ILLOCUTION ON TWITTER: THE CONSTRUCTION AND ANALYSIS OF A SOCIAL MEDIA SPEECH ACT CORPUS

A Dissertation
submitted to the Faculty of the
Graduate School of Arts and Sciences
of Georgetown University
in partial fulfillment of the requirements for the
degree of
Doctor of Philosophy
in Linguistics

By
Laura Beth Bell, M.S.

Washington, DC
December 6, 2019
ILLOCUTION ON TWITTER: THE CONSTRUCTION AND ANALYSIS OF A SOCIAL MEDIA SPEECH ACT CORPUS

Laura Beth Bell, M.S.

Thesis Advisor: Paul Portner, Ph.D.

ABSTRACT

In the years since J.L. Austin (1962) first proposed Speech Act Theory (SAT), the literature has taken it in many directions. One recurring point of discussion is the extent to which the direct illocutionary force of an utterance is overtly encoded (Searle & Vanderveken, 1985; Bach & Harnish, 1979; Kissine, 2009; among others). A related question, which has received notably less attention, is the exact identity of the elements proposed to be responsible for such encoding. This dissertation began with the goal of conducting an empirical investigation into the existence and identity of these illocutionary force indicating devices (IFIDs, Searle & Vanderveken, 1985).

This investigation was realized through the construction and analysis of a corpus of posts from the social media platform Twitter. The data was annotated, using Amazon Mechanical Turk, for a variety of features including direct illocutionary act category, indirect illocutionary act presence and category, and hashtag functionality. The annotated corpus was analyzed using logistic regression and random forest methodologies, with direct act category as the response variable, in an attempt to identify potential IFIDs. Additional hypotheses that arose during the course of planning the project were also assessed. For example, the corpus was utilized to investigate stylistic accommodation on Twitter as a secondary goal.

The corpus analysis identified several features as having a significant effect on the model, though no features were identified by the model as individually significant for direct
illocutionary act classification. The dissertation proposes that IFIDs may be realized as templates of typical features as opposed to categorical entities. The analysis also identified several stylistic features to which Twitter users appear to accommodate over time. The dissertation proposes that several of these features may be explained by a style that promotes authenticity and is uncertain about its own level of formality, both attributes of Twitter attested in the literature.
ACKNOWLEDGEMENTS

I could not decide where to start. Do you start with the greatest contributors, to let everyone know that, “These guys deserve top billing”? Or do you save them for last, letting the weight and suspense build until the final reveal?

I ultimately went for neither, as I do not like the idea of ranking the support of family and friends anyway (what is this, MySpace?). I therefore now acknowledge the contributions of the following parties, in an intentionally randomized order.

My sincerest thanks:

To my cats, for emotional healing and excellent tuition as regards proper stepwise feature dropping.

To my committee members, for your crucial support and guidance.

To Meredith, for always reminding me of what is most important in life and demonstrating true dedication to hygge.

To Sarah, for sending me so many encouraging phone calls, texts, and care packages that it never actually felt like you lived 3,000 miles away.

To my pilot annotators, for your honest and constructive feedback.

To my advisor Paul, for encouraging me to try something new and being patient as I blazed that trail.

To my brother Scott, for pursuing a more stable line of work so that Mom only has to worry about one of us.

To Stacy and Shannon, for escaping with me into the land of André and Hallmark Christmas Movies when I just couldn’t look at Twitter anymore.
To my spouse Scott, for daily pep talks, unwavering faith, and a level of support that can only be classified as monumental.

To the ladies of my D&D group, for the bonus inspiration die and an additional +1 to constitution checks.

To my parents, for the innumerable opportunities that you have made possible with your constant love, support, and generosity.

To Babby, for waiting until after the defense to let me know you were there.

To Russ and Merrilee, for defying every “in-law” stereotype possible and truly loving and supporting me as one of your own.

To my anonymous annotators, for your critical participation and good-faith efforts.

To all members of the Ryals, Kuchta, and Bell crews, for giving me the sort of love and security that most people can only dream of.

And to Dan, for introducing me to Scott and letting us get cats even though you’re allergic.

The completion of this work is dedicated to the memory of Kim. You were a paragon of empathy and groundedness. I feel lucky to have benefited from your mentorship for as long as I did.
# TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION AND BACKGROUND ..............................................1

1.0 Introduction ..............................................................................................1

1.0.1 Outline and Key Findings .................................................................5

1.1 Overview: Speech Act Theory ...............................................................6

1.1.1 Austin (1962) .....................................................................................7

1.1.2 The Family Tree...............................................................................10

1.1.2.1 Literalist Accounts of Illocutionary Force ...................................11

1.1.2.1.1 Literalist Mentalist Accounts: Speaker Only ................................11

1.1.2.1.2 Literalist Mentalist Accounts: Speaker and Hearer ..................15

1.1.2.1.3 Literalist Non-Mentalist Accounts ...........................................20

1.1.2.2 Contextualist Accounts of Illocutionary Force .............................24

1.1.2.2.1 Contextualist Accounts: Reduced Role of the Proposition......24

1.1.2.2.2 Contextualist Accounts: No Role for Proposition .....................27

1.1.2.3 Summary Discussion ...................................................................29

1.1.3 Direct Acts vs. Indirect Acts .............................................................39

1.1.4 Summary ...........................................................................................44

1.1.4.1 Definitions of Key Terms ...............................................................45

1.1.4.2 Motivation for Hypotheses 1 and 2 .............................................46

1.1.4.3 Motivation for Hypothesis 3 ..........................................................47

1.2 Overview: Technological Affordances of the Medium ..........................48

1.2.1 Emoticons and Emoji ......................................................................50
1.2.1.1 Background .............................................................................................................50

1.2.1.1.1 Affecting Interpretation of Adjacent Text .........................................................50

1.2.1.1.2 Standing in for Words/Concepts/Phrases .........................................................55

1.2.1.2 Speech Act Status ..................................................................................................56

1.2.2 Hashtags and @-tags .................................................................................................58

1.2.2.1 Background ............................................................................................................58

1.2.2.1.1 Tagging Usages ..................................................................................................59

1.2.2.1.2 Non-tagging Usages ..........................................................................................64

1.2.2.2 Speech Act Status ..................................................................................................69

1.2.3 Quotations ...................................................................................................................74

1.2.3.1 Background ............................................................................................................74

1.2.3.1.1 The Name Theory .............................................................................................78

1.2.3.1.2 The Demonstrative Theory ..............................................................................79

1.2.3.1.3 The Identity Theory ..........................................................................................81

1.2.3.2 Speech Act Status ..................................................................................................83

1.2.4 Likes and Retweets .....................................................................................................85

1.2.4.1 Background ............................................................................................................85

1.2.4.1.1 Posting Functions of Likes and Retweets ..............................................................85

1.2.4.1.2 Linguistic Function of Likes ...............................................................................87

1.2.4.1.3 Linguistic Function of Retweets .........................................................................88

1.2.4.2 Speech Act Status ..................................................................................................90

1.3 Summary: The Annotation Task So Far ......................................................................90
CHAPTER 2: ANNOTATION INFORMATION AND GUIDELINES .................................. 94

2.0 Introduction ........................................................................................................ 94

2.1 Annotation Guidelines ...................................................................................... 94

2.1.1 Theoretical Motivation .................................................................................. 96

2.1.2 Comparability Motivation ............................................................................. 98

2.1.2.1 Speech Act Studies in Computer-Mediated Communication .................. 99

2.1.2.1.1 Twitter ............................................................................................... 99

2.1.2.1.2 Facebook .......................................................................................... 102

2.1.2.1.3 Online Discussion Forums .................................................................. 103

2.1.2.1.4 Instant Messaging ............................................................................. 107

2.1.2.1.5 Email ................................................................................................. 109

2.1.2.2 Speech Act Studies in Spoken Conversation .......................................... 110

2.1.2.3 Summary ................................................................................................. 112

2.1.3 Accessibility Motivation .............................................................................. 114

2.1.3.1 Taxonomy .............................................................................................. 114

2.1.3.2 Category Definitions .............................................................................. 115

2.1.3.3 Key Concept Definitions ....................................................................... 116

2.1.4 Summary ........................................................................................................ 119

2.2 Segmentation Guidelines and Final Form of the Task .................................... 120

2.2.1 Speech Act Units in Previous Literature ..................................................... 120

2.2.2 Segmentation Guidelines ............................................................................ 125

2.2.2.1 Motivating Segmentation Guidelines .................................................... 125
2.2.2.1.1 Full Sentences ........................................................................................................125
2.2.2.1.2 Sentential Fragments ............................................................................................125
2.2.2.1.3 Conjoined and/or Coordinated Elements ............................................................126
2.2.2.1.4 Complement Clauses ...........................................................................................128
2.2.2.1.5 Subordinate Clauses and Restrictive Relative Clauses .........................................129
2.2.2.1.6 Non-restrictive Relative Clauses and Other Parenthetical Elements ...................131
2.2.2.1.7 Speech Act Units Interrupting Other Speech Act Units .......................................133
2.2.2.1.8 Recursion Guideline ............................................................................................134
2.2.2.1.9 Discourse Particles .............................................................................................135
2.2.2.1.10 Interjections ........................................................................................................137
2.2.2.1.11 Hashtags and @-tags .........................................................................................137
2.2.2.1.12 Direct Address ....................................................................................................139
2.2.2.1.13 Greetings, Goodbyes, and Thank Yous ..............................................................141
2.2.2.1.14 Emoji/Emoticons ..............................................................................................141
2.2.2.1.15 Laughter Particles ..............................................................................................142
2.2.2.1.16 Quotations ..........................................................................................................142
2.2.2.2 Summary of Guidelines and Sample Segmented Tweets ........................................144
2.2.3 Summary of Segment Types and Final Form of the Task ........................................147
2.3 Guideline Revision and Motivation for Hypothesis 4 ..................................................152
2.4 Motivation for Hypothesis 5 ..........................................................................................153
CHAPTER 3: CONSTRUCTING THE CORPUS .......................................................................159
3.0 Introduction ..................................................................................................................159
4.1.4 Influence of the Agreement Results on the Analysis ........................................... 221

4.2 Analysis and Discussion by Hypothesis ............................................................... 224

4.2.1 Analysis and Discussion of Hypothesis 1 ......................................................... 224

4.2.1.1 Analysis ........................................................................................................... 225

4.2.1.1.1 Explanatory Variables ............................................................................... 227

4.2.1.1.2 Model Selection ......................................................................................... 237

4.2.1.1.3 Regression Analyses ................................................................................. 240

4.2.1.1.4 Random Forest Analyses ........................................................................ 242

4.2.1.1.5 Summary of Repeat Variables .................................................................. 245

4.2.1.2 Discussion ...................................................................................................... 248

4.2.1.2.1 Variable-specific Hypotheses and Results ................................................. 248

4.2.1.2.2 Summary and Interpretation .................................................................... 290

4.2.2 Analysis and Discussion of Hypothesis 2 ......................................................... 295

4.2.2.1 Analysis ........................................................................................................ 296

4.2.2.1.1 Explanatory Variables ............................................................................. 297

4.2.2.1.2 Regression Analyses ............................................................................... 300

4.2.2.1.3 Random Forest Analyses ........................................................................ 301

4.2.2.1.4 Summary of Repeat Variables .................................................................. 303

4.2.2.2 Discussion ...................................................................................................... 305

4.2.2.2.1 Variable-specific Hypotheses and Results ................................................. 306

4.2.2.2.2 Summary .................................................................................................. 331

4.2.3 Analysis and Discussion of Hypothesis 3 ......................................................... 335
LIST OF FIGURES

Figure 1.1. Speech act theory family tree .................................................................7
Figure 1.2. Sample hashtag game ...........................................................................63
Figure 1.3. Tag flow chart .......................................................................................69
Figure 1.4. Sample hashtag trend ...........................................................................73
Figure 1.5. Tweet featuring quote with @-tag citation .............................................75
Figure 1.6. Tweet featuring quote with dash citation ..............................................75
Figure 1.7. Tweet featuring quote with no citation ...................................................76
Figure 1.8. Tweet featuring mixed quotation .............................................................76
Figure 1.9. Tweet featuring recursive quotation .......................................................76
Figure 2.1. Sample tweet #1 ..................................................................................145
Figure 2.2. Sample tweet #2 ..................................................................................146
Figure 2.3. Sample tweet #3 ..................................................................................146
Figure 2.4. Sample tweet #4 ..................................................................................146
Figure 2.5. Sample tweet #5 ..................................................................................147
Figure 2.6. Sample tweet #6 ..................................................................................147
Figure 3.1. Annotation task welcome block .............................................................166
Figure 3.2. Annotation task demographic block .......................................................167
Figure 3.3. Annotation task reference block ............................................................171
Figure 3.4. Percentage of attention and quality control questions of each speech act category ..173
Figure 3.5. Annotation task thank you block ............................................................175
Figure 3.6. Annotation task failure message ............................................................175
Figure 3.7. Percentage of segments with potential indirect acts that each Turker coded as having no indirect act ................................................................. 180
Figure 3.8. Segments with no indirect acts by most frequent Turkers ........................................... 181
Figure 3.9. Percentage of human intelligence tasks with X number of respondents ............ 184
Figure 3.10. Number of respondents for 10 human intelligence tasks that received 10 or more respondents ........................................................................................................ 185
Figure 3.11. Direct force for non-backgrounded non-hashtag segments ....................... 191
Figure 3.12. Direct force for non-backgrounded non-hashtag segments in majority annotation version of the corpus ........................................................................................................ 192
Figure 3.13. Conversation #590 ............................................................................................... 193
Figure 3.14. Indirect act response totals ..................................................................................... 196
Figure 3.15. Indirect act response totals except “none” ................................................................. 196
Figure 3.16. Indirect act response totals in majority annotation version of the corpus .......... 197
Figure 3.17. Indirect act response totals except “none” in majority annotation version of the corpus ........................................................................................................ 198
Figure 3.18. Responses to “Does this hashtag perform an emotive or meta-commentative act? If so, what kind?” ........................................................................................................ 200
Figure 3.19. Responses to “Does this hashtag perform an emotive or meta-commentative act? If so, what kind?” for the majority annotation version of the corpus ................................................. 200
Figure 3.20. Responses to “Is this a hashtag that would actually be used?” ...................... 201
Figure 3.21. Responses to “Is this a hashtag that would actually be used?” for the majority annotation version of the corpus ........................................................................................................ 201
Figure 4.1. Calculation for percent chance perfect agreement ........................................205

Figure 4.2. Calculation for percent agreement .............................................................205

Figure 4.3. Calculation for percent perfect agreement ....................................................205

Figure 4.4. Association for Tweet length and Segment position, data subset #1 ............247

Figure 4.5. Association for Direct act category and Punctuation, data subset #1 ...........250

Figure 4.6. Association for Direct act category and Segment position, data subset #1 .......252

Figure 4.7. Association for Direct act category and Wh-word fraction (batched), data subset #1 ..................................................................................................................253

Figure 4.8. Association for Direct act category and Verb fraction (batched), data subset #1 ..................................................................................................................254

Figure 4.9. Association for Direct act category and Indirect act choice, data subset #1 ......256

Figure 4.10. Association for Direct act category and Serious yes/no, data subset #1 ........258

Figure 4.11. Association for Direct act category and “Ask” fraction, data subset #1 ..........260

Figure 4.12. Association for Direct act category and “Tell” fraction, data subset #1 ............262

Figure 4.13. Association for Direct act category and Proper noun fraction (batched), data subset #1 ..................................................................................................................263

Figure 4.14. Association for Direct act category and Numeral fraction (batched), data subset #1 ..................................................................................................................264

Figure 4.15. Association for Direct act category and Nominal verb contraction fraction (batched), data subset #1 ..................................................................................................................265

Figure 4.16. Association for Direct act category and Determiner fraction (batched), data subset #1 ..................................................................................................................266
Figure 4.17. Association for Direct act category and Segment length (batched), data subset #1…………………………………………………………………………………………267

