONTRAST and included more highly stanced utterances beyond the sentence level.

The case studies presented in Chapter 5 investigated discourse navigation. Just as some aspects of the conventionalized kinesic forms for the aspect-marking auxiliaries *keep* and *continue* and *stop* and *quit* from Chapter 3 were shown to be differentially enacted, the utterances that mark asides and returns were also differentially enacted. The highly idiomatic *which is fine* construction*,* for example, was shown to be enacted quite differently from the epistemically stronger construction, *which is true.* The body knows, and conveys, the difference between 'near neighbours' such as *which is fine,* and *which is true*; *what can I say?* and *what can I do?*; *stop* and *quit*; *keep (on)* and *continue*; *on (the) one hand* and *on the other hand;* and *however* and *anyway(s)*. The differences that were found in the kinesic enactments for these pairs demonstrate that utterances convey very finely grained differences in linguistic function and meaning, and that these distinctions can occur at the grammatical level in aspect-marking constructions, in phrasal expressions that mark contrast, and in stance and discourse navigation markers. The findings presented here support a view of language in which there is an intricate interaction between constructional force and lexical force, i.e., constructions exhibit commonalities of structure with differences in their discourse, semantic, and/or pragmatic meaning dictated by the lexicon (as in the case of *fine* and *true* in the [which is ADJ] construction). One of the principal aims of this dissertation has been to explore how, far from being restricted to speech or text renderings of language, this constructional specificity is in full gear in spontaneous, embodied language use.

The findings presented here raise important theoretical implications for a cognitive view of language and may be applied to diverse research agendas. In the remainder of this chapter, I first discuss theoretical implications of this research (§6.2). Issues that I address include the impact of the convention and variation in the multimodal signal on a cognitive view of language and on the nature of gesture, as well as the implications for a view of language as inherently multimodal. I also preview some applications of this research in areas such as language documentation, pedagogy, and acquisition, as well as technological applications ranging from clinical practice to entertainment (§6.3). The chapter ends with directions for future research and final observations in §6.4.

6.2 Implications for a constructionalist view of language

Convention in gesture has been argued to arise, at least partially, from real world physical actions; however, gestures are also known to be highly abstracted forms. In this section, I discuss the high degree of both conventionalization and variation present in the multimodal profiles that emerged in the studies presented throughout this dissertation. Many of the forms do have iconic roots in real-world action; however, I also suggest that there is a high degree of schematization that is involved, in both hand and upper body forms. In this section, I review the findings in the context of convention and variation in form (§6.2.1). I then close with a discussion of schematicity as a process of grammaticalization of kinesic forms (§6.2.2).

6.2.1 Convention and variation

The interplay between convention and variation in linguistic form has been a key interest in cognitive linguistics since its inception in the 1980s. Language is based on conventionalization. The basic unit of analysis in a cognitive linguistic view, namely the construction, rests on the understanding of language as "a structured inventory of *conventional* units" (Langacker 1991a: 15, italics mine), units that emerge and are routinized out of specific usage events.
As corpus studies of actual language usage have shown, spoken language is characterized by highly conventionalized behaviour that consists of 'chunks' – full and partial patterns of language usage. That characteristic of spoken language usage is also evident in co-speech behaviour. While gesture and other co-speech behaviours have been described as idiosyncratic, opportunistic, and local, they have also been shown to be a source of conventional, routine, and pre-fabricated co-speech movements (Streeck 2009: 5; Müller, Cienki, et al. 2013; Müller et al. 2014; Müller 2017). The findings presented throughout this dissertation suggest that there are some multimodal profiles that are strongly

conventionalized. When the speech signal featured a highly fixed utterance, such as *I quit, on (the) one hand/on the other hand,* or *which is fine*, findings showed a high degree of conventionalization in the body signal as well. The profile for *quit* was dominated by an outward moving motion on the lateral axis with both hands featuring one action phase. The expression *on (the) one hand/on the other hand* featured a high degree of symmetrical execution (i.e. the same hand form on either side of the utterance) and of the use of bilateral space. This profile accounted for 80% of the enactments. In the discourse navigation chapter, *which is fine* featured a high proportion (69%) of palm-down or palm-away gestures. The same was true for *what can I say* and *what can I do,* in which the former was shown to have a relatively fixed profile compared to the latter. In sum, these (and many more) utterances were shown to have highly conventionalized embodied profiles.

Of course, on the other side of the conventionality coin one finds variation. I believe the inherent variation in the multimodal signal is largely the reason that the interest in including co-speech behaviour in linguistic analysis has developed somewhat reluctantly: there is simply so much variation in the multimodal signal. Gestures are "an unruly bunch" (Streeck 2009: 181), with a great deal of idiosyncrasy built in. They are impacted by a multitude of factors ranging from a speakers' propensity to gesture and the local interactional environment to factors such as what is most salient in any given utterance context. However, variation is not new to language analysis. Historically, linguists have tackled the problem of variation in language structure differently depending on their theoretical commitments. In its adherence to idealized language as the object of study, mainstream (i.e. generative) linguistics in this century has viewed variability (indeed, actual usage phenomena of any kind) as irrelevant to 'real language'. On the other end of the spectrum are sociolinguists, field linguists, and interactional linguists (including conversation analysts), who have acknowledged the variation and made it their object of study head on; however, the interplay of convention and variation in language usage and the importance of both to the description and analysis of language has largely been the domain of cognitive and corpus linguists. I propose here that multimodal analysis opens up a new arena in which to explore how these factors are harnessed for expressive power in language. Rather than being distracted by the sheer variety of articulations and the 'unruly' nature of the type of gestures examined here, I take the position that the interplay between convention and variation is not unique to multimodal enactments, but rather is a property of language as a whole.

Each of the profiles presented here revealed some degree of variation. While many target utterances had kinesic profiles that were highly fixed, there was never 100% execution of a form. Even the more fixed constructions mentioned above were shown to exhibit some degree of variation in the linguistic context, the degree of enactment, and the forms the articulators took. For example, one instance of *quit talking* was gestured with six cyclic gestures – a movement profile that characterizes the profiles for *continue* and *keep* and is quite at odds with the conventionalized form of *quit,* the palm-down, bilateral, outward-moving gesture. Rather than profiling the quitting action, in this embodied utterance, the speaker profiles the ongoing nature of the talking in the event construal. To use an example from discourse navigation (Chapter 5), while most examples of *which is fine* featured palm-down and holding away, a handful featured quite the opposite form, the PUOH. This imbues *which is fine* with the sense of presenting a more neutral stance than the openly concessive palm-down stance; however, while the kinesic profiles varied widely, at times, much of this variation can be accounted for.

Looking at spontaneous discourse brings into the analytical arena a vast array of contextual factors that contribute to variation. It makes the study of language much messier. It moves the discussion beyond the structural units at the lexico-syntactic level that were shown to influence kinesic enactments, such as specific inflections, verb senses, aspectual construals, and adjectival construction fillers, to name a few. Rather, taking natural discourse as the object of study demands that language analysis take into account subjective forces such as viewpoint, stance, and semantic prosody, as well as alignment with interlocutors. On top of these factors come the influence of genre, personal idiolects, and many other contextual factors. While these clearly play a role in form-meaning pairings at a text-based level, they become increasingly difficult to ignore at a multimodal level. I believe that the perception of variation as problematic is due largely to an undue emphasis, historically, on idealized, decontextualized, and constructed examples of linguistic phenomena. Even while cognitive and functional linguists deconstructed the notion of idealized language and opened the door to language in use, this theoretical openness has not been matched with tools and methods with which to do justice to the inherent specificities of spoken and embodied language use.

In using corpus methods, the studies presented here have abstracted away from the qualitative descriptions of individual enactments that have characterized gesture studies to date. By looking across 20-50 tokens of each utterance type and including utterances from several domains of expression, the case studies captured the ways in which co-speech behaviours can be schematically related in these domains. I suggest that, in addition to the sources of variation described above, the schematization of kinesic form accounts for a large degree of variation in kinesic profiles. I explore this in the next section.

6.2.2 Schematicity, iconicity, and grammaticalization in the body signal

Iconic gestures have been described as those manual gestures that depict concrete objects and spatial relationships, while metaphoric gestures are those that refer to abstract notions. In addition to iconicity and metaphoricity, convention is often seen as a third way in which gestures capture a meaning relationship. Some gesture researchers acknowledge the multiple ways that gestures can refer, but in general the field continues to distinguish referential gestures, which relate to events, characters, and ideas, from pragmatic gestures, which operate at the level of the interaction itself. Iconicity in gesture has, for a long time, been the purview of referential gestures, while pragmatic gestures have been shown to be rooted in metaphoric (i.e. more abstract) processes. An example of a metaphorically based pragmatic gesture is the PUOH form, which metaphorically represents the presentation of a discourse object to a receiver. Similarly, the holding away gesture metaphorically represents a blocking action that can reflect the speaker's dismissing an element in the discourse.