Figure 4.18. Association for Direct act category and Modal fraction (batched), data subset #1…………………………………………………………………………………………268

Figure 4.19. Association for Direct act category and Adjective fraction (batched), data subset #1…………………………………………………………………………………………270

Figure 4.20. Association for Direct act category and Vulgar word fraction (batched), data subset #1…………………………………………………………………………………………271

Figure 4.21. Association for Direct act category and Preposition fraction (batched), data subset #1…………………………………………………………………………………………272

Figure 4.22. Association for Direct act category and “Give” fraction, data subset #1 ………273

Figure 4.23. Association for Direct act category and Common noun fraction (batched), data subset #1…………………………………………………………………………………………274

Figure 4.24. Association for Direct act category and Adverb fraction (batched), data subset #1…………………………………………………………………………………………275

Figure 4.25. Association for Direct act category and Original/reply status, data subset #1 ………276

Figure 4.26. Association for Direct act category and Coordinate conjunction fraction (batched), data subset #1…………………………………………………………………………………………277

Figure 4.27. Association for Direct act category and Emoji in tweet fraction (batched), data subset #1…………………………………………………………………………………………279

Figure 4.28. Association for Direct act category and Interjection fraction (batched), data subset #1…………………………………………………………………………………………280
Figure 4.29. Association for Direct act category and Tweet length, data subset #1 ………..281
Figure 4.30. Association for Direct act category and Emotion word fraction (batched), data
subset #1……………………………………………………………………………………………………288
Figure 4.31. Association for Direct act category and Incorporated emoji, data subset #1 ………289
Figure 4.32. Association for Direct act category and Direct act serious yes/no, data
subset #1……………………………………………………………………………………………………304
Figure 4.33. Association for Hashtag in tweet and Non-incorporated hashtag adjacent, data
subset #1……………………………………………………………………………………………………305
Figure 4.34. Association for Indirect act yes/no and Direct act category, data subset #1 ………308
Figure 4.35. Association for Indirect act yes/no and Proper noun fraction (batched), data
subset #1……………………………………………………………………………………………………309
Figure 4.36. Association for Indirect act yes/no and Original/reply status, data subset #1……..310
Figure 4.37. Association for Indirect act yes/no and Punctuation, data subset #1………………311
Figure 4.38. Association for Indirect act yes/no and Segment has incorporated emoji, data
subset #1……………………………………………………………………………………………………312
Figure 4.39. Association for Indirect act yes/no and Segment position, data subset #1………..313
Figure 4.40. Association for Indirect act yes/no and Direct act serious yes/no, data
subset #1……………………………………………………………………………………………………314
Figure 4.41. Association for Indirect act yes/no and Segment has tone-type emoji adjacent, data
subset #1……………………………………………………………………………………………………315
Figure 4.42. Association for Indirect act yes/no and Common noun fraction (batched), data
subset #1……………………………………………………………………………………………………316
Figure 4.43. Association for Indirect act yes/no and Segment length (batched), data subset #1.........................................................................................................................317

Figure 4.44. Association for Indirect act yes/no and @-tag in tweet, data subset #1 ..............318

Figure 4.45. Association for Indirect act yes/no and Adjective fraction (batched), data subset #1........................................................................................................................................319

Figure 4.46. Association for Indirect act yes/no and Adverb fraction (batched), data subset #1........................................................................................................................................321

Figure 4.47. Association for Indirect act yes/no and Emotion word fraction (batched), data subset #1 ........................................................................................................................................322

Figure 4.48. Association for Indirect act yes/no and Vulgar word fraction (batched), data subset #1........................................................................................................................................323

Figure 4.49. Association for Indirect act yes/no and Interjection fraction (batched), data subset #1........................................................................................................................................324

Figure 4.50. Association for Expressive direct/indirect and Adjective fraction (batched), data subset #1........................................................................................................................................325

Figure 4.51. Association for Expressive direct/indirect and Vulgar word fraction (batched), data subset #1 ........................................................................................................................................326

Figure 4.52. Association for Expressive direct/indirect and Emotion word fraction (batched), data subset #1 ........................................................................................................................................327

Figure 4.53. Association for Expressive direct/indirect and Adverb fraction (batched), data subset #1 ........................................................................................................................................328
Figure 4.54. Association for Expressive direct/indirect and Interjection fraction (batched), data subset #1 ..........................................................329

Figure 4.55. Association for Expressive direct/indirect and Segment has incorporated emoji, data subset #1 ..........................................................330

Figure 4.56. Percentage of direct acts of each category, full version of corpus ...............338

Figure 4.57. Percentage of direct acts of each category, majority-annotation version of corpus ..........................................................339

Figure 4.58. Percentage of highest acts of each category, full version of corpus ............340

Figure 4.59. Percentage of highest acts of each category, majority-annotation version of corpus ..........................................................341

Figure 4.60. Percentage of all non-automatic acts of each category, full version of corpus .....342

Figure 4.61. Percentage of all non-automatic acts of each category, majority-annotation version of corpus ..........................................................343

Figure 4.62. Percentage of non-serious direct acts of each category, full version of corpus .....344

Figure 4.63. Percentage of non-serious direct acts of each category, majority-annotation version of corpus ..........................................................345

Figure 4.64. Event topic type Vosoughi & Roy (2016) ........................................347

Figure 4.65. Entity topic type Vosoughi & Roy (2016) ........................................348

Figure 4.66. Long-standing topic type Vosoughi & Roy (2016) ..............................349

Figure 4.67. Averages across topic types Vosoughi & Roy (2016) ............................350

Figure 4.68. Event topic type Zhang et al. (2011, 2012) .......................................351

Figure 4.69. Entity topic type Zhang et al. (2011, 2012) .......................................352
Figure 4.70. Long-standing topic type Zhang et al. (2011, 2012) ...........................................353
Figure 4.71. Averages across topic types Zhang et al. (2011, 2012) .................................354
Figure 4.72. Frequency of speech acts in Carr et al. (2012) ..................................................355
Figure 4.73. Speech act distributions in Qadir & Riloff (2011) ........................................357
Figure 4.74. Frequency of speech acts in Nastri et al. (2006) ........................................359
Figure 4.75. Speech act distributions in Jurafsky et al. (1997) ........................................361
Figure 4.76. Speech act distributions in Shriberg et al. (2004) ........................................362
Figure 4.77. Percentage assertive, by paper/subset and medium ........................................367
Figure 4.78. Percentage commissive, by paper/subset and medium .....................................368
Figure 4.79. Percentage directive, by paper/subset and medium ..........................................369
Figure 4.80. Percentage exercitive, by paper/subset and medium .........................................370
Figure 4.81. Percentage expressive, by paper/subset and medium ........................................371
Figure 4.82. Percentage interrogative, by paper/subset and medium .....................................372
Figure 4.83. Tweet from @realDonaldTrump .................................................................379
Figure 4.84. Response tweet to @realDonaldTrump ..........................................................379
Figure 4.85. Association for Emotive/meta-commentative yes/no and Incorporation, majority-annotation version of corpus .................................................................382
Figure 4.86. Association for Emotive/meta-commentative yes/no and Incorporation, full version of corpus ..............................................................383
Figure 4.87. Association for Emotive/meta-commentative choice and Incorporation, full version of corpus ..............................................................384
Figure 4.88. Association for Contribution and Incorporation, full version of corpus ...............385
Figure 4.89. Association for Fraction interrogative direct act and Incorporation, full version of corpus………………………………………………………………………………………………………386

Figure 4.90. Association for Segment position and Incorporation, majority-annotation version of corpus………………………………………………………………………………………………………388

Figure 4.91. Distribution of direct act category over time, batches by number of observations ………………………………………………………………………………………………………………………………………397

Figure 4.92. Distribution of direct act category over total tweets sent, batches by number of observations ………………………………………………………………………………………………………………………………………398

Figure 4.93. Likelihood that hashtag is affiliating over time on Twitter ………..400

Figure 4.94. Likelihood that hashtag is affiliating over total tweets sent ………..401

Figure 4.95. Likelihood that hashtag is affiliating over total tweets sent, x>75,000 excluded ……..402

Figure 4.96. Likelihood that hashtag is contributing over time on Twitter …………..403

Figure 4.97. Likelihood that hashtag is contributing over total tweets sent, x>75,000 excluded ………………………………………………………………………………………………………………………………………404

Figure 4.98. Likelihood that hashtag is contributing/affiliating over time on Twitter ………….405

Figure 4.99. Likelihood that hashtag is contributing/affiliating over total tweets sent, x>75,000 excluded ………………………………………………………………………………………………………………………………………406

Figure 4.100. Likelihood that hashtag is emotive/meta-commentative over time on Twitter ….407

Figure 4.101. Likelihood that hashtag is emotive/meta-commentative over total tweets sent, x>75,000 excluded ………………………………………………………………………………………………………………………………………408

Figure 4.102. Hashtag emotive/meta-commentative distribution over time on Twitter …………..409

Figure 4.103. Hashtag emotive/meta-commentative distribution over total tweets sent ………410
Figure 4.104. Number of hashtags over time on Twitter .................................................411
Figure 4.105. Number of hashtags over total tweets sent ............................................412
Figure 4.106. Number of hashtags over total tweets sent, 23 outliers removed ..........413
Figure 4.107. Number of @-tags over time on Twitter ................................................414
Figure 4.108. Number of @-tags over total tweets sent, 23 outliers removed ..........415
Figure 4.109. Tweet length over time on Twitter .........................................................416
Figure 4.110. Tweet length over time on Twitter, y>10 excluded .................................417
Figure 4.111. Tweet length over total tweets sent, y>10 and x>150,000 excluded ....418
Figure 4.112. Segment length over time on Twitter ....................................................419
Figure 4.113. Segment length over total tweets sent, x>150,000 excluded ...............420
Figure 4.114. Presence of tweet-final punctuation over time ......................................421
Figure 4.115. Presence of tweet-final punctuation over total tweets sent, x>150,000 excluded .................................................................421
Figure 4.116. Standard capitalization over time on Twitter .......................................422
Figure 4.117. Standard capitalization over total tweet sent, x>150,000 excluded ........423
Figure 4.118. Number of vulgar words over time on Twitter ....................................424
Figure 4.119. Vulgar words count over total tweets sent, y>2 and x>150,000 excluded ....424
Figure 4.120. Third person pronoun count over time on Twitter ...............................425
Figure 4.121. Third person pronoun count over total tweets sent, x>150,000 excluded ....426
Figure 4.122. Total count of personal pronouns over time on Twitter .....................427
Figure 4.123. Total count of personal pronouns over total tweets sent, x>150,000 excluded ....427
Figure 4.124. Number of unique repliers over number of followers ......................431
Figure 4.125. Number of likes over number of followers ........................................432
Figure 4.126. Number of retweets over number of followers.................................433
Figure 4.127. Number of unique repliers over number of followers, x>2,500 (10% of data) excluded .................................................................434
Figure 4.128. Number of likes over number of followers, x>2,500 (10% of data) excluded ..........435
Figure 4.129. Number of retweets over number of followers, x>2,500 (10% of data) excluded .................................................................436
Figure 4.130. Number of unique repliers over number of followers, x>67 (90% of data) excluded .................................................................437
Figure 4.131. Number of likes over number of followers, x>67 (90% of data) excluded ......438
Figure 4.132. Number of retweets over number of followers, x>67 (90% of data) excluded ....439
Figure 4.133. Association for Fraction of possible indirect acts that occur and Number of unique repliers ........................................................................................................443
Figure 4.134. Association for Fraction of direct acts that are interrogative and Number of unique repliers ........................................................................................................445
Figure 4.135. Association for Fraction of direct acts that are expressive and Number of unique repliers ........................................................................................................446
Figure 4.136. Association for Total number of vulgar words and Total number of likes ........448
Figure 4.137. Association for Total number of hashtags and Total number of retweets .........449
Figure 4.138. Association for Total number of vulgar words and Total number of responses ...450
Figure 4.139. Association for Total number of third person pronouns and @-tagged user response yes/no ........................................................................................................452
Figure 4.140. Association for Total number of personal pronouns and @-tagged user response
yes/no .................................................................452

Figure A.1. Annotation task introduction block .................................................470

Figure A.2. Task speech act category explanation .............................................471

Figure A.3. Task direct and indirect explanation ..................................................472

Figure A.4. Task direct and indirect example ......................................................473

Figure A.5. Task serious and non-serious direct act explanation .........................474

Figure A.6. Task literal and non-literal direct act explanation ..............................475

Figure A.7. Task backgrounded explanation ......................................................476

Figure A.8. Task final notes .................................................................................477

Figure A.9 Task sample annotation tweet introduction ....................................478

Figure A.10. Task sample tweets #1-2 ...............................................................479

Figure A.11. Tasks sample tweets #3-4 .............................................................480

Figure A.12. Excerpt from conversation #307 ..................................................481

Figure A.13. Conversation #322 ......................................................................482

Figure A.14. Excerpt from conversation #326 ..................................................483

Figure A.15. Conversation #378 ......................................................................484

Figure A.16. Conversation #430 ......................................................................485

Figure A.17. Conversation #431 ......................................................................485
## LIST OF TABLES

Table 1.1. Direction of fit definitions from Searle & Vanderveken (1985) .........................13

Table 1.2. I-Rules from Alston (2000) ........................................................................31

Table 1.3. Findings from Clark (1979) ........................................................................41

Table 1.4. Hashtag tagging sub-types ...........................................................................70

Table 1.5. @-tag functionality .......................................................................................73

Table 1.6. Hashtag functionality ...................................................................................74

Table 2.1. I-Rules from Alston (2000) ........................................................................97

Table 2.2. Speech act category definitions ...................................................................116

Table 3.1. Total number of segments in the corpus .......................................................163

Table 3.2. Annotation task hover-text definitions .........................................................169

Table 3.3. Updated total number of segments in the corpus .........................................187

Table 4.1. Agreement scores for “Is this segment backgrounded?” ...............................206

Table 4.2. Agreement scores for direct act category .....................................................207

Table 4.3. Agreement scores for direct act seriousness ................................................208

Table 4.4. Agreement scores for indirect act presence ..................................................209

Table 4.5. Agreement scores for indirect act category ..................................................209

Table 4.6. Agreement scores for act importance to overall message ............................210

Table 4.7. All agreement scores from section 4.1.1 .......................................................213

Table 4.8. Agreement scores for @-tag threading .........................................................214

Table 4.9. Agreement scores for @-tag direct address functionality .............................215

Table 4.10. All agreement scores from section 4.1.2 .....................................................216
Table 4.11. Agreement scores for hashtag usage .........................................................217
Table 4.12. Agreement scores for hashtag affiliation functionality ...............................217
Table 4.13. Agreement scores for hashtag contribution functionality ............................218
Table 4.14. Agreement scores for hashtag evaluative functionality ...............................219
Table 4.15. Agreement scores for evaluative act category .............................................219
Table 4.16. All agreement scores from section 4.1.3 .....................................................221
Table 4.17. Bigrams for direct act analysis .....................................................................239
Table 4.18. Variables that appeared in the final model of more than one regression ........240
Table 4.19. Classification accuracy of final regression models for test portions of data ....241
Table 4.20. Category-specific classification accuracy of final regression models for test portions of data ........................................................................................................241
Table 4.21. Variables that appeared in the top 5 for more than one random forest final model .........................................................................................................................243
Table 4.22. Classification accuracy of final random forest models for test portions of data ....243
Table 4.23. Category-specific classification accuracy of final random forest models for test portions of data ........................................................................................................244
Table 4.24. Bigrams for indirect act analysis .....................................................................298
Table 4.25. Variables that appeared in the final model for more than one regression for indirect act analysis ........................................................................................................300
Table 4.26. Classification accuracy of final regression models for test portions of data, for indirect act analysis ........................................................................................................301
Table 4.27. Category-specific classification accuracy of final regression models for test portions of data, for indirect act analysis .................................................................301
Table 4.28. Variables that appeared in the top 5 for more than one random forest final model, for indirect act analysis .................................................................302
Table 4.29. Classification accuracy of final random forest models for test portions of data, for indirect act analysis .................................................................302
Table 4.30. Category-specific classification accuracy of final random forest models for test portions of data, for indirect act analysis .................................................................303
Table 4.31. Summary of hypothesis 2 analysis results .................................................................331
Table 4.32. Category rankings by paper/subset .................................................................364
Table 4.33. Summary of findings from sections 4.2.4.1.1 and 4.2.4.1.2 .................................389
Table A.1. Summary of results from section 4.2.1 .................................................................487
CHAPTER 1: INTRODUCTION AND BACKGROUND

1.0 Introduction

In 1985, two of the foremost scholars of traditional speech act theory offhandedly left a mission to their academic descendents. While introducing the account of illocution that they would explore throughout the pages of Foundations of illocutionary logic, John Searle and Daniel Vanderveken made plain that their goal was to provide the theory. Any necessary concrete elements proposed by the theory were beyond their concern and left to the purview of “empirical linguistics” (p.1). The present project began with the intention of assuming such an empirical mantle and continuing where Searle & Vanderveken, among others, have left off. To that end, this dissertation will detail an empirical investigation of speech acts through the construction and analysis of an annotated corpus.