Throughout this dissertation, I have focused on exploring a higher-order iconicity that is not reduced to the distinction between referential and pragmatic gesture. For example, in the chapter on ASPECT marking (Chapter 3), results showed that the timing of a gesture is iconically related to the aspect type it is representing. In the CONTRAST and DISCOURSE NAVIGATION studies showcased in Chapters 4 and 5, handshape was shown to play a role in conventionalized expressions, but so too, were the lateral use of space to indicate CONTRAST and shifts of the upper body to indicate discourse junctures. With the exception of iconic handshapes, all these cases are examples of a type of iconic representation that has gone largely unrecognized in gesture and co-speech behaviour patterns. I have argued that the patterns in the kinesic behaviour presented here should challenge us to expand our notion of iconicity and its locus in kinesic patterns of expression. We must expect iconicity to be more schematic, rather than limited to the obviously depictive and particular meanings that can be expressed through gestural forms.

In addition to being heralded as largely iconic forms, all gestures are inherently metonymic, evoking "salient aspects of a certain entity, scenario, action routine, or thought process" (Mittelberg 2018: 2). That is, gestures are necessarily partial representations. This characteristic renders them excellent candidates for reduction, one of the primary characteristics of the process known as grammaticalization, in which particular grammatical markers evolve historically from particular lexical items (as was seen with the auxiliary verbs featured in Chapter 3). The process of grammaticalization generally involves both the phonological reduction and pragmatic strengthening of a (lexical) form (Heine & Kuteva 2002; Heine 2013). Take, for example, the collocation *you know* when used as a discourse marker, often pronounced *y'know.* As a discourse marker *you know* no longer participates in a proposition, instead conveying pragmatic force on the utterance. In the same vein of grammaticalization, in which an utterance is both phonologically reduced and takes on a pragmatic force, I propose that gesture forms are grammaticalized from their prototypical forms to a reduced articulation. To take a simple example, a prototypical shrug may involve bilateral PUOH gestures, a shoulder movement up and then down, a mouth shrug, and a head tilt; a reduced version would consist simply of a very slight movement of one shoulder. In fact, this type of reduced shrug form is frequently observed with the reduced speech string, *I dunno,* itself grammaticalized from its full form, *I don't know.*

From the utterances examined in this dissertation, the enactments of *which is fine; which, by the way;* and *but anyway(s),* from the set of discourse navigation markers investigated in Chapter 5, illustrate grammaticalization in the manual gesture form. The hand form for each of these utterances is dominated by the palm-down open-hand (PDOH) gesture and holding away gesture, to varying degrees. Both of these recurrent gestures feature a flat hand shape; in the PDOH, the forearm is horizontal such that the palm is oriented downwards, while in the prototypical holding away gesture is executed with a vertical forearm such that the palm faces forwards. However, in the multimodal profiles that characterized the utterances mentioned above, the holding away gestures generally featured an angled wrist or forearm such that the hand was at approximately 45 degrees (not full-fledged, forward facing holding away gestures). The PDOH forms featured here likewise did not feature a horizontal forearm and wrist; rather, the wrist was angled upwards slightly, only not as much as for the holding away gestures. In gesture literature, palm-down and holding away gesture forms have been treated as distinct entities within the open-hand prone family. This may indeed hold when they are used in propositional contexts, such as when the holding away gesture depicts a metaphorical holding away of an object. However, in their phonologically reduced forms as seen in the enactments of [WIF], [WBTW], and [but anyway(s)], these two forms are less distinct (e.g. differing only by the slight difference in the angle of the wrist or forearm). These multimodal utterances also exhibit a diffuse pragmatic force, another marker of grammaticalization. Rather than conveying negation and pushing back, respectively, the PDOH and holding away gestures in their reduced forms may display less of a distinction between these two pragmatic forces.

The findings presented here suggest that upper body movements can also be considered in light of grammaticalization. As discussed in relation to the findings in Chapters 4 and 5, upper body movements are iconic in more abstract ways than gesture. Think for example of the lateral and symmetrical forms that emerged for contrast marking. It may be somewhat intuitive that speakers used the presentational gesture form, the PUOH, to propose or offer two options when expressing real world contrast. The conventionalized use of head movements to mark bilateral space in the irrealis domain represents a schematization of the more overt and depictive use of lateral space that characterized gestures of contrast. In a similar way, head movement was shown to indicate a juncture in discourse. The two examples given here, of the reductions in the holding away and palm-down gestures that render them less distinct from one another, and the conventionalized use of head movement to mark lateral space, exemplify a degree of abstraction that is highly relevant to meaning construction and meaning extension.

I began this section with a discussion of the tension between convention and variation. I have suggested that this tension is normal and the disparity between convention and variation is how language extends itself. Both the conventionalization and variation in co-speech enactments comes about because of reduction at the form level. Here I have proposed that reduction, in addition to the diffusion of pragmatic force, holds for gestures and other kinesic articulations in the same way as for the verbal utterance. The features I have identified – the high degree of convention and variation and grammaticalization of the signal – are phenomena that have been shown to motivate language structure and language use. As I hope to have shown, these phenomena also motivate usage patterns in the multimodal signal.

In this section, I have discussed several theoretical contributions of this work. In the next section, I examine the applications of this work to other areas of research in linguistics and in other fields.

6.3 Applications in linguistics and beyond

The research presented throughout this dissertation has shown a tight relationship between specific conceptual domains, a set of utterances that characterize the domain, and a constrained set of bodily enactments. The findings have shown that the multimodal enactments are nuanced and can mark subtle distinctions in linguistic forms that, in traditional grammars, are frequently ascribed near-synonymous functions. Therefore, in addition to the theoretical implications for what we understand about human language, the research has applications in linguistic subfields such as language documentation and language pedagogy. It also has applications further afield, such as in research areas and industries in which human-like language use is delivered by artificial or digital avatars, typically called embodied conversational agents. These include clinical practice and artificial intelligence, among others.

6.3.1 Language documentation and language pedagogy

It has been estimated that as many as half of the world's approximately 7,000 languages are expected to be moribund by the end of this century (Evans 2010). Quite aside from this startling rate of loss, the vitality of the remaining languages is also at risk, with an aged population of native speakers in many speech communities. Even those languages that are well-studied and have more robust speaker communities require language resources for education, among other forms of investment in their language and culture.

Before modern technological advances, field linguists recorded language use and their analyses using pen and paper. As audio recordings became available, these became the standard for language documentation efforts, alongside vocabulary lists and other documentation tools. The theoretical concerns of linguists, however, impeded the adoption of tools to record language use in interaction. In past decades, the types of phenomena that have been the subject of this dissertation, such as periphrastic expressions, fixed and idiomatic expressions, highly subjective and multimodal instances of language use, have not been central objects of linguistic enquiry. They have, therefore, not been documented in field work on Indigenous languages (ILs). With high quality video recording now available on a handheld device such as a cell phone, field linguists have a great opportunity to record video of spontaneous conversations when they are in the field. As Rice states, by approaching ILs from the perspective of meaning and usage,

> [w]e will better convey the import and function of markers of stance, viewpoint, and inter-subjectivity, of discourse particles and repair devices, and of signals of conversational floor-holding or floor-yielding in the prosody, the morphosyntax, and the body, as used by an IL speaker. These are the phenomena that truly bring a language alive. Hopefully, CL can play a bigger role in reviving threatened ILs which have either never been described or have been underanalyzed in unconvincing, overly structural, de-contextualized, and universalist ways, never quite capturing the imagination of those who prefer to encounter a language through a semantic and interactional lens.
>
> (Rice 2017b: 58)

The shift to more and specifically multimodal data will promote the study of phenomena that are at the heart of language in use for Indigenous languages.

It is not only the teaching of Indigenous languages that can benefit from the cognitive and interactional approach supported in this dissertation. Given the specific, conventionalized pairings of fixed utterances and co-speech behavior, the research presented here could also serve language pedagogy. In a second-language classroom, learners are tasked with acquiring the nuanced, highly idiomatic behaviours of a new language. Text-based corpora have been adopted into language pedagogy to some extent to provide illustrations with actual instances of language use. If second language learners had the opportunity to learn the idiomatic patterns of a language, including their embodied profiles, how could this affect their speed or accuracy in L2 acquisition? While there is research on applying cognitive linguistic approaches in the second language classroom (e.g. teaching specific constructions and usage practices) (Achard & Niemeier 2004; Verspoor & Tyler 2009) and a small literature on teachers' use of gesture in the second language classroom (e.g. teachers deploy gestures to clarify and disambiguate meaning (Lazaraton 2004)), there appear to be no studies on the pedagogical effect of teaching conventionalized kinesic patterns alongside conventionalized utterances. A shift to including the multimodal signal in a usage-based language pedagogy could lead to an improvement in oral proficiency, one of the primary goals of second language instruction.