These concrete elements are what Searle & Vanderveken term “illocutionary force indicating devices” (IFIDs). In short, IFIDs are proposed to be responsible for the overt encoding of the illocutionary force of a given utterance. Other speech act scholars have theorized similar overt elements, as well, though none of them provide much in the way of concrete examples either. By building a corpus of data annotated for speech act category and employing computational methods, I will be searching for evidence of such overt elements. If found, this evidence would accomplish two related, theoretically-motivated goals: to confirm the existence of overt force encoding, and to help elucidate its qualities and limitations.

The degree to which illocutionary force is overtly encoded is a matter of debate in the field. The exact nature of any overt elements is likewise unclear. Do these elements categorically encode one force? Or do they exist on a spectrum of usage (e.g. encoding X force
most of the time, but Y force some of the time)? Are they reliable by themselves, or are they best understood as part of a set of features? These are some of the questions that will be explored by this dissertation. Specific questions and an explicit hypothesis concerning overt direct force encoding will be motivated later in this chapter, in section 1.1.4.2.

An investigation of illocutionary force encoding is also ripe for explorations into the boundary between direct and indirect speech acts. While the literature seems to agree that indirect acts are inferred while direct acts are overt, how this difference manifests is not entirely clear. Are indirect acts truly associated with no overt signs, or simply with fewer such signs than are direct acts? Or are the signs themselves simply less consistent for indirect acts? Such questions will be examined in the pages to come, and a testable hypothesis concerning indirect force encoding will also be motivated in section 1.1.4.2.

The methodology used for the investigation at hand was to construct a corpus of English language Twitter conversations, enlist native English speakers to annotate this corpus for speech acts, and then utilize computational methods to search the annotated corpus for potential IFIDs. Specifically, multinomial logistic regression and random forest analyses were both employed in this search. The annotated speech act category was the response variable, and the potential explanatory variables were all either theoretically-motivated or pulled from similar previous studies.

Twitter was chosen to be the medium for this study for several reasons. First and foremost, as a medium under the umbrella of computer-mediated communication (CMC), the language of Twitter is technically written but stylistically quite spoken (see section 2.4). The benefit of written data is that any overt clues to force category will be inherently recorded
(assuming the speaker wants to be understood), while the benefit of casual conversational data is that it is often less formal and more abridged. The elements necessary for interpretation should be included, but without as much formal superfluity. That particular aspect is compounded by the Twitter length restriction, which at the time of data collection was an infamous 140 characters. Hence, if overt IFIDs are to be found, Twitter seems like a good place to look for them.

Two additional reasons for choosing Twitter concern the annotators. First, the structure of Twitter is such that the impact of local conversational context is substantially reduced, or at least potentially contained. Thread-initial tweets, like conversational-initial turns, should theoretically provide any necessary scene-setting information. Meanwhile, any tweet made in reply to another tweet will be shown to be part of a thread, digitally linked to its preceding context. Second, tweets from any public account have an enormous potential audience. All of Twitter could theoretically see a given public tweet. If speakers want to be understood by a wide audience, they will be sure to include the aforementioned necessary scene-setting information. Together, both of these facts mean that annotators should have a good chance of correctly interpreting a given tweet (or, at the least, garnering an interpretation from it, as opposed to having no idea what the speaker is saying).

Using Twitter as the medium of study also allowed for the investigation of some medium-specific questions. What are people doing on Twitter, linguistically? That is, how does the speech act landscape of Twitter compare with that of other media? A specific hypothesis concerning the speech act landscape of Twitter will be motivated in section 1.1.4.3 of this chapter.
In sum, Chapter 1 will result in the motivation of three testable hypotheses. The segmentation and annotation guidelines laid out in Chapter 2 will lead to another hypothesis, based on a phenomenon of hashtag distribution that I observed in the course of motivating those guidelines. This phenomenon will be discussed and formulated into a fourth testable hypothesis in section 2.3.

Finally, the collection of variables utilized to test these four hypotheses combined with the availability of user-specific variables from the Twitter interface created an opportunity to investigate style in the medium. Is there a discernible communicative style to Twitter, and has the Twitter community exhibited accommodation to this style? The fifth and final hypothesis, concerning communicative style on Twitter, will be motivated in section 2.4.

The primary contributions of this project are to the field of pragmatics. At the time of writing, speech act theory has been around for about 60 years, but Twitter has only been around for about 13. Studies of speech acts on Twitter have remained infrequent. The present study thus contributes to both speech act literature and Twitter linguistic literature by investigating the intersection of the two topics. It also takes some novel steps. While previous studies have built manually-annotated corpora of speech acts on Twitter in the course of developing automatic speech act classifiers (Vosoughi & Roy, 2016; Zhang et al., 2011 and 2012), these studies did not seek to separate direct acts from indirect acts, nor did they accommodate the possibility of a tweet containing multiple speech act units. This dissertation does both. Additionally, uncovering possible answers to questions about the realization and nature of illocutionary
encoding could contribute to syntactic theories of the left periphery. Specifically, evidence of overt IFIDs could support proposals regarding the location of the force phrase.¹

1.0.1 Outline and Key Findings

Chapter 1 will begin with an overview of speech act theory, the goal of which is to determine the specific version of the theory which will underlie our investigation of illocutionary speech acts on Twitter. This version will also help determine the taxonomy used for the annotation, and provide specific definitions of key terms such as “illocutionary,” “direct,” and “indirect.” Chapter 1 will continue with a review of some of the technological affordances of the medium of Twitter. The goal of this section is to determine whether these affordances affect or perform speech acts themselves, which will be relevant when deciding on the exact form of the annotation task in chapter 2.

Chapter 2 will determine the exact form that the corpus will take by motivating speech act annotation and segmentation guidelines for Twitter data.

Chapter 3 will detail the construction of the corpus in three stages: data collection, data segmentation, and data annotation. The chapter will conclude with some initial results and observations of the corpus.

Chapter 4 will begin with an assessment of the agreement scores for the annotation results, before moving on to the analysis and discussion of each of the five hypotheses. The chapter, and the entire dissertation, will close with a discussion of overarching conclusions, primary contributions of the study, and recommendations for future directions. Some of the key findings to come out of this project are as follows:

¹ Such an avenue is not pursued in this work.
We will see support for the proposal that direct illocutionary force is at least partially overtly encoded, specifically as a template of typical features.

We will see that Twitter users form a linguistic community, complete with norms to which they accommodate over time and/or through experience.

We will see support for crowdsourced annotation as a feasible method of obtaining linguistic data.

We will also see, however, that this feasibility varies, seemingly according to the level of inference required to address a given phenomenon. Consequently, inferred phenomena may especially benefit from higher numbers of raters, or raters with more training.

1.1 Overview of Speech Act Theory

In this section, I will review and assess the various expansions and developments that have been made in the study of speech acts in the 60 years since J.L. Austin’s original proposal. The goal of this review is to determine the specific version of the theory which will underlie this investigation of illocutionary speech acts on Twitter. This version will also help determine the taxonomy used for the annotation, and provide specific definitions of key terms such as “illocutionary,” “direct,” and “indirect.”

Instead of assessing the evolution of “illocutionary act” from a purely chronological perspective, I have organized the major sources into a sort of family tree, seen in figure 1.1. The main branches of this tree (literalist vs. contextualist, and mentalist vs. non-mentalist) are fundamental groupings suggested by Kissine (2009) and Tsohatzidis (2010), respectively. The literalist/contextualist distinction indicates the extent to which a “literal meaning” of an utterance exists independently of extra-linguistic factors in that theory (which in turn affects how
illocutionary force is to be defined), while the mentalist/non-mentalist metric indicates the relative importance of the mental state of the speaker vs. normative elements in defining and determining illocutionary force under that theory. I will be looking at each of these main branches (and their sub-branches) in turn, but first it seems pertinent to provide a brief summary of the major aspects of their common ancestor, Austin (1962).

![Speech act theory family tree](image)

**Figure 1.1. Speech act theory family tree.**

### 1.1.1 Austin (1962)

In *How To Do Things With Words*, Austin provides a tripartite classification of speech acts using terms that are still used today: locutionary acts, illocutionary acts, and perlocutionary acts. These correspond respectively to acts of *speaking* (the speaking itself), acts done *in speaking* (the function or purpose of saying what you did), and acts done *by speaking* (the results of the saying).
Focusing in on the illocutionary act, Austin introduced the notion of separating the “force” of an utterance from the “meaning.” The meaning of an utterance is a part of the locutionary act, since a speaker speaks a specific word with a specific “sense” and “reference” in mind. Meanwhile, the purpose of the utterance is conventionally encoded, in the syntactic form or word choice. This is the illocutionary force with which the propositional content of the utterance is stated (Austin, 1962, pp.98-99). The force being “conventional” in this way meant that, according to Austin, an utterance could always be restated (if not already stated) with an explicit performative verb that details (and performs) the exact force. For example, “The stove is hot” does not contain a performative verb, but its force can be understood anyway. Austin seems to suggest that this is because “The stove is hot” is elliptical for a more complete version of the utterance that contains the appropriate performative verb, depending on the context (e.g. “I warn you that the stove is hot,” or “I state that the stove is hot”).

Austin’s original account, while undeniably foundational, had some notable insufficiencies. Perhaps foremost is that he only considers some utterances to perform acts. Namely, he only considers those which are not usually considered either true or false. The rest (those which can usually be assessed for truth or falsity, e.g. “The sky is blue,”) he deems non-performative “constatives.” The problem is that this delineation creates an immediate inconsistency in the theory due to his “performative formula”, the idea that every act can be restated/performated with an auto-descriptive performative verb. I can replace “The sky is blue” with “I state that the sky is blue,” and there is no inherent reason to say that “stating” is not an act of the same kind as, say, “asking.”
Additionally, Austin’s “performative formula” is easily challenged by a number of illocutionary acts, for example threatening: “I threaten you with a failing grade” would hardly be said to be a felicitous threat (Sadock, 2008, p.56). It would be quite odd to say, and it is too formal to actually be fear-inducing. A more likely threat is something like, “If you don’t shape up, I’m going to give you an F.”

Austin’s account also leaves some important questions unanswered, or at least not sufficiently addressed. One such question is where exactly the locutionary/illocutionary and illocutionary/perlocutionary boundaries lie, and if they are hard or soft. For example, to solve the problem of differentiating the act of “warning” from “attempting to warn” (which could both be performed by the exact same utterance), Austin requires the securing of uptake to classify something as a successful illocutionary act. But as later scholars point out, uptake should be classified as a perlocutionary act, per Austin’s own definitions. As we will see, most later speech act scholars adhere to the idea that illocutionary acts that fail to be understood by an adequately equipped observer have no inherent constitutional differences from those that happen to be understood.

Another big question that Austin leaves ultimately unanswered is what to do about any utterance other than a straightforward one. He states that actions like “joking” or “acting a part” are not illocutionary acts, but are instead marginal, exploitative uses of language. This does not mean that jokes do not perform illocutionary acts, as he clarifies that an utterance may be analyzed for force regardless of whether it is intended as a joke. However, he believes that the joking intention itself has no bearing on that force, and that the exact nature of the connection between the joking intention and the utterance is unclear (p.104). While later scholars generally
agree with the former point (that joking, sarcasm, and irony are not quite illocutionary acts themselves), we will see that several of them attempt to pick up where Austin left off in terms of finding a space within speech act theory for non-serious uses of language.

The extent to which later scholars depart from Austin’s original account varies considerably, as do the different directions these departures take. But they all share the primary goals of defining illocutionary force and delineating the relationship between the force and meaning of an utterance, often making reference to the speaker and/or hearer in the process. As I make my way through the various branches of the speech act theory family tree, my goal is the same: to arrive at a motivated definition of “illocutionary act” which will underlie my investigation.

1.1.2 The Family Tree

The most fundamental split in the speech act theory tree, literalist vs. contextualist, is ultimately a result of different ideas about the semantics-pragmatics interface, and how it interacts with notions of illocutionary acts. Kissine (2009) summarizes this difference by discussing how sentence meaning is related to “what is said.” A literalist holds that there is some level of sentence meaning that exists independent of context, and that this meaning is generally equivalent to “what is said.” From there, the illocutionary force (and therefore “what is meant”) is determined based on the interaction of “what is said” and the context. Meanwhile, a contextualist disagrees that “what is said” can exist outside of context (p.123). In Austinian terms, speaking certain words of a language L (e.g. performing a phatic act) must be done with a certain sense and reference in mind to be considered a rhetic act. Seeing as the same phatic act can perform different rhetic acts in different contexts, contextualists view an utterance as
uninterpretable outside of context. Essentially, as Recanati (1989) phrases it, contextualists believe that, “The meaning of the sentence...seriously underdetermines what is said” (p.297).

1.1.2.1 Literalist Accounts of Illocutionary Force

A literalist account includes reference to a proposition that can be defined in formal semantic terms with minimal pragmatic input (such as the contextual determination of the content of indexicals, deictic elements, and the like), as well as a definition of illocutionary force that connects it to the literal meaning of the sentence. Connecting the illocutionary force to the literal meaning of the sentence may mean that there is an explicit piece of language that encodes the force (Searle, Searle & Vanderveken), that the literal sentence meaning places it into a category compatible with a predetermined set of forces (Bach & Harnish), or that the literal sentence meaning conveys a set of conditions (corresponding to a speech act category) that the speaker is taking responsibility for (Alston).

1.1.2.1.1 Literalist Mentalist Accounts: Speaker Only

The accounts in this branch qualify as “mentalist” because the success of an illocutionary act under these theories depends in some way on a cognitive state (as opposed to a normative feature), and specifically only on the cognitive state of the speaker.

Searle (1969) proposes that “...for any possible speech act there is a possible linguistic element the meaning of which (given the context of the utterance) is sufficient to determine that its literal utterance is a performance of precisely that speech act” (pp.20-21). His stipulation of “given the context of the utterance” refers to rules of utterance that the speaker must follow. For example, uttering a propositional content with the force of, say, “promising,” means that a speaker must determine that certain conditions prerequisite to promises hold true before uttering
the explicit piece of language (the illocutionary force indicating device, a.k.a. IFID) that signals the “promising” force. If the proper conditions are met, the speech act is performed successfully and felicitously and can be formally represented as the standard $F(p)$, where $F$ stands for the force as indicated by the IFID, and $p$ stands for the propositional content of the utterance (which includes all referring and predicating) (Sadock, 2008, pp.59-60; Searle, 1969, p.23).