In this section to this point, I have discussed the potential impact of multimodal linguistic research on Indigenous language documentation and second language teaching, both areas which rely on spoken usage in a natural environment. In the next section, I address quite the opposite: language behaviour in a technologically-driven environment.

6.3.2 Embodied conversational agents

Clinical practice

It is a moment that both doctors and patients fear: the delivery of bad news. As a patient, the moment in which a dire medical diagnosis is shared with you by your doctors may be stressful and heartbreaking. The effects of the conversation may be compounded by the manner in which it is communicated. Although doctors will be well-intentioned in these scenarios, they are often not adequately trained in the difficult task of breaking bad news (abbreviated in medical circles as BBN). A project at the Center for Virtual Reality of the Mediterranean, a joint research technology laboratory that is part of the Institute of Movement Sciences at Aix-Marseille University in France, is developing an embodied conversational agent with high-level natural language communication skills to training doctor to break bad news.[171] Doctors in medical training already receive training in BBN. These generally involve simulations between experienced or novice physicians and an actor playing the role of the patient. This project, *Acorformed*, has created a virtual training model in which the patient-actor role is played by a

[171] http://crvm.ism.univ-amu.fr/en/acorformed.html

virtual agent. The project intends to test it by having both the doctor and the patient roles played by avatars and submitting the interactions to observers who will assess which behaviours are deemed most appropriate, understandable, and desirable for BBN disclosure.[172] Figure 6.1 shows a screenshot from the project in which a doctor is wearing a virtual reality headset and interacts with an avatar patient.

Figure 6.1. Screenshot of virtual reality training scenario for doctor-patient interaction in 'breaking bad news' disclosures

The *Acorformed* project directly explores the question of multimodal natural language interaction, its theoretical basis and its experimentation in a virtual reality environment thanks to the development of an embodied conversational agent.

In a project that addresses elder loneliness using artificial intelligence, a digital chat companion is under development in Canada in a partnership between the City of Edmonton and the University of Alberta[173]. The Automated Nursing Agent (ANA) "aims to create an AI

---

172 https://www.youtube.com/watch?v=hZ0jZxYZnEcandfeature=emb_logo

173 http://twitter.com/ualbertaScience/status/1190036229787656199

companion able to hold conversations and fight loneliness" (Lyle 2019). It is designed to use artificial intelligence to identify which emotions the speaker is expressing and to provide emotion in response. While the social and ethical issues of addressing loneliness through disembodied technological creations remain paramount and at this point the project is text only, this type of digital companion must, eventually, become three dimensional. When it does, the human-like multimodal expression of the chat companion will be critical to its success.

Gaming and entertainment

Finally, embodied conversational agents and avatars are used in entertainment. The gold standard in video-game design is motion-capture (MoCap) technology in which actors are used to capture full scenes. This is frequently used for pre-programmed dialogue and the actions of the main characters; however, in large narrative-driven video games that can take weeks to play, there is too much interactional content to capture with MoCap technology. Moreover, the game player makes choices throughout the game that define the interactions that take place. Therefore, games rely on procedural animation in which dialogue and co-speech behaviour is generated 'on the fly' from a library of gestures. This is where the quality of human-like behaviour decreases markedly, creating a visible difference in the dialogue that is produced using the motion capture data and the procedural animations. To create ultimately believable characters that viewers and players can relate to, game companies are trying to close this gap; however, without building linguistic co-speech behaviour into the procedural animation system, the output will continue to be impoverished. For example, in Figure 6.2 and Figure 6.3, I provide examples of procedurally generated dialogue from a story-telling game, *Dragon Age: Inquisition*, produced by Electronic Arts studio BioWare, in Edmonton, Alberta. The first figure shows a character marking the utterance *Understood* with a head nod, a conventionalized signal associated with agreement. In the next sequence, in Figure 6.3, the male character Varric utters *well* with a head movement backwards, in keeping with the types of backward nods often seen in contexts involving a hedge or shrug. This seems a possible co-speech behaviour to pair with *well; however,* the procedural gesture system displays its weakness in the next frame. Here, the character Varric produces a bimanual PUOH gesture with the utterance *Bianca's excited,* while beginning to walk forward.[174] The effect is stilted and jarring and seems out of place in the context. These co-speech behaviours are selected from the library of procedural animations that are tagged with a limited set of basic semantic and gestural categorizations.

[174] Dialogues from Dwarf Inquisitor in *Dragon Age: Inquisition*, Part I, accessed Nov. 26, 2019 from https://www.youtube.com/watch?v=37ceU-5M-6c.

S: A. *Indeed, I find it difficult to imagine any mage having such power.* | S: B. *Understood.*

G: (n/a, speaker A off screen) | G: head nod | G: return to neutral

Figure 6.2. Cassandra in dialogue with Solas, *Dragon Age: Inquisition*

S: *Well…* | *Bianca's excited.*

G: head movement backwards | bilateral PUOH gesture

Figure 6.3. Leliana and Varric in dialogue, *Dragon Age: Inquisition*

We know that for humans in spontaneous interaction (or experimental setups that have served as proxies for such interactions), observing semantically aligned co-speech gesture increases comprehension, while observing unnatural body movements, mismatches, and inconsistencies in speech and gesture can lead to decreased comprehension (Cassell et al. 1999; Gunter et al. 2015). The application of interactional research could improve the human-like rendering of speech and co-speech behaviour in character dialogues in video games, leading to improved emotional resonance of dialogues and subsequent player satisfaction with the game. Gaming companies will need to assess whether they see (or prioritize) the potential for growth in their market and whether they will be involved in pursuing this line of research and development.

6.4 Final observations

Spoken (or signed) conversation lies at the centre of the human capacity for language. Conversation is at the centre of play and of work. It is where children acquire their native language and how language is passed between speakers from generation to generation (Enfield 2017). In comparing human and animal cognitive abilities, it's been said that "many species have this 'predicate (x)' cognition, but only humans are motivated to make it public" (Hurford 2003: 264). As one might expect from the primary modality of communication, the spoken language signal is incredibly rich. Speakers use a vast array of elements to convey precisely their desired meanings over a stunningly broad range of contexts. These elements can include the standard repertoire of lexical and grammatical material in a language, as well as fixed idioms, interjections, and prosodic features such as intonational contour, to name only a few. Corpus studies of spoken language usage events have shed light on some of the patterns and partial patterns that are at the heart of language, such as the utterances I have explored throughout this dissertation. As the findings presented here suggest, the kinesic signal – a modality that has long been ignored as 'paralinguistic', or outside of the linguistic signal – displays the same type of non-compositional yet patterned behaviour as the speech signal. Furthermore, the rich interplay between the patterns in the body and the patterns in speech suggested in this dissertation implies that multiple modalities and multiple channels, rather than the speech signal alone, constitute the linguistic signal. The primacy of embodied interaction necessitates a view of language in which the construction is considered to be multimodal in nature and in which the study of dialogue in face-to-face interaction is the essential starting point of linguistic analysis.

REFERENCES

Achard, Michel, and Niemeier, Susanne (Eds.). 2004. *Cognitive Linguistics, Second Language Acquisition, and Foreign Language*. Berlin: Mouton de Gruyter.

Barsalou, Lawrence W. 1999. Perceptual symbol systems. *Behavioral and Brain Sciences* 22: 577-660. https://doi.org/10.1017/S0140525X99002149.

Barsalou, Lawrence W. 2010. Grounded cognition. *Annual Review of Psychology* 59: 617-645. https://doi.org/10.1146/annurev.psych.59.103006.093639.

Bavelas, Janet, and Chovil, Nicole. 2018. Some pragmatic functions of conversational facial gestures. *Gesture* 17(1): 98-127. https://doi.org/10.1075/gest.00012.bav.

Bavelas, Janet, Chovil, Nicole, Coates, Linda, and Roe, Lori. 1995. Gestures specialized for dialogue. *Personality and Social Psychology Bulletin* 21: 394-405. https://doi.org/10.1177/0146167295214010.

Bavelas, Janet, and Gerwing, Jennifer. 2007. Conversational hand gestures and facial displays in face-to-face dialogue. In *Social Communication*, Klaus Fiedler (Ed.), 283-308. New York: Psychology Press.

Beeching, Kate, and Detges, Ulrich. 2014. Introduction. In *Discourse Functions at the Left and Right Periphery: A Crosslinguistic Investigation of Language Use and Language Change*, Kate Beeching and Ulrich Detges (Eds.), 1-24. Leiden: Brill.

Bergen, Benjamin. 2012. *Louder Than Words*. New York: Basic Books.

Bergen, Benjamin, and Chang, Nancy. 2005. Embodied Construction Grammar in simulation-based language understanding. In *Construction Grammars: Cognitive Grounding and Theoretical Extensions*, J.-O. Östman and M. Fried (Eds.), 147-190. Amsterdam: Benjamins.

Bergen, Benjamin, and Chang, Nancy. 2013. Embodied Construction Grammar. In *The Oxford Handbook of Construction Grammar*, Thomas Hoffman and Graeme Trousdale (Eds.), 168-190. New York: Oxford University Press.

Bergen, Benjamin K., Lindsay, Shane, Matlock, Teenie, and Narayanand, Srini. 2007. Spatial and linguistic aspects of visual imagery in sentence comprehension. *Cognitive Science* 31: 733-764. https://doi.org/10.1080/03640210701530748.

Bergen, Benjamin, and Wheeler, Kathryn. 2010. Grammatical aspect and mental simulation. *Brain and Language* 112(3): 150-158. https://doi.org/10.1016/j.bandl.2009.07.002.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, and Finegan, Edward (Eds.). 1999. *Longman Grammar of Spoken and Written English*. London: Pearson Education Limited.

Boas, Hans C., and Sag, Ivan A. 2012. *Sign-based Construction Grammar*. Stanford: CSLI Press.

Bolinger, Dwight. 1989. *Intonation and its Uses: Melody in Grammar and Discourse*. London: Edward Arnold.

Bressem, Jana. 2013. A linguistic perspective on the notation of form features in gestures. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1), C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. Mcneill, and S. Teßendorf (Eds.), 1079-1098. Berlin: De Gruyter Mouton.

Bressem, Jana, and Müller, Cornelia. 2014. The family of away gestures. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 2), C Müller,

A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and S. Teßendorf (Eds.), 1592-1604. Berlin: De Gruyter Mouton.

Brinton, Laurel. 1988. *The Development of English Aspectual Systems*. Cambridge: Cambridge University Press.

Brinton, Laurel. 2017. *The Evolution of Pragmatic Markers in English: Pathways of Change*. Cambridge: Cambridge University Press.

Brône, Geert, Oben, Bert, Jehoul, Annelies, Vranjes, Jelena, and Feyaerts, Kurt. 2017. Eye gaze and viewpoint in multimodal interaction management. *Cognitive Linguistics* 28(3): 449-483. https://doi.org/10.1515/cog-2016-0119.

Brookes, Heather. 2004. A repertoire of South African quotable gestures. *Journal of Linguistic Anthropology* 14(2): 182-224. https://doi.org/10.1525/jlin.2004.14.2.186.

Bybee, Joan. 2006. From usage to grammar: The mind's response to repetition. *Language* 82(4): 711-733.

Bybee, Joan. 2010. *Language, Usage and Cognition*. New York: Cambridge University Press.

Calbris, Geneviève. 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. https://doi.org/10.1075/gest.3.1.03cal.

Calbris, Geneviève. 2011. *Elements of Meaning in Gesture*. Amsterdam: John Benjamins.

Casasanto, Daniel. 2008a. Conceptual affiliates of metaphorical gestures. *Paper presented at the International Conference on Language, Communicaiton, And Cognition*.

Casasanto, Daniel. 2008b. Similarity and proximity: When does close in space mean close in mind? *Memory & Cognition* 36(6): 1047-1056. https://doi.org/10.3758/MC.36.6.1047.

Casasanto, Daniel, and Bottini, Roberto. 2014. Mirror reading can reverse the flow of time. *Journal of Experimental Psychology: General* 143(2): 473-479.

Casasanto, Daniel, and Chrysikou, Evangelia G. 2011. When left is "right": Motor fluency shapes abstract concepts. *Psychological Science* 22(4): 419-422. https://doi.org/10.1177%2F0956797611401755.

Casasanto, Daniel, and Henetz, Tanya. 2012. Handedness shapes children's abstract concepts. *Cognitive Science* 36: 359–372. https://doi.org/10.1111/j.1551-6709.2011.01199.x.

Casasanto, Daniel, and Jasmin, Kyle. 2010. Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *Plos One* 5(7): 1-5. https://doi.org/10.1371/journal.pone.0011805.

Casasanto, Daniel, and Jasmin, Kyle. 2012. The hands of time: Temporal gestures in English. *Cognitive Linguistics* 23(4): 643-674. https://doi.org/10.1515/cog-2012-0020.

Cassell, Justine, McNeill, David, and McCullough, Karl-Erik. 1999. Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information. *Pragmatics & Cognition* 7(1): 1-34. https://doi.org/10.1075/pc.7.1.03cas.

Chira, Susan. 2017, June 14, 2017. The universal phenomenon of men interrupting women. *The New York Times*. Retrieved from https://www.nytimes.com/2017/06/14/business/women-sexism-work-huffington-kamala-harris.html

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge: MIT Press.

Chovil, Nicole. 1989. *Communicative functions of facial displays in conversation* (Doctoral dissertation). University of Victoria, Victoria, Canada.

Chung, Sandra, and Timberlake, Alan. 1985. Tense, aspect, and mood. In *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon* (Vol. 3), Timothy Shopen (Ed.), 202-258. Cambridge: Cambridge University Press.

Cienki, Alan. 1997. Some properties and groupings of image schemas. In *Lexical and Syntactical Constructions and the Construction of Meaning*, Marjolijn Verspoor, Kee Dong Lee, and Eve Sweetser (Eds.), 3-16. Amsterdam: John Benjamins.

Cienki, Alan. 1998. STRAIGHT: An image schema and its metaphorical extensions. *Cognitive Linguistics* 9(2): 107-149. https://doi.org/10.1515/cogl.1998.9.2.107.

Cienki, Alan. 2005. Image schemas and gesture. In *From Perception to Meaning: Image Schemas in Cognitive Linguistics*, Joseph E. Grady and Beate Hampe (Eds.), 421-441. Berlin: Mouton de Gruyter.

Cienki, Alan. 2013. Image schemas and mimetic schemas in cognitive linguistics and gesture studies. *Review of Cognitive Linguistics* 11(2): 417-432. https://doi.org/10.1075/rcl.11.2.13cie.

Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.

Comrie, Bernard. 1976. *Aspect: An Introduction to the Study of Verbal Aspect and Related Problems*. Cambridge: Cambridge University Press.

Cooperrider, Kensy. 2018. Foreground gesture, background gesture. *Gesture* 16(2): 176-202. https://doi.org/10.1075/gest.16.2.02coo.

Cooperrider, Kensy, and Goldin-Meadow, Susan. 2017. Gesture, language, and cognition. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.), 118-134. Cambridge, UK: Cambridge University Press.

Cooperrider, Kensy, and Núñez, Rafael. 2009. Across time, across the body: Transversal temporal gestures. *Gesture* 9(2): 181-206. https://doi.org/10.1075/gest.9.2.02coo.

Cooperrider, Kensy, and Núñez, Rafael. 2012. Nose-pointing: Notes on a facial gesture of Papua New Guinea. *Gesture* 12(2): 103-129. https://doi.org/10.1075/gest.12.2.01coo.

Croft, William. 2001. *Radical Construction Grammar*. Oxford: Oxford University Press.

Croft, William. 2012. *Verbs: Aspect and Causal Structure*. Oxford: Oxford University Press.

Croft, William, and Cruse, D. Allan. 2004. *Cognitive Linguistics*. Cambridge: Cambridge University Press.

Dancygier, Barbara. 2012. Negation, stance verbs, and intersubjectivity. In *Viewpoint in Language: A Multimodal Perspective*, Barbara Dancygier and Eve Sweetser (Eds.), 69-93. Cambridge: Cambridge University Press.

Dancygier, Barbara, and Sweetser, Eve. 2005. *Mental Spaces in Grammar: Conditional Constructions*. Cambridge: Cambridge University Press.

Danesi, Marcel. 2009. Opposition theory and the interconnectedness of language, culture, and cognition. *Sign Systems Studies* 37(1/2): 11-41.

Danesi, Marcel, and Perron, Paul. 1999. *Analyzing Cultures*. Bloomington: Indiana University Press.

Davies, Mark. 2008. The Corpus of Contemporary American English (COCA): 560 million words, 1990-present. Available online at http://corpus.byu.edu/coca/.

de Saussure, Ferdinand. 1916. *Course in General Linguistics* (Wade Baskin, Trans.). New York: McGraw Hill.

Debras, Camille. 2013. *L'expression multimodale du positionnement interactionnel ('multimodal stance-taking'): Étude d'un corpus oral vidéo de discussions sur l'environnement enanglais britannique* (Doctoral dissertation). Sorbonne Nouvelle – Paris 3 University, Paris, France.

Debras, Camille. 2017. The shrug: Forms and meanings of a compound enactment. *Gesture* 16(1): 1-34. https://doi.org/10.1075/gest.16.1.01deb.

Debras, Camille, and Cienki, Alan. 2012, September 3-5. *Some uses of head tilts and shoulder shrugs during human interaction, and their relation to stancetaking.* Paper presented at the International Conference on Social Computing, Amsterdam, Netherlands.