In Searle & Vanderveken (1985), the authors divide illocutionary force into seven components. Six of these components are various types of conditions required for felicitous utterance and performance of the given act (conditions on $p$ as well as contextual specifications, many of which have to do with the speaker and his cognitive state, particularly his intentions), while the seventh is the “illocutionary point” (pp.12-20). In their words, the illocutionary point is “that purpose which is essential to [a given illocutionary act]'s being an act of that type” (p.14). The illocutionary point is where the meat of the utterance’s conversational function comes from. For example, speech acts of ordering and requesting both have the illocutionary point of directive – the six remaining components of illocutionary force are responsible for differentiating between what is an order versus a request.

Searle & Vanderveken propose a five-category taxonomy of these illocutionary points (familiar from Searle’s previous work), which they support with what they call “direction of fit.” The job of the illocutionary point, in their words, is to relate $p$ to the world of utterance, which can only be done in so many different ways -- thus the limited taxonomy of illocutionary points (p.52). The categories and their directions of fit can be seen in table 1.1.
Table 1.1. **Direction of fit definitions from Searle & Vanderveken (1985).** Definitions are verbatim (pp.53-54).

<table>
<thead>
<tr>
<th>Direction of fit</th>
<th>Illocutionary Point</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Word-to-world</em>: In achieving success of fit the propositional content of the illocution fits an independently existing state of affairs in the world.</td>
<td>Assertive</td>
</tr>
<tr>
<td><em>World-to-word</em>: In achieving success of fit the world is altered to fit the propositional content of the illocution.</td>
<td>Commissive (responsibility lies with speaker)</td>
</tr>
<tr>
<td><em>Double direction of fit</em>: In achieving success of fit the world is altered to fit the propositional content by representing the world as being so altered.</td>
<td>Directive (responsibility lies with hearer)</td>
</tr>
<tr>
<td><em>Null or empty</em>: There is no question of achieving success of fit between the propositional content and the world, because in general success of fit is presupposed by the utterance.</td>
<td>Expressive</td>
</tr>
</tbody>
</table>

**Discussion of Literalist Mentalist Accounts: Speaker Only**

Searle & Vanderveken’s account is initially appealing due to its neat packaging and intuitive formulation. There is a concise and theoretically motivated taxonomy of illocutionary forces, which are applied to semantic propositions to yield illocutionary acts. The presence of a given force is indicated by an explicit element, though this element (IFID) is often left implicit if unneeded due to context. It is no wonder that this taxonomy (though not necessarily the motivations behind it) has been used in several other speech act annotation projects.

However, this approach runs into some of the same problems that we will see repeated elsewhere. The most obvious is that these authors emphasize that their account is meant to cover “serious and literal” utterances. The force is supposedly literally encoded, and signals the
speaker’s intention. But if a speaker’s intention is to use a piece of language creatively (again, non-literally and/or non-seriously), we do not have a way, within Searle & Vanderveken’s account, to accommodate this information. It is possible that Searle & Vanderveken actually hold the same view of this matter as did Austin. That is, their statement that they are only accounting for “serious and literal” utterances could simply mean that they consider joking, sarcasm, and irony to be marginal elements that do not directly interact with the force of an utterance. But a theory that links speaker intention to force needs to find a way to incorporate different genres of intention if it hopes to address how people actually communicate.

Additionally, requiring that force be literally encoded means requiring that, in the case of indirect acts, the direct acts with which they are performed are also performed. That is, for example, a rhetorical question may be asserting something but it is truly also a question. Or an utterance of Can you pass the salt? may be indirectly requesting that one pass the salt, but is also truly asking whether one is able to do so.

In a related vein, as Alston (2000) points out, the contextual conditions that Searle lays out for the “sincere and non-defective” performance of a given act can easily be too strict and too easily violated. Alston critiques Searle’s (1969) discussion of the act of promising to demonstrate that violating such conditions does not necessarily yield an insincere or defective act (though perhaps a perlocutionally unsuccessful one). For example, Searle posits that one of the conditions necessary for a sincere and non-defective promise is that the hearer “would prefer [the speaker]’s doing [whatever is being promised] to his not doing [it]” (Alston, 2000, pp.51-52). But if I promise to buy ice cream for a friend, a later discovery that she is lactose-intolerant should not be considered to render my initial promise defective or insincere. As we will see
shortly, Alston’s account (and the other theories in his branch of the tree) solves this problem by putting the contextual requirements in less strict terms.

The second main issue with Searle & Vanderveken’s theory, which we will see is also true for several of the contextualist accounts, is that certain key elements are motivated in the theory but only a few (if any) concrete examples are provided. In the contextualist accounts, we will see that the clarification of which pragmatic processes and contextual information are primary (filling out the explicature) and which are secondary (determining implicatures and indirect acts) is missing. In Searle & Vanderveken’s account, this missing element is the exact realizations of the IFID. They declare early on that they are not interested in exploring this question, summing up the matter with the statement that the IFID shows up “in the syntax of actual languages in a variety of ways, e.g. mood, punctuation, word-order, intonation, contour, and stress,” and is “literally used to indicate...that [the utterance] has a certain illocutionary force or range of forces” (Searle and Vandeveken, 1985, pp.1-2). Thus, while it is clear that the element is (or at least can be) explicit and is endowed with power through its literal meaning, Searle & Vanderveken generally concern themselves with the logic of illocutionary acts and leave the task of identifying exact IFIDs to “empirical linguists.”

1.1.2.1.2 Literalist Mentalist Accounts: Speaker and Hearer

Strawson (1964) took Austin’s account and re-interpreted a key element to generate a new, intention-oriented branch of the speech act family tree. In short, he states that Austin must have meant something other than the “most natural” meaning of “convention” in his explication of illocutionary acts (p.445). When Austin states that an act is conventional “in the sense that it at least could be made explicit by the performative formula,” Strawson proposes that what is
made explicit in such circumstances is the speaker’s intentions. Hence, he incorporates Grice (1957) into his account of illocutionary acts. In producing an utterance, the speaker intends to elicit a certain response from the hearer “by means of recognition on the part of the audience the intention to produce that response, this recognition to serve as part of the reason that the audience has for its response, and the intention that this recognition should occur being itself intended to be recognized” (p.450). These intentions must be overt and avowable to result in an illocutionary act (p.454).

At this point, Searle (1969) points out a problem with such an account. Essentially, he writes that a Gricean account of illocutionary force reduces the role of convention (in the way meant by Austin) by too large of a margin. Grice reduces it so much, according to Searle, that the meaning of the words is now merely another piece of contextual information. He finds this problematic, insisting that speaker meaning and word meaning merit distinction. Searle demonstrates his point by constructing the following example: an American soldier (A) attempts to get an Italian soldier (B) to think that he (A) is a German soldier. To do this, A speaks the only German he knows (a poetic quotation), hoping that B does not know enough German to realize what exactly A is saying. A’s intention is to convey the meaning that “I am a German soldier,” but this intention should not be understood to erase the conventional meaning of the words (i.e. they still mean something else entirely, in German). However, according to Searle, that is exactly where a Gricean account ends up (pp.44-45).

Searle does acknowledge, however, that this problem can be addressed by allowing for the intention of conventional use. That is, convention may be incorporated into a Gricean account by allowing the speaker’s conventional use of a word to be one of the intentions he seeks
to convey to the hearer (and therefore one of the conditions that the hearer must recognize or infer). Strawson takes this exact route. He states that there are indeed specific linguistic and cultural conventions (in the sense meant by Austin) that are employed in some illocutionary acts, but he emphasizes that the Gricean inference dance is still necessary for interpretation in such cases. The speaker must convey his intention to use the convention, and the hearer must be familiar with the convention and recognize the speaker’s intention to invoke it. Regardless of whether a linguistic or cultural convention is used, it is the speaker’s responsibility to make his intentions overt (p.458).

Bach & Harnish (1979), influenced by Strawson (1964), proposed that the direct and literal illocutionary force potential of a given utterance is the set of forces that are compatible with the mood and meaning of the locution e (immediately addressing Searle’s initial critique of Grice). In this sense, the illocutionary force is “underdetermined” by the explicit act. While “L-compatibility” is itself determined by formal elements (e.g. literal requests and orders are L-compatible with imperative sentences), the final understanding of which exact potential act is being performed is a matter of inference on the part of the hearer. The speaker, for his part, intends for this inferencing to occur and in fact counts on it when forming his utterance (Bach & Harnish, 1979, p.11). The authors couch the inferencing in terms of what they call “mutual contextual beliefs” – mutually-held and salient (on the parts of the speaker and hearer) beliefs (p.5).

Harnish (1994) succinctly details the inferencing as a two-step process: 1) the hearer infers the speaker's expressed attitude from the form of the sentence, and 2) he then infers from the attitude to the intended force. The “expressed attitude” can only be one of four options,
based on the four basic sentence types (declarative, imperative, yes/no interrogative, and wh-interrogative). The inference from the attitude to the intended force relies on compatibility between the two. For example, the expressed attitude of an imperative is that the speaker “desires/intends that [the hearer] make it the case that \( p \).” From this, the hearer can only posit illocutionary forces wherein the speaker desires/intends that the hearer make it the case that \( p \) (e.g. ordering, requesting, etc.) (Harnish, 1994, p.433).

Bach & Harnish also attempt to address non-literal and non-serious utterances, something which is an especially difficult (and oft ignored) task for the literalist branches of the speech act family tree. Bach & Harnish propose a four-way distinction (two orthogonal spectra) composed of “direct vs. indirect” and “literal vs. non-literal.” The first spectrum characterizes whether or not the act happens by way of another, while the latter characterizes the exact words used. For example, a direct and literal act would be *Please pass the salt*, while an indirect but literal act would be *Could you pass the salt?* Meanwhile, a direct but non-literal act would be *My mind got derailed*. Finally, Bach’s example of an act that is both non-literal and indirect is a sarcastic performance of *I love the sound of your voice* meant to indirectly request that the hearer stop singing (Bach, 2004, p.470).

**Discussion of Literalist Mentalist Accounts: Speaker and Hearer**

Like Searle & Vanderveken’s theory, the Strawson/Bach & Harnish branch of the tree has some initial appeal in its concise and theoretically motivated taxonomy. Additionally, it succeeds in a major area where Searle & Vanderveken’s account is lacking, namely the over-reliance on conventions and specific contextual conditions for successful illocutionary act performance. This is especially demonstrated by the fact that Bach & Harnish at least make an
effort to incorporate non-serious and non-literal utterances into their account. We will revisit such utterances in section 1.1.3, where I discuss direct vs. indirect acts.

While this branch of the tree is quite sound from a theoretical perspective, I believe that there is at least one element that can be improved upon. Strawson/Bach & Harnish define illocutionary acts in terms of both the speaker’s intentions and the hearer’s uptake. The inclusion of the latter element performs a crucial act of mitigation. If only a speaker’s intentions counted, they could be said to perform any act at all through the use of any utterance at all. By including the “uptake” stipulation, we reign in the speaker’s freedom and more accurately reflect the reality that language is a system of cooperative communication wherein words have meaning.

But instead of using hearer uptake to reflect a system of cooperative communication, one could reference the system directly. That is, one could define illocutionary acts in terms of the speaker’s intentions as well as the speaker’s consideration that they are operating within a system of cooperative communication. I believe that such a definition is an improvement on the “uptake” definition because it does not allow for interference or complications due to hearer-specific variables. A given utterance heard by two hearers from the same general linguistic community may be taken up by one and not by the other. If we define illocutionary acts in terms of hearer uptake, it is unclear whether such an utterance has successfully performed such an act. But if we instead define illocutionary acts in terms of that general linguistic community, we can avoid such confusion. Such a definition makes clear that the speaker’s intention to have his intended meaning recognized by the hearer does not rely on knowledge of that hearer’s mind, but instead on the speaker’s assumption that the hearer is a member of a given linguistic community, one that the speaker knows the rules of. This also more accurately
reflects how we communicate with someone new. We do not need to have any specific knowledge of this new hearer’s mind to have intentions of being understood, as long as we think that the new interlocutor is a member of a linguistic community that we are familiar with.

In short, the Strawson/Bach & Harnish branch of the speech act family tree emphasizes the importance of speaker intention while also incorporating linguistic and cultural convention in the form of hearer uptake. One could instead incorporate linguistic and cultural convention by making reference to the cooperative communication system that established and/or upholds that convention. We will see at least one approach that takes this route in the next section.

1.1.2.1.3 Literalist Non-Mentalist Accounts

Non-mentalist approaches de-prioritize the role that the mental states of the speaker and/or hearer play in the performance of illocutionary acts, emphasizing social or socio-contextual elements instead. Normativist approaches, which revolve around “normative features,” and dynamic update models, which revolve around a structured discourse context, both fall under this umbrella. Each will be discussed in turn.

Normativist Approaches

Normativist accounts consider social “normative features” to be fundamental to the performance of illocutionary acts. Alston (2000) puts this “normativity” in terms of what a speaker can be expected to take responsibility for, based on his utterance. To perform an illocutionary act is to subject your utterance to a certain normative rule, which specifies those events or states of affairs that you would be held responsible for given the utterance. These “I-rules” are “socially entrenched,” (p.58), and outline the conditions necessary for the felicitous use of a given utterance to perform a given illocutionary act (p.7). The different conditions of
the I-rules are what determine the different categories of possible illocutionary acts. Meanwhile, the fact that the same utterance could mean different things at different times is merely due to an interaction of the sentence meaning and the context which results in different I-rules potentially being available at a given moment (p.188).

Gauker (2007) has a similar approach. But instead of focusing directly on the norm of what a speaker could reasonably be held accountable for as a result of their utterance, Gauker centers his view around normative conversational goals that are borne out of what conversation is for. On his definition, a given conversation is simply a situation wherein a set of interlocutors have the same such goal (p.132). In a conversation, a speaker produces an utterance with the intention that it contributes to the conversational goals in one of a few possible ways. Each of these ways corresponds to a speech act category. For example, an utterance that is meant to contribute to the conversational goals by contributing something reasonable to the common ground would make the utterance an assertive speech act.

Essentially, these two approaches both make the aforementioned improvement on Bach & Harnish’s approach. They incorporate cooperation directly into the theory instead of using hearer uptake as a proxy.

**Dynamic Update Models**

In dynamic update models of illocution, different types of illocutionary acts update different aspects of the structured discourse context in different ways. Declaratives (or speech acts that are usually accomplished by declaratives) update the common ground, which is a set of propositions of which all discourse participants are aware. Interrogatives (or speech acts that are usually accomplished by interrogatives) update the question set, which is the set of questions
whose answering would be felicitous at the moment of utterance. Finally, imperatives (or speech acts that are usually accomplished by imperatives) update the “to-do list function,” which assigns “a set of properties to each participant” (Portner, 2018, p.227).

Gazdar (1981) proposes such a dynamic account, wherein speech acts are “functions from contexts into contexts.” Instead of a function from a propositional content to a speech act, he defines illocutionary force as a function from a sentence meaning (a.k.a. “content”) to a speech act. The speech act takes the context and alters it by accounting for the content of the sentence in a certain way, as dictated by the force (Gazdar, 1981, pp.66-70). For example, Gazdar classifies an assertion as a function that takes the context and changes it to include the speaker’s commitment to the truth of the content (p.69). That is, an assertion updates the common ground.

In Kaufmann’s (2012) dynamic proposal, she posits a force function that takes the content as argument, outputting a function that takes the initial context and alters it in a certain way (taking into account the content argument), outputting that altered context. Thus far, her theory has the same basic apparatus as Gazdar’s theory. However, Kaufmann makes (at least) two major changes: 1) she incorporates a “pre-context” of felicity conditions that does some initial narrowing of possible speech act assignments, and 2) she further speculates on how exactly the semantic type of the content argument alters the context (something Gazdar had only mentioned). Basically, the semantic type also narrows the possible speech act assignments. If the content is a proposition, the output context is one where that content is entailed. If the content is a question, the output context is partitioned according to the possible answers. On this account, the pre-context of felicity conditions carries quite a bit of sway in terms of determining
what speech act occurs. The most significant example of this is that, in Kaufmann's account, the pre-context is entirely responsible for differentiating between speech acts traditionally accomplished by declarative sentences and those traditionally accomplished by imperative sentences (as she suggests that the content arguments of both are of the proposition semantic type) (Portner, 2018, pp.169-174).