Divjak, Dagmar, and Gries, Stefan Th. 2006. Ways of trying in Russian: clustering behavioral profiles. *Corpus Linguistics and Linguistic Theory* 2(1): 23-60. https://doi.org/10.1515/CLLT.2006.002.

Du Bois, John W. 2007. The stance triangle. In *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, Robert Englebretson (Ed.), 138-182. Amsterdam: John Benjamins.

Dudis, Paul G. 2004. Body partitioning and real-space blends. *Cognitive Linguistics* 15(2): 223-238.

Duncan, Susan. 2002. Gesture, verb aspect, and the nature of iconic imagery in natural discourse. *Gesture* 2(2): 183-206. https://doi.org/10.1075/gest.2.2.04dun.

Efron, David. 1972. *Gesture, Race and Culture*. The Hague: Mouton and Co.

Ekman, Paul. 1993. Facial expression and emotion. *American Psychologist* 48(4): 384-392. http://doi.org/10.1037/0003-066X.48.4.384.

Ekman, Paul, and Friesen, Wallace. 1969. The repertoire of nonverbal behaviour: Categories, origin, usage, and coding. *Semiotica* 1: 49–98.

Ekman, Paul, and Friesen, Wallace V. 1981. The repertoire of nonverbal behavior: Categories, origins, usage and coding. In *Nonverbal Communication, Interaction, and Gesture*, Adam Kendon (Ed.), 57-105. The Hague: Mouton and Co.

Enfield, Nick. 2013. A 'Composite Utterances' approach to meaning. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1), C. Muller, E. Fricke, A. Cienki, and D. McNeill (Eds.), 689-707. Berlin: Mouton de Gruyter.

Enfield, Nick. 2017. *How We Talk: The Inner Workings of Conversation*. New York: Basic Books.

Enfield, Nick. 2019. *Enchrony*. Paper presented at the 15th International Cognitive Linguistics Conference, Nishinomiya, Japan.

Evans, Nicholas. 2010. *Dying Words: Endangered Languages and What They Have to Tell Us*. Indianapolis: Wiley-Blackwell,.

Evans, Nicholas, and Wilkins, David. 2000. In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language* 76(3): 546-592.

Feyaerts, Kurt, Brône, Geert, and Oben, Berndt. 2017. Multimodality in interaction. In *The Cambridge Handbook of Cognitive Linguistics*, B. Dancygier (Ed.), 135-156. Cambridge: Cambridge University Press.

Fillmore, Charles J. 1982. Frame semantics. In *Linguistics in the Morning Calm*, The Linguistic Society of Korea (Ed.), 111-137. Seoul: Hanshin.

Fillmore, Charles J. 1985. Frames and the semantics of understanding. *Quaderni di Semantica* 6(2): 222-254.

Fillmore, Charles J., Kay, Paul, and O'Connor, Mary Catherine. 1988. Regularity and idiomaticiy in grammatical constructions: the case of Let Alone. *Language* 64(3): 501-538.

Ford, Cecilia. 2000. The treatment of contrasts in interaction. In *Cause – Condition – Concession – Contrast: Cognitive and Discourse Perspectives*, Elizabeth Couper-Kuhlen and Bernd Kortmann (Eds.), 283-311. Berlin: De Gruyter.

Frawley, William. 1992. *Linguistic Semantics*. New York: Routledge.

Fricke, Ellen. 2014. Kinesthemes: Morphological complexity in co-speech gestures. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 2), Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, and S. Tessendorf (Eds.), 1618-1629. Berlin: De Gruyter Mouton.

Gaby, Alice Rose, and Sweetser, Eve. 2017. Space-time mappings beyond language. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.), 635-650. Cambridge, UK: Cambridge University Press.

Geeraert, Kristina. 2016. *Climbing on the bandwagon of idiomatic variation: A multi-methodological approach* University of Alberta,

Gibbs, Raymond W. Jr. 2005. *Embodiment and Cognitive Science*. New York: Cambridge University Press.

Gibbs, Raymond W. Jr. 2017. Embodiment. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.). Cambridge, UK: Cambridge University Press.

Givens, David. 1977. Shoulder shrugging: a densely communicative behavior. *Semiotica* 19 (1-2): 13–29. https://doi.org/10.1515/semi.1977.19.1-2.13.

Glenberg, Arthur M., and Kaschak, Michael P. 2002. Grounding language in action. *Psychonomic Bulletin & Review* 9(3): 558-565.

Goldberg, Adele E. 1995. *Constructions*. Chicago: University of Chicago Press.

Goldberg, Adele E. 2003. Constructions: a new theoretical approach to language. *Trends in Cognitive Sciences* 7(5): 219-224. https://doi.org/10.1016/S1364-6613(03)00080-9.

Goldberg, Adele E. 2006. *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Gries, Stefan Th. 2006. Corpus-based methods and cognitive semantics: The many meanings of to run. In *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*, Stefan Th. Gries and Anatol Stefanowitsch (Eds.), 57-99. Berlin: Mouton de Gruyter.

Gries, Stefan Th., Hampe, Beate, and Schönefeld, Doris. 2005. Converging evidence: bringing together experimental and corpus data on the association of verbs and constructions. *Cognitive Linguistics* 16(4): 635-676. https://doi.org/10.1515/cogl.2005.16.4.635.

Gries, Stefan Th., Hampe, Beate, and Schönefeld, Doris. 2010. Converging Evidence II: More on the association of verbs and constructions. In *Empirical and Experimental Methods in Cognitive/Functional Research*, Sally Rice and John Newman (Eds.), 59-72. Stanford: CSLI

Gries, Stefan Th., and Stefanowitsch, Anatol. 2004. Extending collostructional analysis: a corpus-based perspective on 'alternations'. *International Journal of Corpus Linguistics* 9(1): 97-129. https://doi.org/10.1075/ijcl.9.1.06gri.

Grosz, Barbara J., and Sidner, Candace L. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12(3): 175-203.

Guilbeault, Douglas. 2017. How politicians express different viewpoints in gesture and speech simultaneously. *Cognitive Linguistics* 28(3): 417-447. https://doi.org/10.1515/cog-2016-0086.

Gunter, Thomas C., Weinbrenner, J. E. Douglas, and Holle, Henning. 2015. Inconsistent use of gesture space during abstract pointing impairs language comprehension. *Frontiers in Psychology* 6(80). https://doi.org/10.3389/fpsyg.2015.00080.

Hampe, Beate, and Grady, Joseph E. (Eds.). 2005. *From Perception to Meaning: Image Schemas in Cognitive Linguistics*. Berlin: Mouton de Gruyter.

Harrison, Simon. 2013. *The temporal coordination of negation gestures in relation to speech*. Paper presented at the Tilburg Gesture Research Meeting (TiGER), Tilburg, NL.

Harrison, Simon. 2018. *The Impulse to Gesture: Where Language, Minds, and Bodies Intersect*. Cambridge: Cambridge University Press.

Heine, Bernd. 2013. On discourse markers: Grammaticalization, pragmaticalization, or something else? *Linguistics* 51(6): 1205-1247. https://doi.org/10.1515/ling-2013-0048.

Heine, Bernd, and Kuteva, Tania. 2002. *World Lexicon of Grammaticalization*. Cambridge: Cambridge University Press.

Hilton, Katherine 2018. *What Does an Interruption Sound Like?* (Doctoral dissertation). Stanford University, Stanford, USA.

Hinnell, Jennifer. 2018. The multimodal marking of aspect: The case of five periphrastic auxiliary constructions in North American English. *Cognitive Linguistics* 29(4): 773-806. https://doi.org/10.1515/cog-2017-0009.

Hinnell, Jennifer. 2019. The verbal-kinesic enactment of contrast in North American English. *The American Journal of Semiotics* 35(1-2): 55-92. https://doi.org/10.5840/ajs20198754.

Hinnell, Jennifer, and Rice, Sally. 2014. *Inflection-specific gestural constructions in English: The case of the catenative auxiliary*. Paper presented at the International Society of Gesture Studies, San Diego, USA.

Hinnell, Jennifer, and Rice, Sally. 2016. *'On the one hand… ': Opposition and optionality in the embodied marking of stance in North American English*. Paper presented at the 7th meeting of the International Society for Gesture Studies, Paris, France.

Hinnell, Jennifer, and Rice, Sally. 2019. *The embodied marking of stance in North American English: Stacked and idiomatic*. Paper presented at the 15th International Association of Cognitive Linguistics, Nishinomiya, Japan.

Hoffmann, Thomas. 2017. Construction grammars. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.), 310-329. Cambridge: Cambridge University Press.

Horne, Merle, Hansson, Petra, Gösta, Bruce, Frid, Johan, and Filipsson, Marcus. 2001. Cue words and the topic structure of spoken discourse: The case of Swedish men 'but'. *Journal of Pragmatics* 33(7): 1061-1081. https://doi.org/10.1016/S0378-2166(00)00044-8.