While Kaufmann (2012) allowed the semantic type of the content argument to narrow the field of possible speech act assignments, Krifka’s (2014) dynamic approach gives that power to a syntactic force operator. Four operators are possible: ASSERT, PERFORM, COMMAND, and ASK. The force operator takes a proposition $p$ and assigns a function that maps the context of utterance onto another function (the speech act) which specifically maps the world-time index component of the context onto whatever world-time index results from incorporating $p$ into the context. The exact way that $p$ is taken into account (i.e. which part of the structured discourse context is updated) varies by operator choice (as with, in Kaufmann's theory, it varied by semantic type) (Portner, 2018, pp.174-176).

**Discussion of Literalist Non-Mentalist Accounts**

In a way, the approaches in this branch of the speech act family tree split the difference between the convention-heavy Searlean branch, and the intention-heavy Strawsonian branch. They escape the problem of overly strict or specific meanings, and thereby the problems of a literal force approach, by invoking sentence meaning in terms of utterance usage on the whole as opposed to strict semantic meaning of an individual element. Meanwhile, they improve upon the approach of Bach & Harnish by invoking intention and intuition not in terms of an effect on the hearer, but instead in terms of an effect on the discourse. The approaches in this branch treat the
discourse itself almost like another interlocutor, but one whose actions and reactions are more reliable, and predictable, than an actual human interlocutor.

1.1.2.2 Contextualist Accounts of Illocutionary Force

The contextualist approaches to illocutionary force fundamentally diverge from the literalist ones in that they consider pragmatics to have a much greater role in the construction of the underlying proposition than do the non-contextualists. For some theorists in this branch, the role of something like a semantically-determined proposition is simply reduced. For others, it is almost entirely eliminated.

1.1.2.2.1 Contextualist Accounts: Reduced Role of the Proposition

The approach discussed in this section is that of Kissine (2009). As we will see shortly, it bears superficial similarities to literalist accounts of speech acts. It includes a taxonomy that we have seen in most branches of the speech act theory family tree, along with definitions of the categories in that taxonomy that are ultimately based on the speaker’s and hearer’s relationships to a proposition-like object $p$ (by way of the common ground). But, as prefaced above, the constitution of $p$ is where the main differences lie.

At this point, it is pertinent to mention how contextualist theories analyze utterance meaning and speaker meaning. In particular, Kissine’s approach makes reference to certain concepts from relevance theory. Relevance theory (RT) takes Grice’s inferential machinery and reduces it down to just one maxim: the maxim of relevance. According to relevance theorists, the assumption that a speaker’s utterance is relevant to the utterance context is all that the hearer needs to infer the speaker’s meaning. Furthermore, relevance is so fundamental that it also participates in the determination of utterance meaning in RT. That is, RT concludes that
linguistically-coded content alone is not sufficient to yield an interpretable utterance. While this perspective is not particular to RT (or even to contextualism), RT diverges from literalist approaches by conflating the pragmatic processes that participate in the determination of utterance meaning and those that participate in the determination of speaker meaning. RT terms the contextually-determined utterance meaning the “explicature” (Sperber & Wilson, 2008, pp.615-626).

Recanati (1989), another contextualist, makes a similar proposal. Just as Grice proposes a pragmatic system to bridge the gap between “what is said” and “what is meant,” Recanati (1989) proposes that pragmatics must also bridge the gap between “sentence meaning” and “what is said.” He then outlines criteria for delineating when a pragmatically-determined aspect of utterance meaning is a conversational implicature (in the space between what is said and what is meant), versus “constitutive of what is said” (joining with the sentence meaning to determine what is said) (p.302). This is similar to what relevance theorists propose, the main difference being that, instead of the two sequential pragmatic steps that Recanati proposes, relevance theorists propose one pragmatic step that determines what is said (the “explicature”) and anything else that may have been communicated (the “implicatures”) at the same time.²

Thus, instead of a primarily semantically-determined proposition \( p \), we have Recanati (1989)’s “contextually-determined what is said” and the relevance-theoretic “explicature.” Kissine takes this concept and gives the \( p \) in his theory a third name: “locutionary act.”

---

² Recanati, for his part, later specifies that the sequential nature of his theory is only logical, and not temporal, and hence agrees with the relevance theoretic idea that the two levels may mutually affect each other’s output (Carston, 2007, p.24).
Once this account of $p$ is in place, Kissine (2009) moves back into more familiar territory. He proposes that illocutionary forces are qualities that may be ascribed to utterances in the event that certain conditions are met, suggesting a five-supercategory taxonomy of these forces: 1) assertive, 2) directive, 3) commissive, 4) interrogative, and 5) institutional (e.g. verdictive, declarative). His theory does not apply to institutional speech acts, by definition, seeing as their performance hinges on social conventions as opposed to pragmatics (pp.123-124). For the remaining categories, however, he goes on to explain the contextual conditions that “trigger” those illocutionary forces, excepting interrogatives.\footnote{The reason for this exception is not made clear.} For example, for assertive illocutionary force, Kissine proposes that “a successful assertion that $p$ is a reason for [the hearer] to believe that $p$.” As mentioned above, $p$ refers to the content of what he has termed the “locutionary act” (i.e. the “contextually-determined what is said,” the “explicature” of relevance theory) (p.129). Kissine puts this in terms of possible worlds. As long as there is a possible world in the common ground wherein the combination of this utterance with the contextual conditions of that world would be enough to infer that the content $p$ is true in that world, then the utterance performs a direct and literal assertive speech act (p.130).

**Discussion of Contextualist Accounts: Reduced Role of the Proposition**

While this type of contextualism has some theoretically appealing aspects, it is unclear how such a theory would be experimentally tested using Twitter. At the least, assuming this theory to underlie our study of illocutionary acts on Twitter would not look much different than assuming a theory that references a proposition $p$ from the start. Thread-initial tweets should theoretically be written to be understood by themselves, without the need for explicit knowledge
of the tweeter’s life or any previous utterances. In fact, possessing such explicit knowledge does not even necessarily help with interpretation. Kruger et al. (2005) found that friends and strangers were equally accurate at judging certain emotions in emails, while Riordan & Trichtinger (2017) found the same result for judging the presence of sarcasm in emails. It appears, therefore, that even though friends share a larger common ground with the speaker, it does not necessarily help them when interpreting computer-mediated messages.

This is all to say that thread-initial tweets are already about as context-independent as they could possibly be. Extracting them further “out of context” in order to see how complete of a proposition remains might not even be possible, with the possible exception of removing the context that they are, in fact, tweets. However, the annotation goal is to gather native speaker intuitions about illocutionary classifications on Twitter, not to test whether the content of tweets is interpretable if extracted from the context of “this is a tweet.”

1.1.2.2 Contextualist Accounts: No Role for Proposition

Sbisa (2006) proposes an even further departure from literalist accounts, accusing them of misinterpreting crucial aspects of Austin’s original theory. She believes that Austin’s failure to explicitly refer to “propositions” in his work was not unintentional, and that the notion of “proposition” was inserted into the theory by Searle and Strawson (and propagated by their theoretical descendents) in order to make the locutionary/illocutionary boundary explicit and to provide an easily identifiable locus of truth/falsity (p.156). This inserted notion of proposition, she states, weakened the real world power of speech acts that Austin originally intended. She believes that actions accomplished through linguistic means are just like non-linguistic actions, and we must eliminate the notion that using words is so different from using other tools at our
Sbisa proposes that we must find a way to break down illocution without making reference to propositions, and suggests that one way in which this could be done is to use “action attribution and responsibility.” That is, instead of analyzing an illocutionary act by saying that A’s utterance applied force X to proposition Y, we should say that A’s utterance has the effect that she can be attributed responsibility for Z, a “state of affairs or event in the world” (p.170). The boundaries between locution, illocution, and perlocution are at least blurred and possibly entirely eliminated in Sbisa’s account.

Sbisa is not alone in proposing that propositions should not be a central component of speech act theories. Barker (2007) follows suit by abandoning the force/sense distinction all together, while still respecting the boundary between illocution and perlocution (pp.192-193). Illocutionary force, he proposes, is not applied to propositions but rather to “proto-acts.” For example, Barker proposes that asserting amounts to advertising an intention to defend a certain cognitive state, as opposed to committing to a (propositional) belief, or contributing a (propositional) piece of information to the common ground. This advertisement is the “proto-act,” which can be accompanied by an actual defense of that state (resulting in a literal assertion) or a rejection of the advertised intention (resulting in a non-literal assertion). A literal assertion is then judged as true, by the hearer, if she would also defend that state of mind in herself. Significantly, these cognitive states are not only belief states. That is, the cognitive state of mind that one advertises by saying, say, “Haggis is tasty” is not simply a belief state that haggis is delicious, but rather a “gustatory preference state” for haggis (p.199).

Interestingly, this notion of “what the speaker can be held accountable” for is similar to what Alston (2000) develops in his account of illocution, which still makes use of propositions and rejects the conflation of the illocutionary and perlocutionary levels.
Barker’s approach does not tackle the notion of proposition only as it applies to illocution, but across the board -- putting all of compositional semantics into these same cognitive (and ultimately, pragmatic) terms. For example, negation is reexamined as “an advertisement of the intention to defend the rejection of a certain cognitive state,” while universal quantification is “an advertisement of the intention to defend any advertisements of intentions made by singular statements of a particular kind” (Tsohatzidis, 2010, p.359).

Discussion of Contextualist Accounts: No Role for Proposition

These proposition-free approaches to illocution raise interesting questions, and a deeper dive into expressivism would be merited if the primary goal of this project was to ascertain the most sound illocutionary machinery. However, the primary purpose of this overview of illocutionary theories has been to determine the best version of speech act theory to underlie the annotation project. While sound theory is important to that end, so is a fully fleshed out taxonomy and definitions that are approachable and intuitive to potential annotators.

1.1.2.3 Summary Discussion

The main reason to choose one specific illocutionary theory from those outlined above is to determine a sound and specific definition of “illocutionary speech act,” a specific taxonomy of illocutionary speech act categories, and specific definitions of those categories that are accessible to the annotators. The literalist non-mentalist branch of the tree particularly recommends itself from this perspective. The approaches in this branch have evolved in response to the groundwork of the literalist mentalist approaches, and made theoretical improvements on them. Meanwhile, the contextualist approaches discussed above focus more on inferential machinery than definitions and taxonomies.
Within the non-mentalist branch, Alston’s account recommends itself by being fully fleshed out and including a detailed and familiar taxonomy with specific definitions for each category. Alston’s definition of illocutionary act breaks an illocutionary act into a meaningful propositional content and a force. Therefore, we can still use a base model of \( F(p) \). The biggest effect on the annotation task will take place in how we characterize the force, \( F \) (i.e. how we define the categories for the annotators). As explained above, Alston defines illocutionary force by the socially entrenched rules that a speaker submits their utterance to -- rules which stipulate the necessary conditions “for permissible utterance only within a certain sphere of activity, only when [the speaker] is playing the appropriate ‘language-game’ [as determined by the speaker],” (p.62). More concretely, Alston proposes the following taxonomy of illocutionary act potentials, with associated illocutionary rules:

[Note: \( U \) = the speaker; \( H \) = the hearer; \( S \) = the utterance or utterance surrogate; \( R \) = taking responsibility for \( X \) condition being satisfied. All rules in this table are reproduced verbatim.]
<table>
<thead>
<tr>
<th>Category</th>
<th>Associated I-Rules</th>
</tr>
</thead>
</table>
| Assertive  | U asserted that $p$ in uttering $S$ iff:  
1. $U$ R’d that $p$.  
2. $S$ explicitly presents the proposition that $p$, or $S$ is uttered as elliptical for a sentence that explicitly presents the proposition that $p$.  
(p.120)   |
| Directive  | $U$ D’d in uttering $S$ (where ‘$D$’ is a term for some directive illocutionary act type, a purporting to be producing a certain kind of obligation on $H$ to do $D$) iff in uttering $S$, $U$ R’d that:  
1. Conceptually necessary conditions for the existence of the obligation are satisfied. (These include such things as that it is possible for $H$ to do $D$, that $D$ has not already been done, etc.)  
2. Circumstances of the utterance of $S$ are appropriate for the production of the obligation in question. (This includes the appropriate authority for orders, the right kind of interpersonal relationship for requests, etc.)  
3. By uttering $S$, $U$ lays on $H$ a (stronger or weaker) obligation to do $D$.  
4. $U$ utters $S$ in order to get $H$ to do $D$.  
(pp.102-103)   |
| Commissive | $U$ C’d in uttering $S$ (where ‘$C$’ is a term for a commissive illocutionary act type, a purporting to produce an obligation on $U$ to do $D$) iff in uttering $S$, $U$ R’d that:  
1. Conceptually necessary conditions for $U$’s being obliged to do $D$ are satisfied.  
2. $H$ has some interest in $U$’s doing $D$.  
3. $U$ intends to do $D$.  
4. By uttering $S$, $U$ places herself under an obligation to do $D$.  
(p.97)   |
| Exercitive | $U$ O’d in uttering $S$ (where ‘$O$’ is a term for purporting to be producing a particular conventional effect, $E$) = df. In uttering $S$, $U$ R’d that:  
1. Conceptually necessary conditions for $E$.  
2. $U$ has the authority to produce $E$.  
3. Conditions are appropriate for the exercise of that authority.  
4. By uttering $S$, $U$ is bringing about $E$.  
(p.93)   |
| Expressive | $U$ expressed a $P$ (some psychological state) in uttering $S$ iff in uttering $S$, $U$ R’d that $U$ has a $P$.  
(p.109)   |
Before I delve into a little more about Alston’s theory and the specific categories in his
taxonomy, I want to mention that his explicit reference to “utterance surrogates” is extremely
relevant to my study, and really any study of modern CMC data. In that same way that a nod of
the head can be elliptical for, say, *Yes I’d love the last donut*, so might a thumbs-up emoji. Or a
smiley face. Or a click of the “like” button. Alston’s inclusion of these surrogates is crucial.
Whether or not an action such as “retweeting” can also be interpreted as a surrogate (for
something like, *I would like to repeat X or I also believe X*) will be addressed in section 1.2.

Nonassertive illocutionary acts, in Alston’s theory, can be summarized as the taking of
responsibility for certain conditions to hold, “plus the production of a certain kind of
conventional effect,” (p.89). Meanwhile, assertive illocutionary acts are reduced to taking
responsibility for it being the case that \( p \), as well as “explicitly presenting” the proposition that \( p \).
This second stipulation is added in order to differentiate between what is asserted and what is
presupposed, as Alston is of the opinion (contra Russell) that presuppositions are not asserted
(e.g. *The chair in the corner is broken* presupposes that *There is a chair in the corner* but does
not assert it) (pp.115-120).

As can be seen from the taxonomy above, Alston does not include so-called
“institutional” illocutionary acts in his taxonomy.\(^5\) This is not unusual, of course, as these types
of acts have been singled out as significantly different from the rest going back all the way to
Strawson. But it is relevant to note how exactly Alston’s own definitions of the terms at hand
exclude institutional illocutionary acts, and how his notion of “exercitive” differs. Namely, in an

\(^5\) Alston does not in fact call these “institutional.” I am using this term because they are the acts
that invoke an extralinguistic institution for their successful performance. As can be gleaned
from the paragraph that follows, Alston’s term would be something more like “non-purported
exercitive.”
institutional illocutionary act, the “conventional effect” that the act brings about actually occurs -- it is not only taken responsibility for by the speaker, but it actually happens. This is because institutional acts are the only ones wherein the various conditions that a speaker has to take responsibility for must actually be true (in an objective and extralinguistic sense) for the effect to take place. For example, I can declare that I hereby christen a ship with the name “Boaty McBoatFace,” and in doing so I have taken responsibility for the condition that I have the power to do so. However, even if I actually believe that I have the power to do so, the ship is not thusly-named unless I really do -- according to some objective judge -- have the authority to name it. Alston summarizes this difference by saying that in both cases I purport to christen the ship, but only in the case where I objectively have the authority is the ship actually christened (pp.89-93). Therefore, purporting to perform an institutional act is included in Alston’s taxonomy -- but not the actual performance of one.