Hostetter, Autumn B., and Alibali, Marta W. 2008. Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review* 15(3): 495-514. https://doi.org/10.3758/PBR.15.3.495.

Huette, Stephanie, Winter, Bodo, Matlock, Teenie, Ardell, David H., and Spivey, Michael. 2014. Eye movements during listening reveal spontaneous grammatical processing. *Frontiers in Psychology* 5: 1-7.

Hurford, James, R. 2003. The neural basis of predicate-argument structure. *Behavioral and Brain Sciences* 26(3): 261-283. https://doi.org/10.1017/S0140525X03000074.

Ibbotson, Paul, Lieven, Elena V.M., and Tomasello, Michael. 2013. The attention-grammar interface: Eye-gaze cues structural choice in children and adults. *Cognitive Linguistics* 24(3): 457-481. https://doi.org/10.1515/cog-2013-0020.

Jakobson, Roman. 1962. *Selected Writings*. The Hague: Mouton.

Janda, Laura A. 2017. The quantitative turn. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.), 498-514. Cambridge: Cambridge University Press.

Janzen, Terry. 1998. *Topicality in ASL: Information Ordering, Constituent Structure, and the Function of Topic Marking* (Doctoral dissertation). University of New Mexico, Albuquerque, USA.

Janzen, Terry. 1999. The grammaticization of topics in American Sign Language. *Studies in Language* 23(2): 271-306. https://doi.org/10.1075/sl.23.2.03jan.

Janzen, Terry. 2012. Two ways of conceptualizing space: Motivating the use of static and rotated vantage point space in ASL discourse. In *Viewpoint in Language*, B. Dancygier and E. Sweetser (Eds.), 156-176. New York: Cambridge.

Janzen, Terry. 2017. Composite utterances in a signed language: Topic constructions and perspective-taking in ASL. *Cognitive Linguistics* 28(3): 511-538. https://doi.org/10.1515/cog-2016-0121.

Jehoul, Annelies, Brône, Geert, and Feyaerts, Kurt. 2017. The shrug as marker of obviousness. *Linguistics Vanguard* 3(s1): 1-9. https://doi.org/10.1515/lingvan-2016-0082.

Johnson, Mark. 1987. *The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason*. Chicago, IL: University of Chicago.

Johnson, Mark. 2017. *Embodied Mind, Meaning, and Reason: How our bodies give rise to understanding*. Chicago: University of Chicago Press.

Kemmer, Suzanne, and Barlow, Michael (Eds.). 2000. *Usage-Based Models of Language*. Stanford: CSLI Publications.

Kendon, Adam. 1980. Gesture and speech: Two aspects of the process of utterance. In *Nonverbal Communication and Language*, Mary Ritchie Key (Ed.), 207–288. The Hague: Mouton.

Kendon, Adam. 2004. *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press.

Kuteva, Tania. 2001. *Auxiliation: An Enquiry into the Nature of Grammaticalization*. Oxford: Oxford University Press.

Ladewig, Silva H. 2011. Putting the cyclic gesture on a cognitive basis. *CogniTextes* 6: 1-23.

Ladewig, Silva H. 2014. Recurrent gestures. In *Body – language – Communication: An international handbook on multimodality in human interaction* (Vol. 2), C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. McNeill, and J. Bressem (Eds.), 1558–1574. Berlin: De Gruyter Mouton.

Lakoff, George. 1990. The invariance hypothesis: Is abstract reason based on image-schemas? *Cognitive Linguistics* 1(1): 39-74. https://doi.org/10.1515/cogl.1990.1.1.39.

Lakoff, George, and Johnson, Mark. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.

Lakoff, George, and Johnson, Mark. 1999. *Philosophy in the Flesh: The Embodied Mind and its Challenge to Western Thought*. New York: Basic Books.

Langacker, Ronald W. 1987. *Foundations of Cognitive Grammar: Theoretical Prerequisites* (Vol. 1). Stanford: Stanford University Press.

Langacker, Ronald W. 1991a. *Concept, Image, and Symbol: The Cognitive Basis of Grammar*. Berlin and New York: Mouton de Gruyter.

Langacker, Ronald W. 1991b. *Foundations of Cognitive Grammar: Descriptive Application* (Vol. 2). Stanford: Stanford University Press.

Langacker, Ronald W. 2001. Discourse in cognitive grammar. *Cognitive Linguistics* 12(2): 143-188. https://doi.org/10.1515/cogl.12.2.143.

Langacker, Ronald W. 2008. *Cognitive Grammar: A Basic Introduction*. Oxford: Oxford University Press.

Lanwer, Jan. 2017. Apposition: A multimodal construction? The multimodality of linguistic constructions in the light of usage-based theory. *Linguistics Vanguard* 3(s1): 1-12. https://doi.org/10.1515/lingvan-2016-0071.

Lazaraton, Anne. 2004. Gesture and speech in the vocabulary explanations of one ESL teacher: A microanalytic inquiry. *Language Learning* 54(1): 79-117. https://doi.org/10.1111/j.1467-9922.2004.00249.x.

Lenk, Uta. 1998. *Marking Discourse Coherence: Functions of Discourse Markers in Spoken English*. Tübingen: Gunter Narr Verlag.

Lyle, Andrew. 2019, October 31, 2019. Avoiding elder loneliness using artificial intelligence. *University of Alberta Faculty of Science Blog*. Retrieved from https://www.ualberta.ca/science/news/2019/october/avoiding-elder-loneliness-with-ai

Matlock, Teenie. 2011. The conceptual motivation of aspect. In *Motivation in Grammar and the Lexicon*, K. Panther and G. Radden (Eds.), 133-147. Amsterdam: John Benjamins.

Matlock, Teenie, Sparks, David, Matthews, Justin L., Hunter, Jeremy, and Huette, Stephanie. 2012. Smashing new results on aspectual framing. *Studies in Language* 36(3): 700-721. https://doi.org/10.1075/sl.36.3.09mat.

McClave, Evelyn. 2000. Linguistic functions of head movements in the context of speech. *Journal of Pragmatics* 32(2000): 855-878. https://doi.org/10.1016/S0378-2166(99)00079-X.

McNeill, David. 1992. *Hand and Mind: What Gestures Reveal about Thought*. Chicago: University of Chicago Press.

McNeill, David. 2005. *Gesture and Thought*. Chicago: University of Chicago Press.

Michaelis, Laura. 2010. Sign-based construction grammar. In *The Oxford Handbook of Linguistic Analysis*, B. Heine and H. Narrog (Eds.). Oxford: Oxford University Press.

Mittelberg, Irene. 2006. *Metaphor and metonymy in language and gesture: Discoursive evidence for multimodal models of grammar* (Doctoral dissertation). Cornell University, Cornell, USA.

Mittelberg, Irene. 2007. Methodology for multimodality: One way of working with speech and gesture data. In *Methods in Cognitive Linguistics*, M. Gonzalez-Marquez, I. Mittelberg, S. Coulson, and M. J. Spivey (Eds.), 225-248. Amsterdam: John Benjamins.

Mittelberg, Irene. 2008. Peircean semiotics meets conceptual metaphor: Iconic modes in gestural representations of grammar. In *Metaphor and Gesture*, A. Cienki and C. Müller (Eds.), 115-154. Amsterdam: John Benjamins.

Mittelberg, Irene. 2010. Geometric and image-schematic patterns in gesture space. In *Language, Cognition, and Space: The State of the Art and New Directions*, Vyvyan Evans and Paul Chilton (Eds.), 351-385. London: Equinox.

Mittelberg, Irene. 2013a. Balancing acts: Image schemas and force dynamics in pictures by Paul Klee and their gestural enactments. In *Language and the Creative Mind*, Mike Borkent, Barbara Dancygier, and Jennifer Hinnell (Eds.), 325-346. Stanford, CA: Centre for the Study of Language and Information.

Mittelberg, Irene. 2013b. The exbodied mind: Cognitive-semiotic principles as motivating forces in gesture. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 1), Cornelia Müller, Alan Cienki, Ellen Fricke, Silva H. Ladewig, David McNeill, and Stefanie Teßendorf (Eds.), 750-799. Berlin: De Gruyter Mouton.

Mittelberg, Irene. 2014. Gestures and iconicity. In *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction* (Vol. 2), C. Müller, A. Cienki, E. Fricke, S. H. Ladewig, D. Mcneill, and S. Teßendorf (Eds.), 1712-1732. Berlin: De Gruyter Mouton.

Mittelberg, Irene. 2018. Gestures as image schemas and force gestalts: A dynamic systems approach augmented with motion-capture data analyses. *Cognitive Semiotics*: 1-21. https://doi.org/10.1515/cogsem-2018-0002.