Even without Alston’s exclusion of institutional acts from his primary taxonomy, it is a natural move to leave institutional acts out of a taxonomy used to analyze social media data. As we will see in section 2.1.2.1, several other studies that have looked at illocution in CMC have found a lack of institutional acts. In my own pilot, as well, I did not find any acts that I categorized as institutionally verdictive or declarative (see section 2.3).6

Another important element to note is Alston’s treatment of expressives, seeing as there is a wide spectrum of views on this particular category. Essentially, some scholars would treat an

---

6 It is interesting to note, however, that the concept of institutional illocutionary acts on social media has recently gained cultural relevance and attention. The outcome of the 2016 United States presidential election placed a very active Twitter user in a position of objective and extralinguistic authority. Whether, and to what extent, declarations made by this user on Twitter have real-world power will be further discussed in section 4.2.3.
utterance of “I am disgusted by X” as an expressive act, while others classify it as an assertion, arguing that an expressive must be something like “Ugh!” (said in response to seeing X). But Alston’s account considers both to be expressive, because both utterances would be submitted under the same rule, and make the same requirements of the speaker in terms of accountability.

This view of expressives is particularly appropriate for the analysis of written data, I would argue. The idea that “Ugh!” is a true expressive while “I am disgusted by X” is not is related to the notion of expressives as consisting of something like pure emotion, spontaneously exiting our mouths in a moment of panic/passion/surprise/etc. However, none of my data is spontaneous. One does not see a meme on Twitter and spontaneously type “haha,” or see a news headline and spontaneously type “Fuck!” with no control over one’s fingers. Typing those elements does not “express” in the same way as letting them bubble forth uncontrollably from one’s larynx.7

In short, Alston’s treatment of expressives is related to his treatment of exercitive acts. That is, an expressive illocutionary act is one wherein the speaker purports to have a particular psychological state. Real-world evidence (“reliable signs” in Alston’s terminology) that he actually does currently possess that state (such as the spontaneity of an “Ugh!”, or perhaps an uncontrollable vomit of disgust) is not what makes the utterance an expressive illocutionary act, in the same way that adherence to a real-world institution is not what makes an utterance an exercitive illocutionary act. In both cases, the illocutionary acts are purportations. Once

7 To this point, Goffman (1981) differentiates between spoken and written “response cries” (e.g. “Oops!” or “Fuck!”) as well. Spoken response cries convey information through direct expression, “a natural overflowing...of previously contained feelings.” Meanwhile, written response cries convey information through an intentional codification of the signals conveyed by their spoken correlates (p.99-114).
objective facts have taken us outside the realm of purportation, we are outside the realm of illocution and into something else.\textsuperscript{8}

There is one last comment regarding Alston’s view of expressives that seems pertinent to include. On his definition, how is the category of assertives differentiated from that of expressives, since they both involve taking responsibility for a certain psychological state?\textsuperscript{9} The difference is a matter of whether it is my expression of the state or the state itself which can come under fire. That is, an infelicitous expressive could be met with an utterance of \textit{You don’t feel that way} while an infelicitous assertive could be met with \textit{That isn’t true}. Meanwhile, the other nonassertive acts are differentiated from expressives by not only taking responsibility for the fact that I have a certain psychological state (e.g. intent), but by the addition of other stipulations (pp.109-113).

Finally, any discussion of expressives would be incomplete without a consideration of Potts (2007a). Potts details six properties that expressive elements (e.g. \textit{damn, bastard}) exhibit, and proposes that these properties suggest and support an analysis of expressive meaning that places it in a dimension parallel to descriptive meaning. This is important for any discussion of speech acts because, as Potts points out, these expressive elements perform an expressive act simply in being uttered (p.180). This means that an expressive act can be performed alongside a descriptive act, by the same utterance. For example, consider the utterance in example 1.1:

\begin{quote}
(1.1) That bastard Kresge is famous.
\end{quote}

\textsuperscript{8} If my data were not written, it would be key to tease apart the line between purportations to express and expressives of the pure-spontaneous-emotion type. But as it stands, I consider this a non-issue for my study.

\textsuperscript{9} For an assertive, the psychological state involved is a belief in the truth of the proposition presented.
This utterance conveys the expressive meaning that *Kresge is a bastard/bad person* as well as the descriptive meaning that *Kresge is famous* (p.168). Moreover, in terms of denyability (the metric I will use to distinguish direct from indirect acts, see section 1.1.3), both of these acts appear to be direct as neither can be subsequently felicitously denied. Consider example 1.2:

(1.2)  
a. That bastard Kresge is famous.

b. #But I think he’s a good person.

c. #But he is not famous.

Neither 1.2b nor 1.2c are felicitous follow-ups to 1.2a, at least not without shifting the meanings of the words *bastard* and *famous.* This raises an important question for the current project: does the annotation need to specifically address these expressive elements?

Ultimately, incorporating an annotation of the expressive dimension would significantly complicate an already complicated task, drawing precious annotator attention away from the project’s primary research questions. It would also require the merging of multiple frameworks, as the relationships among direct, indirect, descriptive, and expressive would all need to be clearly delineated (and distilled into accessible definitions and guidelines). For these reasons, I will not be concerned with the difference between the expressive dimension and the descriptive dimension in designing the annotation task.

At this point, we may turn to my main disagreement with Alston’s account: though I think his general account of illocution (e.g. taking responsibility for certain conditions, etc.) is

---

10 That is, 1.2b may be felicitous if the speaker’s goal was to point out that they are using *bastard* in a non-standard way, likewise for 1.2c and *famous.*

11 A post-annotation review of this decision further supports it, considering the difficulty that the task exhibited as-is. As we will see in sections 4.2 and 4.3, however, the expressive dimension may have made itself known in the corpus nonetheless.
convincing, his taxonomy is not as theoretically-grounded as that of, say, Searle & Vanderveken’s or Bach & Harnish’s. Alston does not provide a unifying notion or dimension within which the only options yield the final categories (on the other hand, Searle’s direction-of-fit model, for example, does this beautifully). Instead, Alston simply starts with the taxonomy originally provided by Austin and follows later scholars changes to it, before making some final tweaks of his own regarding the precise boundaries and names of the categories. This is not to say that his taxonomy is in some way lacking. After all, it is essentially composed of those same categories that are so neatly motivated by Searle. But, from a parsimony perspective, Alston’s whole proposal would be stronger with a unified dimension (or intersection of dimensions) wherein these five categories are synced with the only five outputs that the dimension or intersection yields.

Finally, it is crucial to ask how Alston addresses non-literal, ironic, and indirect utterances. The first two categories he treats like any other “sentence surrogate,” and proposes that, in these cases, “we move from what was uttered to some sentence not uttered and take the I-rule governing the latter as supplying conditions R’d in the IA performed.” He uses Winston Churchill’s utterance of *An iron curtain has descended across the continent* as an example, saying that Churchill would have made it obvious that a move away from a literal interpretation was necessary to correctly interpret his utterance. So in this way, Alston has avoided the problems of a “literal force” interpretation of illocution. But what exactly it means to “move” from the explicit utterance to another utterance is not well-defined. Especially in the case of non-literal utterances, Alston himself acknowledges that there are no clear instructions for finding the meaning, and that the burden is ultimately placed on the creativity of the hearer. The
“move” is slightly easier with ironic utterances, as there is usually an “opposite” component to their interpretation (i.e. take the literal utterance and find its -- or an -- opposite) (pp.233-234).

In both non-literal and ironic utterances, Alston is taking it for granted that the hearer has identified the utterance as such, evidently because the speaker made it “obvious.” Presumably there are different ways to make it obvious that one is speaking non-literally or ironically, most of which are contextual, but Alston does not really explore the matter. It is simply included under what he calls the “tacit rider” that illocutionary act performances also require the speaker to R that they have provided enough information in such cases (p.136).

Meanwhile, he does not explicitly detail how indirect acts are to be dealt with, other than to state that part of what a speaker must make “obvious” when performing a direct act is that they are not performing an indirect act, and intend a “straightforward, direct use” of a given utterance. But he does not then explore what happens when the speaker does intend to perform an indirect act, other than to acknowledge that there is disagreement as to whether both acts are performed, or only the indirect act (p.186). Because some account of indirect acts is necessary for my annotation task, I will be exploring the issue of direct or indirect next, in section 1.1.3. For the moment, I will consider Alston’s formulations to be sufficient for direct acts, including non-literal and ironic utterances.

In sum, I propose that Alston’s account of illocutionary acts is well-motivated and will provide the theoretical basis for this project. How exactly this choice will affect the form of the annotation task will be further discussed in section 2.1.
1.1.3 Direct vs. Indirect Acts

In this section, I will explore the delineation between direct illocutionary acts and indirect illocutionary acts. To do this, we will have to step outside of Alston’s account, seeing as he does not give indirect acts much attention. However, we will generally be staying within, at the least, the traditional, literalist branch of the speech act family tree. The goal of this section is to arrive at a formulation of indirect acts that respects the foundations of Alston’s theory. It seems therefore misguided, from the start, to entertain any definition of indirect acts that stems from a non-literalist account.

There are many questions that come to mind when we approach the topic of indirect acts. What counts as “indirect”? How exactly does it differ from “non-literal”? How does an interlocutor know that an indirect act has been performed? These are just a few of the issues within this topic.

Let’s begin with the most basic of these questions: what is an indirect speech act? Within traditional speech act theoretic models, the view that indirect speech acts are non-overt acts which are inferred from an utterance based on common ground knowledge and rational judgment is generally accepted, with some variation in the details (Searle, 1975; Clark, 1979; Gordon & Lakoff, 1975; Bach & Harnish, 1979; among others).

Searle (1975) simply says that indirect speech acts are those wherein “one illocutionary act is performed...by way of performing another.” The speaker conveys that he means more than what he explicitly said by relying on the common ground shared with the hearer, as well as assuming that this hearer possesses basic inferential capabilities (pp.60-61). That is, indirect speech acts on this definition are essentially Gricean implicatures. A minor exception is indirect
acts of a highly conventionalized nature, such as indirect requests formed as questions using “can” or “could” (e.g. “Can you tell me the time?”). In those cases, the only inferencing that needs to occur is whether it seems likely that the speaker in fact meant to invoke the convention. For example, if A asks B “Can you tell me the time?”, B infers based on the context whether A meant this as an indirect request to be told the time (i.e. invoking the convention) or instead meant it as an actual yes/no question for information. Alternatively, A could indicate that she did not mean to invoke the convention by cancelling the indirect act (e.g. “I know what time it is, but I’m wondering if you can see the clock from your seat.”) Searle’s view is therefore that the indirect act is “logically contingent” on the direct act, even in the conventionalized cases. The direct act is performed, after which rational judgment, common ground information, and sometimes convention allow the hearer to determine if any indirect acts were also performed.

Clark (1979) takes Searle’s proposal and attempts to experimentally tease apart some of the factors that might participate in the inferential steps between the direct and indirect acts, at least in the particular case of indirect requests. Clark found that conventions of various types, the obviousness of the answer to the direct question, the transparency of the indirect meaning, and the speaker’s goals as inferred by the hearer all seemed to affect that hearer’s interpretation. His findings are summarized in table 1.3, wherein q represents the probability that the direct meaning (Q) was meant seriously (e.g. if a question of Can you tell me the time? is actually meant to elicit a Yes or No response), and r represents the probability that the speaker intended a salient indirect meaning (R) to be taken seriously. Both q and r are estimated by the hearer, and whether their values were perceived to go up or down was based on the hearer’s response to the indirect request (p.439). A response that only answers the direct question exhibits a high q and low r; a
response that responds to both exhibits relatively equal values for both $q$ and $r$; and a response that only addresses the indirect meaning exhibits low $q$ and high $r$.

Table 1.3. Findings from Clark (1979).

<table>
<thead>
<tr>
<th>Response Type</th>
<th>Relative value of $q$</th>
<th>Relative value of $r$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only answers Q</td>
<td>high</td>
<td>low</td>
</tr>
<tr>
<td>Answers both Q and R</td>
<td>equal</td>
<td></td>
</tr>
<tr>
<td>Only answers R</td>
<td>low</td>
<td>high</td>
</tr>
</tbody>
</table>

Clark’s findings support the view that indirect speech acts are probabilistically contingent on, or inferred from, direct acts for two main reasons.

First of all, nonlinguistic information was able to affect the hearer’s judgments of $q$ and $r$. Contextual information, such as the speaker’s goals and/or how obvious the answer to Q was, affected whether or not the hearer responded to Q, R, or both. In fact, we can even put that information in explicitly Gricean terms. Taking the speaker’s goals into account is essentially a type of relevance-based implicature, while not providing an obvious answer to a question would be a quantity-based implicature (“say no more than you need”).

Secondly, in the cases wherein the hearer responds to both Q and R, the responses come in that order. That is, such a response to *Can you tell me the time?* might be *Yes, I can - it’s six*. As Clark points out, the reverse order of *It’s six - yes, I can* would be quite marked (p.435). This suggests the the Q and R meanings themselves have a distinct order of interpretation, wherein R can only be interpreted after Q. In other words, that R is contingent on Q.

We have yet to address the difference between “indirect” and “non-literal.” As previously mentioned, Bach & Harnish (1979) provide a thorough delineation of these terms.
They propose a four-way distinction (two orthogonal spectra) composed of “direct vs. indirect” and “literal vs. non-literal.” The first spectrum characterizes whether or not the act happens by way of another, while the latter characterizes the exact words used. For example, a direct and literal act would be *Please pass the salt*, while an indirect but literal act would be *Could you pass the salt?* Meanwhile, a direct but non-literal act would be *My mind got derailed.* Finally, Bach’s example of an act that is both non-literal and indirect is *I love the sound of your voice* meant “...to tell someone non-literally (ironically) that she can’t stand the sound of his voice and thereby indirectly to ask him to stop singing,” (Bach, 2004, p.470).

Hence, I need to make clear to my annotators that non-literal and indirect are not the same thing (the first being a matter of semantics, and the latter being a matter of pragmatics). In other words, that *My mind got derailed*, meant as a way to say *I got distracted*, is a non-literal direct speech act, and does not inherently entail the presence of an indirect speech act (though it could spawn one, in the right context, as a separate matter of course).

If we step outside of a traditional speech act theoretic model for a moment, we can find other ideas about what indirect speech acts are, as well as how we should attempt to categorize them.

Asher & Lascarides (2001) provide one such conception, using discourse level information in an attempt to eliminate the need to rely so heavily on inferencing. The idea that a hearer has to gauge the likelihood that the direct meaning of a speech act unit was meant to be

---

12 This, of course, rests on the assumption that indirect acts are dependent on direct acts. Bach & Harnish (1979) are indeed of that opinion, though. “[An indirect illocutionary act] is indirect in the sense that its success is tied to the success of the first act. That is, securing uptake requires [the hearer] to identify the indirect act by way of identifying the first act” (Bach & Harnish, 1979, p.70).
taken seriously does not sit well with these authors. In response, they propose an analysis of indirect speech acts (and speech acts more generally) that is free of any mention of speaker intention (p.188). Their proposal is essentially to focus less on notions of illocutionary point (e.g. whether a given unit is *asserting* something vs. *expressing* something) and more on the rhetorical function of the unit (e.g. fine-tuned classifications such as *explanation* vs. *elaboration*, which they deem “subtypes” of assertion). To this end, they develop a formalization that classifies the default rhetorical function of a unit by taking into account the discourse context (preceding statements) at the lexical and compositional level, in addition to more pragmatic information like speaker goals (p.207). The presence of indirect acts is identified when a piece of information does not properly align with the default (e.g. if a speaker goal seems out of line with the default function assigned to a unit).

However, equating speech acts and rhetorical relations seems inadequate for my project. It is not obvious how initial turns would fit into such a theory. As the vast majority of Twitter consists of initial turns -- which are in fact often the ONLY turns in a Twitter “conversation” (e.g. thread) -- this is a non-trivial issue. The discourse level of speech for Twitter is an entirely different animal than that for spoken conversation, and hence cannot be approached in the same way.

Clapp (2009) discusses two more general problems with the rhetorical relation typology. First of all, it is quite easy to come up with examples of speech acts that do not have the purpose of connecting utterances in a discourse. For example, *promising*, *warning*, and *apologizing* do not necessarily have anything to do with discourse structure. While Asher & Lascarides do
acknowledge the existence of such non-relational acts, they do not acknowledge or address the fact that such acts have quite often been the main focus of previous literature (p.44).

This brings us to the second problem, which is that the authors seem to view non-relational acts as a marginal category or minor inquiry. This departure from previous scholars makes the authors’ proposal that their rhetorical relation typology replace the typology of Austin and Searle all the more misguided. Clearly, Asher & Lascarides are interested in acts of a different nature than were their predecessors (p.44). By trying to use the discourse level to bypass the difficulties (especially indirect acts) of the utterance level, Asher & Lascarides also bypassed the utterances themselves. In other words, at the utterance level, a unit could be an assertion while also, at the discourse level, being an explanation.

Additionally, as previously noted, in order for a formulation of indirect acts to be compatible with Alston’s theory of illocution, it seems best to use one that at least stems from a literalist account of direct acts.

In sum, I believe that the best formulation of indirect acts to use for this study is the traditional one: that they are non-overt acts which are inferred from the direct act based on rational judgment, the common ground, and sometimes convention. Seeing as this formulation is not affected by the exact definition of direct act adhered to, there is no particular reason for it not to coalesce with Alston’s theory of direct acts. How this formulation of indirect acts will affect the annotation task will be discussed in section 2.1.3.

1.1.4 Summary

This section has three parts. In the first, I will briefly summarize the definitions of the key terms of “illocutionary act,” “direct act,” and “indirect act” that I have arrived at, based on
the version of speech act theory that I have chosen as the theoretical basis for this study. In the second and third parts, I will discuss how sections 1.1.2 and 1.1.3 have motivated the first three hypotheses of the project.

1.1.4.1 Definitions of Key Terms

In this project, we will say that an “illocutionary act” can be formally represented as $F(p)$, wherein a force, $F$, is applied to a proposition, $p$. This force is a matter of the socially entrenched rule that a speaker submits their utterance to. In submitting their utterance so, the speaker has taken responsibility for the conditions of this rule. In taking this responsibility, the speaker has felicitously performed the associated illocutionary act. The utterance need not be meant seriously or literally for the act to be felicitously performed, as a speaker may choose to take responsibility for conditions that they know do not hold. However, in such cases, the speaker must convey “sufficient clues” that the utterance is meant non-literally and/or non-seriously. Thus, we can set out the following definitions:

- An “illocutionary act” is the act that results from a speaker submitting their utterance to a given socially entrenched rule while taking responsibility for the conditions of this submission.
- A “direct illocutionary act” or “direct act” is the straightforward and overt act performed by an utterance.
- An “indirect illocutionary act” or “indirect act” is any non-overt act inferred from the direct act based on rational judgment, the common ground, and possibly convention.

How these definitions affect the annotation task, and how I expect annotators to judge the difference between direct and indirect acts, will be discussed in section 2.1.
1.1.4.2 Motivations for Hypotheses 1 and 2

I have said that I intend to investigate illocutionary acts on Twitter by constructing a corpus of Twitter data, annotated for direct and indirect speech acts. And indeed, building a corpus using the definitions summarized in section 1.1.4.1 will provide an interesting tool for such an investigation. But what exactly am I investigating? I have yet to define any specific hypotheses for the study.

The primary hypothesis of an investigation of illocutionary acts, I propose, should address a primary disagreement that repeatedly occurred in section 1.1.2: whether, and to what extent, direct illocutionary force is overtly encoded, as opposed to a matter of speaker (or speaker and hearer) intention and/or inference. Alston speaks of direct acts generally as “straightforward” and overt, while Searle & Vanderveken speak of an “explicit piece of language” that indicates the illocutionary force. Can we find evidence of such an explicit, straightforward, and overt piece of language? Because Alston does not provide a name for such an element, I will borrow Searle & Vanderveken’s term to use in the exact formulation of this primary hypothesis, given in 1.3:

(1.3) Hypothesis 1: Certain overt elements available to Twitter users function as direct illocutionary force indicating devices (IFIDs, from Searle & Vanderveken (1985)).

If hypothesis 1 is supported, it would mean that illocutionary force is at least partially overtly encoded. If it is not supported, this could mean that illocutionary force is not overtly encoded, that some force markers are overt but they were not tested for, that the underlying theory was poorly chosen, and/or that there was a flaw in the annotation task.

Meanwhile, section 1.1.3 suggests a second hypothesis: what about indirect acts? These acts have not received nearly as much attention in the literature as have direct acts, despite a
general consensus among traditional speech act theories that these acts exist. According to the formulation in 1.1.4.1, the process of inferring an indirect act from a direct act may involve conventional elements. If this is true, the nature of such elements is not necessarily clear. They may or may not be overtly identifiable, though Clark’s (1979) study did identify one such possible element: whether or not the direct act was meant seriously. Specifically, what he found was a clue of indirect act presence, not necessarily indirect act force. Can we identify any others in the corpus? Hypothesis 2 is given in 1.4:

(1.4) **Hypothesis 2:** Certain overt elements available to Twitter users signal the presence of an indirect act.

If hypothesis 2 is supported, this may help illuminate the extent to which overtly identifiable elements contribute to the inferencing process that occurs between the direct act and the indirect act. If it is not supported, this could mean that indirect act presence is not overtly encoded, that some overt signals exist but they were not tested for, and/or that there was a flaw in the annotation task.

Hence, hypotheses 1 and 2 seek to uncover overt clues to direct act force and indirect act presence, respectively. By analyzing both direct and indirect speech acts on Twitter, a written medium wherein any possible overt clues should be recorded, I hope to be able to shed light on the ultimate importance of such overt elements in speech act interpretation.

**1.1.4.3 Motivation for Hypothesis 3**

If we step back and examine hypotheses 1 and 2, additional questions arise. One of the most salient is that they are, of course, focused on Twitter. The extent to which results found for language on Twitter could translate to other CMC, let alone non-computer-mediated modes of communication, is not immediately clear. In other words, what is the potential generalizability
for the results of this study? Are the results for Twitter potentially representative of other CMC modes?

The literature would seem to suggest not. In short, computer-mediated communication is not a communicative monolith, either in social discourse or in formal grammar. Herring & Androutsopoulos (2015) note that the different modes of CMC are “socially as well as technologically defined, each having their own unique affordances, histories, and cultures of use,” (p.129). These unique cultures of use can lead to formal differences, as “…in subcultural and interpersonal contexts of intensive, self-contained interaction…[CMC grammar] may evolve at an accelerated pace,” (Herring, 2012, p.7).

In other words, Twitter should be considered its own linguistic community. Instead of looking for ways to generalize the results of this study to other modes, it may be more informative to focus on comparison. Hypothesis 3 is given in 1.5:

(1.5) **Hypothesis 3**: We will see a speech act landscape for Twitter that is distinct from that of other social media and certain spoken genres.

Comparing the distribution and behavior of speech acts on Twitter with such in other media would illuminate the differing, or similar, communicative purposes of these media. It may also help elucidate which conditions are absolutely necessary for the performance of certain acts, conditions which may or may not be available on Twitter. The specific stipulation of comparison with “other social media and certain spoken genres” is meant to cover the variety of studies for which I was able to find comparable data (see section 2.1.2).

1.2 **Overview: Technological Affordances of the Medium**

The goal of this section is to determine whether certain technological affordances perform speech acts or affect speech act interpretation, which will be relevant when deciding on
the exact form of the annotation task in chapter 2. Twitter, as a medium under the umbrella of CMC, possesses several features that face-to-face conversation does not. While some of these features will not be relevant until the post-annotation analysis, several of them need to be assessed with regard to speech act status, as they may affect the annotation task.

The technological affordances include: punctuation, capitalization, explicit quotes (as indicated by quotation marks, other punctuation, or the presence of a quote attribution), emoticons/emoji, tags (both hashtags and @-tags), the “like” button, and the “retweet” button. Only those features which may in fact perform a speech act by themselves need to be addressed in this section, which means that capitalization and punctuation need not be addressed here. These two elements will, however, be considered potential IFIDs in section 4.2.1 and potential indirect act presence indicators in section 4.2.2.

Thus, the four groups of technological affordances I will address in this section are emoticons/emoji, tags, the “like” and “retweet” buttons, and explicit quotes. For each element, I will first review background information and literature about it. I will then determine, based on that background, the speech act status of that element: whether the element should be considered to perform a speech act by itself, and/or whether it may affect the interpretation of an adjacent segment (meaning it should be considered a potential IFID or indirect act presence indicator).

---

13 The test here is whether a tweet, or tweet-like interaction (i.e. like or retweet), could consist of ONLY that element. A tweet cannot, of course, consist of only capitalization. A tweet could, theoretically, consist of only a punctuation mark. In such a situation, however, that punctuation mark would still be punctuating something (a previous tweet, an event in popular culture, an implicit comment). If such a situation arises in the corpus, I will present it to the annotator like any other segment that consists of [content+punctuation mark].
1.2.1 Emoticons and Emoji

1.2.1.1 Background

Emoticons (faces and other pictures made out of traditional keyboard characters) and emoji (small, pre-made, and colorful pictures placed into text using unicode shortcuts or a separate keyboard) can be found in every type of CMC. Given this ubiquity, it is no surprise that numerous scholars have investigated whether, and to what extent, these elements can influence the interpretation of a message. I will presently overview these studies, but to preview their findings: yes, emoticons and emoji can indeed affect the interpretation of a message, in sometimes profound ways.

As we will see, there are two general categories of emoticon and emoji usage: 1) affecting the interpretation of an adjacent piece of text, or 2) standing in for a word or concept. The former has been studied much more than the latter, likely due to the fact that the newer technology of (detailed and varied) emoji has only recently made category two more of an option.

1.2.1.1.1 Affecting Interpretation of Adjacent Text

There have been quite a few studies that ask whether and to what extent emoticons and/or emoji can affect the interpretation of adjacent text, but many of them approach the question in a similar way. Specifically, they tend to investigate the effect of emoticons through valence interaction (e.g. [how] does the presence of a positive emoticon affect the interpretation of a negative/positive/neutral message, etc.), often with an explicit focus on how valence interactions yield or support sarcastic/ironic interpretations.
Derks et al. (2008) is a good representative sample of such a study. The authors looked at emoticon usage in emails by pairing the same message with a variety of emoticons and asking for native speaker judgments about what each utterance conveyed. They found that varying the valence pairing of the emoticon and the message could strengthen the message, create ambiguity, or express sarcasm. The valence of the message itself had a large effect on how the message was rated -- that is, positive messages were mostly rated as positive, and so forth. But emoticons could affect the strength of the positive/negative rating. A matching valence emoticon could strengthen the message, but only on the positive end of the scale (i.e. a positive emoticon tended to increase the perceived positivity of a positive message, but the same is not true for negative emoticons/messages). Meanwhile, emoticons with a valence opposite to that of the message had several effects on message interpretation. Such messages were judged to be more ambiguous and more sarcastic, but also to have an overall weakened valence. That is, a positive message paired with a negative emoticon was judged to be less positive than the pure positive message, but more positive than the pure negative emoticon (and vice versa). Finally, the authors also found that emoticons paired with a neutral message could shift the perception of the message toward the valence of the emoticon (pp.384-385). Their most general conclusion is that, “…to a large extent, emoticons serve the same functions as actual nonverbal behavior,” (p.379).

Lo (2008) also focuses on the “nonverbal” aspect of the information that emoticons tend to convey, using the same basic methodology as Derks et al, and coming to the same general conclusion. The author makes the assumption that, in spoken communication, “the goal of nonverbal cues is to convey emotional messages, and the goal of verbal cues is to communicate

---

14 This is a common finding in these studies, as will become clear.
ideas” (p.595). Lo therefore hypothesized that the presence and valence of an emoticon accompanying an instant message would affect the reception and perceived valence of emotion, attitude, and attention. These hypotheses were supported. Luor et al. (2010) came to essentially the same conclusions about the effects of emoticons on message interpretation in task-based instant messaging situations.

Dresner & Herring (2010) come to similar conclusions about the functions of emoticons in their instant message data, but loosely place these conclusions within a speech act framework. The authors find three basic uses of emoticons: 1) emotional meaning indicators, “mapped directly onto facial expressions,” 2) non-emotional meaning indicators, “mapped conventionally onto facial expressions,” and 3) “illocutionary force indicators that do not map conventionally onto a facial expression” (p.250). For example, the first usage would be shown by a smile emoticon accompanying a positive expressive message, contributing to the expressive illocutionary force. The second usage would be shown by a wink emoticon accompanying a message meant as a joke, seeing as a wink in face-to-face conversation tends to have the same meaning. Finally, the authors provide the following example for the third usage: “I would like a noncircumventing solution ;->”. The wink emoticon here is not meant to represent an actual wink, because the author is not joking. Instead, the wink acts as a mitigating element, indicating that this message should not be interpreted as an order or request, as would seem to be suggested by the phrase “I would like.” The wink suggests something less face-threatening, such as a mere assertion (pp.256-257).

Provine et al. (2007) explore the ability of emoticons to affect the interpretation of adjacent text from a more unique perspective. Specifically, they investigate whether emoticons
adhere to the “punctuation effect” observable in the distribution of laughter in speech, which refers to the fact that laughter occurs:

...during pauses, at phrase boundaries, and before and after statements and questions -- the places where punctuation would be placed in a transcript of a conversation. Such punctuation indicates that language is dominant over laughter in competition for the vocal tract because laughter seldom interrupts spoken phrases (p.299).

The authors found the same distribution for emoticons in their data, a corpus of online message board posts, finding that emoticons appeared in the middle of a phrase in only 1% of cases (out of a total of 836 cases, p.303). They also point out that, since there are many more opportunities for mid-phrase emoticon placement than phrase-terminal placement (i.e. there is an opportunity between every single pair of words in the phrase), the fact that emoticons do not get placed mid-phrase is significant. Additionally, this “punctuation effect” of emoticons is more pronounced than that of laughter because they are not competing with either vocalizing or respiration -- there are no physical constraints on the placement of emoticons. The authors interpret this to mean that language is ultimately dominant over “emotional expression” in CMC, the way that it is dominant over laughter in speech. However, as studies like Dresner & Herring (2010) point out, emoticons are not always meant to convey emotional meaning. I would therefore reframe the conclusion of Provine et al. (2007) to say that this is another study that supports the analysis of emoticons as resources that are conventionalized as opposed to decorative or devoid of communicative power.¹⁵

¹⁵ Notably, the idea that language is dominant over emotional expression in CMC would support the idea that emotional expression in CMC is purported, as opposed to “truly” expressive (i.e. typing “ouch” or a frowning emoticon is purportation to expression, as opposed to spontaneously saying “ouch” when you stub your toe, which is “truly” expressive). This would therefore support classifying “statements” that express as expressive acts in CMC because they are just as expressive as interjections in CMC -- both are purposefully-produced purportations.
Riordan (2017) tackles a different corner of this subject, being the only study I found in this area to explicitly focus on emoji of objects, as opposed to emoticons or emoji of faces. Specifically, in a series of two studies, the author investigates whether emoji of objects convey speaker affect. In the first study, Riordan had college students rate positively and negatively valenced texts which had been paired with 0-3 emoji. The students rated the overall message for valence, which of 8 possible emotions they detected and how much of it, and how confident they were in their rating. The results showed that, regardless of the valence of the message, the presence of nonface emoji increased the amount of positive affect detected by the raters, with a slight effect of “more emoji = more positive” (pp.559-560). In the second study, Riordan essentially repeated the experiment but used Amazon Mechanical Turk to access a “larger and nonstudent population” of raters. The results yielded the same overall trends (p.562).