Mittelberg, Irene. 2019. Visuo-kinetic signs are inherently metonymic: How embodied metonymy motivates form, function, and schematic patterns in gesture. *Frontiers of Psychology*. https://doi.org/10.3389/fpsyg.2019.00254.

Morris, Desmond. 1994. *Bodytalk: A Worldguide to Gesture*. London: Jonathan Cape.

Morris, Desmond, Collett, Peter, Marsh, Peter, and O'Shaughnessy, Marie. 1979. *Gestures: Their Origins and Distribution*. London: Jonathan Cape.

Müller, Cornelia. 1998. *Redegebegleitende Gesten: Kulturgeschichte – Theorie – Sprachvergleich*. Berlin: Arno Spitz Verlag.

Müller, Cornelia. 2004. Forms and uses of the Palm Up Open Hand: A case of a gestural family. In *The Semantics and Pragmatics of Everyday Gestures*, Cornelia Müller and Roland Posner (Eds.), 233-256. Berlin: Weidler.

Müller, Cornelia. 2017. How recurrent gestures mean: Conventionalized contexts-of-use and embodied motivation. *Gesture* 16(2): 276-303. https://doi.org/10.1075/gest.16.2.05mul.

Müller, Cornelia, Bressem, Jana, and Ladewig, Silva H. 2013. Towards a grammar of gesture: A form-based view. In *Body – Language – Communication. An International Handbook on Multimodality in Human Interaction* (Vol. 1), 707-732. Berlin: De Gruyter Mouton.

Müller, Cornelia, Cienki, Alan, Fricke, Ellen, Ladewig, Silva H., McNeill, David, and Bressem, Jana (Eds.). 2014. *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 2 [Handbooks of Linguistics and Communication Science 38]. Berlin: De Gruyter Mouton.

Müller, Cornelia, Cienki, Alan, Fricke, Ellen, Ladewig, Silva H., McNeill, David, and Teßendorf, Stefanie (Eds.). 2013. *Body – Language – Communication: An International Handbook on Multimodality in Human Interaction*, Vol. 1 [Handbooks of Linguistics and Communication Science 38]. Berlin: De Gruyter Mouton.

Müller, Cornelia, and Ladewig, Silva H. 2013. Metaphors for sensorimotor experiences: Gestures as embodied and dynamic conceptualizations of balance in dance lessons. In *Language and the Creative Mind*, 278-307. Stanford, CA: Centre for the Study of Language and Information.

Munsch, Robert. 1998. *Andrew's Loose Tooth*. Markham, ON: Scholastic Canada.

Murphy, Lynne. 2006. Antonyms as lexical constructions: Or, why paradigmatic construction is not an oxymoron. *Constructions* Special volume 1(8): 1-37.

Newman, John, and Rice, Sally. 2006. Transitivity schemas of English EAT and DRINK in the BNC. In *Corpora in Cognitive Linguistics: Corpus-based Approaches to Syntax and Lexis*, Stefan Th. Gries and Anatol Stefanowitsch (Eds.), 225-260. Amsterdam: John Benjamins.

Núñez, Rafael E., and Cooperrider, Kensy. 2013. The tangle of space and time in human cognition. *Trends in Cognitive Sciences* 17(5): 220-229. https://doi.org/10.1016/j.tics.2013.03.008.

Núñez, Rafael E., and Sweetser, Eve. 2006. With the future behind them: Convergent evidence from Aymara language and gesture in the crosslinguistic comparison of spatial construals of time. *Cognitive Science* 30(5): 401–450. https://doi.org/10.1207/s15516709cog0000_62.

Oben, Bert, and Brône, Geert. 2015. What you see is what you do: On the relationship between gaze and gesture in multimodal alignment. *Language and Cognition* 7: 546-562. https://doi.org/10.1017/langcog.2015.22.

Ochs, Elinor, Schegloff, Emanuel A., and Thompson, Sandra A (Eds.). 1996. *Interaction and Grammar*, [Studies in Interactional Sociolinguistics 13. Cambridge: Cambridge University Press.

OED Online (Oxford English Dictionary). 2019 (September 2019 ed.). Oxford: Oxford University Press.

Ogden, Charles. 1932. *Opposition: A Linguistic and Psychological Analysis*. London: Paul, Trench, and Trubner.

Parrill, Fey. 2009. Dual viewpoint gestures. *Gesture* 9(3): 271-289. https://doi.org/10.1075/gest.9.3.01par.

Parrill, Fey. 2010. The hands are part of the package: Gesture, common ground and information packaging. In *Empirical and Experimental Methods in Cognitive/Functional Research*, Sally Rice and John Newman (Eds.), 285-302. Stanford: CSLI Publications.

Parrill, Fey, Bergen, Benjamin, and Lichtenstein, Patricia. 2013. Grammatical aspect in multimodal language production: Using gesture to reveal event representations. *Cognitive Linguistics* 24(1): 135-158.

Parrill, Fey, and Stec, Kashmiri. 2017. Gestures of the abstract: Do speakers use space consistently and contrastively when gesturing about abstract concepts? *Pragmatics and Cognition* 24(1): 33-61. https://doi.org/10.1075/pc.17006.par.

Pelkey, Jamin. 2017. *The Semiotics of X: Chiasmus, Cognition, and Extreme Body Memory*. London: Bloomsbury Publishing.

Perniss, Pamela , and Ozyürek, Asli. 2015. Visible cohesion: A comparison of reference tracking in sign, speech, and co-speech gesture. *Topics in Cognitive Science* 7: 36-60. https://doi.org/10.1111/tops.12122.

Polanyi, Livia, and Scha, R. J. H. 1983. The syntax of discourse. *Text* 3(3): 261-270. https://doi.org/10.1515/text.1.1983.3.3.261.

Precht, Kristen. 2000. *Patterns of stance in English* (Doctoral dissertation). Northern Arizona University, Flagstaff, USA.

Priesters, Matthias. 2012. *Functional Patterns in Gesture Space: Correlating location and function of gestures using 3D motion capture technology* (Unpublished master's thesis). RWTH Aachen, Aachen, Germany.

Redeker, Gisela. 2006. Discourse markers as attentional cues at discourse transitions. In *Approaches to Discourse Particles*, Kerstin Fischer (Ed.), 339-358. Amsterdem: Elsevier.

Rice, Sally. 1987. *Towards a cognitive model of transitivity* (Unpublished doctoral dissertation). University of California San Diego, San Diego, USA.

Rice, Sally. 2017a. Phraseology and polysynthesis. In *The Oxford Handbook of Polysynthesis*, Michael Fortescue, Marianne Mithun, and Nicholas Evans (Eds.), 203-214. Oxford.

Rice, Sally. 2017b. The study of Indigenous languages. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.), 38-58. Cambridge: Cambridge University Press.

Rice, Sally, and Hinnell, Jennifer. 2017. *'…which, by the way...': Multimodal marking of medial asides in North American English*. Paper presented at the 14th International Cognitive Linguistics Conference, Tartu, Estonia.

Rice, Sally, and Newman, John. 2005. *Inflectional islands*. Paper presented at the 9th International Cognitive Linguistics Conference, Yonsei University, Korea.

Rice, Sally, and Newman, John. 2008. *Beyond the lemma: Inflection-specific constructions in English*. Paper presented at the American Association for Corpus Linguistics, Provo, Utah.

Rice, Sally, and Newman, John. 2018. A corpus investigation of English cognition verbs and their effect on the incipient epistemization of physical activity verbs. *Russian Journal of Linguistics* 22(3): 560-580. https://doi.org/10.22363/2312-9182-2018-22-3-560-580.

Richardson, Daniel C., Spivey, Michael J., Barsalou, Lawrence W., and McRae, Ken. 2003. Spatial representations activated during real-time comprehension of verbs. *Cognitive Science* 27: 767-780. https://doi.org/10.1207/s15516709cog2705_4.

Sacks, Harvey, Schegloff, Emanuel A., and Jefferson, Gail. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50(4): 696-735. https://doi.org/10.1016/B978-0-12-623550-0.50008-2.

Sahakian, Barbara J., and Labuzetta, Jamie Nicole. 2013. *Bad Moves: How Decision Making Goes Wrong, and the Ethics of Smart Drugs*. Oxford, UK: Oxford University Press.

Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.

Schoonjans, Steven. 2014. *Modalpartikeln als multimodale Konstruktionen: Eine korpusbasierte Kookkurrenzanalyse von Modalpartikeln und Gestik im Deutschen* (Doctoral dissertation). KU Leuven, Leuven, Belgium.

Schoonjans, Steven, Feyaerts, Kurt, and Brône, Geert. 2013. *Analyzing German modal particles as multimodal constructions*. Paper presented at the Mapping Multimodal Dialogue (MaMuD), Aachen, Germany.

Seyfedinnipur, Mandana. 2006. *Disfluency: Interrupting Speech and Gesture* (Doctoral dissertation). Radboud University Nijmegen, Nijmegen, NL.

Shaffer, Barbara. 2012. Reported speech as an evidentiality stategy in American Sign Language. In *Viewpoint in Language: A Multimodal Perspective*, Barbara Dancygier and Eve Sweetser (Eds.), 139-155. Cambridge: Cambridge University Press.

Sidnell, Jack. 2010. *Conversation Analysis: An Introduction*. Chichester, UK: Wiley-Blackwell.

Sidnell, Jack, and Stivers, Tanya (Eds.). 2013. *The Handbook of Conversation Analysis*. Chichester, UK: Wiley Blackwell.

Sinclair, John 1991. *Corpus, Concordance, Collocation*. New York: Oxford University Press.

Stede, Manfred, and Schmitz, Birte. 2000. Discourse particles and discourse functions. *Machine Translation* 15: 125-147.

Steen, Francis, and Turner, Mark. 2013. Multimodal construction grammar. In *Language and the Creative Mind*, Mike Borkent, Barbara Dancygier, and Jennifer Hinnell (Eds.), 255-274. Stanford, CA: Centre for the Study of Language and Information.

Stefanowitsch, Anatol, and Gries, Stefan Th. 2003. Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8(2): 209-243.

Stefanowitsch, Anatol, and Gries, Stefan Th. 2005. Covarying collexemes. *Corpus Linguistics and Linguistic Theory* 1(1): 1-43.

Stickles, Elise. 2016. *The interaction of syntax and metaphor in gesture: A corpus-experimental approach* (Doctoral dissertation). UC Berkeley, Berkeley, USA.

Streeck, Jürgen. 2009. *Gesturecraft: The Manu-facture of Meaning.*

Sweetser, E. 1998, July 1997. *Regular metaphoricity in gesture: bodily-based models of speech interaction.* Paper presented at the 16th International Congress of Linguists, Paris.

Sweetser, Eve, and Stec, Kashmiri. 2013. Maintaining multiple viewpoints with gaze. In *Viewpoint and the Fabric of Meaning: Form and Use of Viewpoint Tools across Languages and Modalities*, B. Dancygier, Wei-lun Lu, and Arie Verhagen (Eds.). Berlin: DeGruyter Mouton.

Talmy, Leonard. 2000a. *Toward a Cognitive Semantics* (Vol. 1). Cambridge: MIT Press.

Talmy, Leonard. 2000b. *Toward a Cognitive Semantics* (Vol. 2). Cambridge: MIT Press.

Thompson, Sandra A, and Hopper, Paul J. 2001. Transitivity, clause structure, and argument structure: Evidence from conversation. In *Frequency and the Emergence of Linguistic Structure*, Joan Bybee and Paul J Hopper (Eds.), 27-60. Amsterdam: John Benjamins.

Traugott, Elizabeth C. 2014. On the function of the epistemic adverbs *Surely* and *No Doubt* at the left and right peripheries of the clause. In *Discourse Functions at the Left and Right Periphery: Crosslinguistic Investigations of Language Use and Language Change*, Kate Beeching and Ulrich Detges (Eds.), 72-91. Leiden: Brill.

Turner, Mark. 1991. *Reading Minds.* Princeton: Princeton University Press.

Turner, Mark. 2017. Opening commentary: Polytropos and communication in the wild. In *The Cambridge Handbook of Cognitive Linguistics*, Barbara Dancygier (Ed.). Cambridge, UK: Cambridge University Press.

Verspoor, Marjolijn, and Tyler, Andrea. 2009. Cognitive linguistics and second language learning. In *The New Handbook of Second Language Acquisition*, William C. Ritchie and Tej K. Bhatia (Eds.). Bingley, UK: Emerald Group.

Wehling, Elisabeth. 2010. *Argument is gesture war: Function, form and prosody of discourse structuring gestures in political argument.* Paper presented at the 35th Annual Meeting of the Berkeley Linguistics Society, Berkeley, California.

Wehling, Elisabeth. 2017. Discourse management gestures. *Gesture* 16(2): 245-275. https://doi.org/10.1075/gest.16.2.04weh.

Winston, Elizabeth A. 1995. Spatial mapping in comparative discourse frames. In *Language, Gesture and Space*, Karen Emmorey and Judy S. Reilly (Eds.), 87-114. Hillsdale, NJ: Lawrence Erlbaum Associates.

Winston, Elizabeth A. 1996. *Spatial Mapping in ASL Discourse.* Paper presented at the CIT 11th National Convention, Little Rock, Arkansas.

Winter, Bodo, Perlman, Marcus, and Matlock, Teenie. 2013. Using space to talk and gesture about numbers: Evidence from the TV News Archive. *Gesture* 13(3): 377-408. https://doi.org/10.1075/gest.13.3.06win.

Woodin, Greg. 2018. *Degrees of metaphoricity: A quantitative gesture analysis* (MA thesis). University of Birmingham, Birmingham, UK.

Wulf, Alyssa, and Dudis, Paul G. 2005. Body partitioning in ASL metaphorical blends. *Sign Language Studies* 5(3): 317-332.

Zima, Elisabeth. 2017a. Multimodal constructional resemblance: The case of English circular motion constructions. In *Constructing Families of Constructions*, F. Ruiz De Mendoza, A. Luzondo, and P. Pérez-Sobrino (Eds.), 301-337. Amsterdam: John Benjamins.

Zima, Elisabeth. 2017b. On the multimodality of [all the way from X PREP Y]. *Linguistics Vanguard [Special Issue]: Towards a Multimodal Construction Grammar* 3(s1). https://doi.org/10.1515/lingvan-2016-0055.

Zima, Elisabeth, and Bergs, Alexander. 2017a. Introduction: Multimodality and construction grammar. *Linguistics Vanguard [Special Issue]: Towards a Multimodal Construction Grammar* 3(s1). https://doi.org/10.1515/lingvan-2016-1006.

Zima, Elisabeth, and Bergs, Alexander. 2017b. Towards a Multimodal Construction Grammar. *Linguistics Vanguard [Special Issue]* 3(s1).

Zyp, Jasmine. 2018. *'Which is great': The multimodal performance of a parenthetical construction* (Undergraduate honour's thesis). University of Alberta, Edmonton, Canada.

APPENDIX A: Video clips and corpus metadata

A1. Instructions for viewing videos clips

Video clips for each screenshot from the Red Hen archive can be found in the following Google Drive folder:

https://drive.google.com/open?id=1VHCijHu6Ku6Dn8lOK0hoLmU-iKzELkqa

If you cannot access the folder via this link, please contact me directly (see www.jenniferhinnell.com for up to date contact information).

Clips in the folder are organized by chapter and figure number. Each filename includes the figure number followed by the metadata and a timestamp (which follows the comma), as follows:

Fig2.4-2015-09-28_1300_US_CNN_Newsroom,1330.

A B C

A: Figure number in dissertation (here, Figure 2.4 in Chapter 2).

B: Date (2015-09-28), time (1300 in 24 hour clock), network (CNN), program (Newsroom).

C: Timestamp in seconds from onset of program.

A2. Metadata for Red Hen examples

Metadata for the Red Hen examples (given by chapter in figures as stated above) can be found listed in the Excel spreadsheet in the same folder. The spreadsheet lists figure number, metadata, timestamp, and permalink to the file on Red Hen.

APPENDIX B: Red Hen program selections and settings

B1. Red Hen network and series selections

Networks	Series
CNBC*	(any series)
CNN	Vice-presidential and Presidential debates, Larry King, Paula Zahn Now, Piers Morgan Tonight
CSPAN	Tony Snow Briefing, Tony Snow Return
Comedy Central	Colbert Report, Daily Show
Current	Young Turks With Cenk Uygur
ESPN*	(any series)
Fox-news	Fox and Friends, Hannity, Hannity and Colmes, Huckabee, The OReilly Factor
KABC	White House interviews, Barbara Walters Interviews, Live with Regis and Kelly, Michael J Fox Parkinsons, Oprah Winfrey, The View, This week with George Stephanopoulos, Ebert and Roper, Good morning America, Katie Couric, The Doctor Oz Show
KCBS	Late Late Show with Craig Ferguson, Late Late Show with James Corden, Late Show with Dave Letterman, Late Show with Stephen Colbert
KCET	Moyers, Newshour with Jim Lehrer
KNBC	Hardball with Chris Matthews, Meet the Press, The Ellen Degeneres Show, The Jay Leno Show, Today Show, Tonight Show with Conan Obrien, Tonight Show with Jay Leno, Late Night with Conan Obrien, Late Night with Jimmy Fallon, The Chris Matthews Show, The Megan Mullally Show, The Rachel Maddow Show
KOCE*	(any series)
KTLA*	(any series)
MSNBC*	(any series)

* Discourse chapter only

B2. Sample Red Hen search interface