Many other studies have found that emoticons and/or emoji generally have the power to affect the interpretation of adjacent text. Some have even pointed out that the law agrees with such an assessment. While several of these studies note specific patterns of usage for emoticons/emoji (e.g. mixed-valence between the emoticon and the message is meant to convey sarcasm, etc.), for the purposes of my study, the relevant finding is merely that emoticons/emoji


17 Pelletier (2016) and Browning & Seale (2017) are two such studies. In the latter, the authors discuss a case from 2014, wherein a Detroit city official accused internet message board commenters of libel. In that case, the court ruled that the commenters were not guilty of libel because it was “patently clear” that they were joking, as evidenced by the presence of tongue face emoticons, which a “reasonable reader” would interpret as non-serious (p.18).
have the power to affect the interpretation of adjacent text in a general capacity, regardless of the exact convention used. An emoticon/emoji used in this way is therefore a potential IFID.

It is worth noting that the ability of emoticons/emoji to affect the interpretation of adjacent text also designates them as an example of what Gumperz (1992) termed “contextualization cues.” He proposes that certain elements of language function as cues that situate an utterance within the local context, allowing a hearer to “retrieve the presuppositions they must rely on” to assess what the utterance is saying and doing in context (p.229). This function is similar to that of IFIDs. Given that emoticons/emoji, and several other elements posited as contextualization cues by Gumperz (e.g. intonation, stress, pauses, lexical choice) are also posited as potential IFIDs, it is worth asking: is “contextualization cue” another name for “IFID”?

I would propose that IFIDs are not equivalent to contextualization cues, but instead form a subset of them. Specifically, IFIDs are contextualization cues that help retrieve one specific type of presupposition (i.e. the presupposition that an utterance of X being spoken in the form it was, in the context it was, is meant to be interpreted as an act of Y category).

1.2.1.1.2 Standing in for Words/Concepts/Phrases

A few studies have also addressed emoticons and emoji that stand in for words or concepts, whether alone (as a separate segment) or incorporated into a larger semantic unit. Garrison et al. (2011) mention the topic briefly in their discussion of emoticons in IM discourse. The authors give samples from their data wherein a question or comment is responded to with only an emoticon, which then elicits a response from the questioner commenter. While they do not go as far as saying that the emoticons in these examples are standing in for particular
words/phrases/concepts, they do acknowledge that these emoticons seem to have “rhetorical significance” (p.121). The authors express doubt that emoticons could go beyond this simple “rhetorical significance” however, saying that an entire conversation taking place in only emoticons seems “highly improbable” (p.122). And perhaps an entire conversation in only emoticons is indeed unlikely. But as Browning & Seale (2017) point out, an entire conversation in only emoji is not only possible, but accessible enough and in high enough demand that there is now an emoji-only social network (called, appropriately, Emoji; p.14).

Emoji do not have to appear on an emoji-only platform to be interpreted as standing in for words, phrases, or ideas. Browning & Seale also mention several legal cases wherein emoji in texts or internet posts were argued to constitute meaningful communicative segments that provided crucial evidence.18

In short, using emoji (and, to a lesser extent, emoticons) to stand in for words, ideas, or complete messages -- as opposed to merely affecting the interpretation of nearby text -- is an established and widespread practice.

1.2.1.2 Speech Act Status

In light of existing research, emoji and emoticons in my data have three possibilities available to them: 1) standing in for a word or series of words that is semantically incorporated into the preceding/following text (e.g. *I love her 😊* to mean *I love her smile/face*); 2) standing in for a word or series of words that is not semantically incorporated into the

---

18 For example, in Beaver Falls, Pennsylvania, a man accused of homicide had his text messages used as evidence of intention to commit gun violence. The texts read “we’re going to hit that lick” and “Let’s get at it,” followed by emoji of a running man, an explosion, and a firearm. Similarly, in 2015, a young man was arrested and charged with threatening police officers after he posted several Facebook statuses featuring references to particular police precincts followed by emoji of firearms pointed at emoji of police officers (p.16).
preceding/following text, and instead represents a separate unit (e.g. I love herיחוד to mean
something like I love her, fuck off); and 3) affecting the interpretation of the preceding/following
text (or both) (e.g. I love her 😉 to mean something like I don’t really love her).

Given the uses and communicative power of emoticons and emoji, where do they fit into
Alston’s speech act framework, and into a F(p) model? We must look at each broad category of
usage separately.

1) Situations wherein an emoticon/emoji is standing in for a word/phrase are easily analyzed
in Alston’s framework, though we should address “incorporated” and “unincorporated”
types separately. I will hereafter refer to this type of emoticon/emoji as “lexical.”
   a) For incorporated emoticons/emoji of this type (e.g. let’s go get some 🍻 and 🍔
   ), they are merely signs of the same type as any other word in the sentence. They
might be slightly more ambiguous (e.g. the previous example might “translate” to
let’s go get some beers and a burger or let’s go get some drinks and food, etc.), but
ultimately they are primarily participating in the construction of p, the semantic
content of the utterance. The choice to use emoji, and/or the particular choice of
emoji, likely also contributes to the force of the utterance, F, just as any other
word choice does.
   b) For unincorporated emoticons/emoji that are meant to represent complete
   segments, the main difference is that they constitute all of p, instead of only
   contributing to it.

2) For emoticons/emoji that are not meant to stand in for a word/phrase and are only
   affecting the interpretation of adjacent text, they are not contributing to p at all, but only
to F, and to the general interpretation of the segment. I will hereafter refer to this type of emoticon/emoji as “tone-type.”

Thus, lexical emoticons/emoji will be treated like any other word in the corpus. Tone-type emoticons/emoji will be considered potential IFIDs and potential indirect act presence indicators in the analyses of hypotheses 1 and 2. The judgment of which emoticon/emoji are lexical and which are tone-type, as well as which are incorporated and which are not, will be made by me during the segmentation phase of the corpus. This will be discussed further in section 2.2.

1.2.2 Hashtags and @-tags

1.2.2.1 Background

Given the prominence of hashtags in current popular culture, it is no surprise that the various functions of hashtags on Twitter have received quite a bit of attention in the literature. I will first overview these functions, dividing them into two primary categories which I will term tagging usages and non-tagging usages. I will then assess these usages from the perspective of speech act theory, so as to determine whether hashtags should be considered to perform speech acts.

Meanwhile, the other type of tag available on Twitter, the @-tag, has received significantly less attention (likely because of the lack of creativity involved in using an @-tag, which is essentially a technologically-enhanced vocative). Nonetheless, we will see that @-tags have also been argued to have both a tagging usage and a non-tagging usage. Hence, @-tags will be discussed alongside their more popular cousin.
1.2.2.1 Tagging Usages

What I have termed a *tagging* usage is one that is actually meant to invoke the technological affordance of tagging: a hashtag or @-tag that is intended to index the tweet to a given Twitter page (i.e. the search result page for that hashtag or the profile page for that @-tag). A tagging type hashtag or @-tag is one that a reader might actually click on or search for, in order to see other tweets on the same topic or about the @-tagged user, or that a speaker might use with the intention of reaching a certain user/audience. As I will discuss in section 1.2.3.2, these tags rely on the institution of technology in the same way that a judge’s verdictive utterance relies on the institution of law.

Within the tagging category, I have distilled four sub-types of hashtags, and one of @-tags, from the literature. I will address @-tags first.

An @-tag that is meant to tag really only has one purpose: to gain the attention of the @-tagged user or to gain the attention of that user’s followers (or both). In this way, it is the equivalent of loudly saying someone’s name in a room full of people. That person can hear you, and so can everyone else. Either party, or both, could be your intended audience. Bruns & Moe (2014) liken @-tagged exchanges to having a conversation within earshot of uninvolved bystanders that could potentially enter the exchange if they wanted (p.20), while Zappavigna (2011) simply refers to the @ symbol as a deictic marker (p.790). She points out that an @-tag that occurs at the beginning of a tweet often functions as a vocative, while a non-initial @-tag is likely more referential than vocative. However, in both cases, the tagging functionality works the same way (that is, the tweet is called to the attention of the @-tagged user and their followers). Consider the @-tags in examples 1.6 and 1.7:
In both examples, the @-tags ensure that both John and his followers can see the tweet. The syntactic difference between 1.6 and 1.7, the fact that the @-tag in 1.6 seems to be in vocative position, is a matter for segmentation (see section 2.2.2). The actual tagging functionality does not differ between 1.6 and 1.7. In short, the original purpose of the @-tag seems to have remained the primary purpose.

In the case of hashtags, the original purpose has certainly stuck around as well, though not necessarily as the primary usage. It is the first of the four sub-types of tagging type hashtag usages: topic marking and information organization. Heyd & Puschmann (2017) state that Twitter user Chris Messina is considered the first to propose the use of hashtags to mark “channels” on Twitter, a convention borrowed from internet relay chat (another form of CMC), in August of 2007. Other users quickly heeded the suggestion, and in 2009 Twitter made hashtags into clickable hyperlinks (p.55). The idea was to hashtag your tweet with the main topic or category, to help organize information on Twitter and allow users to find tweets in their category of interest. Consider the hashtag in example 1.8, from Zappavigna (2015):

(1.8) Channel 7, reporting that it’s alleged that channel 7 could be paying Corby two million for her story. #SchapelleCorby (p.5)

The tweet in 1.8 is about a news story involving a person named Schapelle Corby, and the hashtag categorizes it as such. The literature confirms that this functionality remains a primary usage of hashtags (Avelar, 2015; Bruns & Moe, 2014; Bruns & Burgess, 2011; Caleffi, 2015; Cislaru, 2015; Heyd & Puschmann, 2017; Huang et al., 2010; Page, 2012; Scott, 2015; Wikstrom, 2014; Yang et al., 2012; Zappavigna, 2011, 2015).
The three other sub-types of tagging type hashtag usages all inherently invoke this first usage, but expand on it in different directions.

The second sub-type is socio-relational: constructing communities and declaring community membership. Bruns & Burgess (2011) call this the creation of “ad hoc publics,” ephemeral communities united by the usage of a given tag (p.1). When such a tag is created, a user can then use the tag to not only index their tweet to the topic page (the original hashtag usage), but also to declare their membership in the community. Zappavigna (2011, 2014) uses the term “ambient affiliation” for this loose association, showing how this type of hashtag often involves establishing a set of community values, when users couple the hashtag with tweets that center around a given ideational stance (2011, pp.799-800). She gives the following example, noting that the hashtag #obama makes the tweet’s topic searchable, and couples it with a clear positive evaluation.

(1.9) From fear, hatred and economic collapse to hope, search for common ground and prosperity again. Good change! #obama. (p.799)

Likewise, Wikstrom (2014) gives the example of the hashtag #MayTheForceBeWithKatie, a tag created to show support for a girl bullied for her love of Star Wars. Using this hashtag is akin to the “chanting of a slogan,” as it makes the user’s support and affiliation with the community clear while also drawing attention to it.

(1.10) Don’t let those boys get you down#maytheforcebewithkatie (Wikstrom, 2014, p.134)

Zappavigna adds that the use of hashtags to signal community membership and/or ideational stance is part of a trend of “interpersonal punctuation,” such as using commas to mark clause boundaries and pauses, or “a smiley emoticon to mark positive affect” (p.803). A variety of
language has been used to describe this function of hashtags,\textsuperscript{19} but the literature agrees that affiliation is a primary purpose of tagging type hashtags.

The third sub-type of tagging type hashtag usages is oriented around play: hashtag games, or “micro-memes” (a term used by Huang et al., 2010). These micro-memes serve the “purpose of establishing a simple but recognizable and unique premise, and being part of a template that must be adhered to for successful participation,” (Wikstrom, 2014, p.135). In these cases, the user uses the hashtag: 1) to mark what game the tweet is part of (i.e. the topic), and 2) to participate in the game by making their contribution searchable. Consider the example in 1.11, wherein the hashtag explains which game the user is participating in (\(#BoringPrequels\)) while also making the tweet searchable under that tag so that other users wishing to peruse contributions to the game may find it.

\textbf{(1.11)} Some Like It Tepid \(#BoringPrequels\) (Wikstrom, 2014, p.134)

This participation/game usage is related to the previously-mentioned affiliation usage, and it is possible if not likely that a given hashtag might be meant to perform both tasks, but they are indeed separate. For example, the hashtag \(#SpookyTalesForLinguists\) is a hashtag game wherein users tweet Halloween-themed linguistic puns. Tweeting such a pun and adding the hashtag results in participation, but a user can also mark affiliation with the associated ad hoc public without participating in the game. For example, consider the following tweet:

\textsuperscript{19} In addition to “signalling” and “performing” affiliation, authors have addressed this functionality of hashtags by mentioning how they can “support movements” (Caleffi, 2015, p.49), “construct...identity” (Page, 2012, p.184), and be a “symbol of a community membership” (Yang et al., 2012, p.1).
The author of this tweet is not actually participating in the game (they do link to a blog post that includes several other users’ contributions, but the hashtag itself does that, too). Instead, the author is affiliating with, and establishing ideational alignment with, the associated ad hoc public.

The fourth and final sub-type of tagging type hashtag usages is strategic: increasing visibility. This usage is exactly as described -- a user adds a hashtag to their tweet with the intent to gain visibility for their tweets and their profile. As with the affiliation and participation usages, the visibility usage inherently invokes the topic indexing usage. However, it does so somewhat vacuously, exploiting the indexing usage without concern for the identity of the tag itself. Bruns & Moe (2014) mention this usage, commenting that a user might include a certain hashtag “speculatively,” not actually caring about joining its associated ad hoc public, but merely seeking to gain visibility (p.18), while Bruns & Burgess (2011) term such indiscriminate hashtag usage “Twitter spam,” (p.5). They provide the example in 1.12, wherein the user includes numerous hashtags, several of which are generic in a way that makes them unlikely candidates for ad hoc publics (e.g. #japan) but still able to increase the searchability of the tweet.

(1.12) #japan #tsunami is the real killer. #sendai #earthquake PGA only 0.82g. 2011 #chch #eqnz 2.2g (p.5)
It can be difficult, and sometimes impossible, to tease the different hashtag usages apart. A tagging type hashtag might be meant to accomplish any combination of indexing, affiliating, participating, or exploiting. Additionally, a given tag might invoke any number of non-tagging type functions alongside the tagging type usages. For example, consider the following tweet taken from Page (2012):

(1.13) Oh come on, announce the squad already #england #worldcup (p.184)

My interpretation, as a Twitter user, is that the hashtags in this tweet are at least marking the topic (indexing) and demonstrating the author’s affiliation. But it is also possible (if not likely) that they are also meant to aid the interpretation of the tweet, a non-tagging function that I will address in the next section, seeing as the tweet itself is rather bare in terms of context.

Section 1.2.3.2 will discuss how these different usages should be interpreted in speech act theory, and how we might deal with the issue of multifunctionality in the annotation task. For the moment, I will move on to the non-tagging usages.

1.2.2.1.2 Non-tagging Usages

As is evident from their name, non-tagging usages are those that are not intended to invoke the actual hyperlinking ability of tagging. A solely non-tagging hashtag or @-tag is not intended to be searched for or clicked on, and is not placed into the tweet with the intention of attracting the attention of a certain audience. As I will discuss in section 1.2.3.2, these tags mockingly invoke the institution of technology, in the same way that someone might mockingly invoke the institution of marriage during a play wedding.

I have distilled three main types of non-tagging hashtag usages, and one type of non-tagging @-tag usage, from the literature. As before, I will address the @-tags first.
Bruns & Moe (2014) claim that, even though @-tags are primarily used to explicitly gain the attention of a certain audience, they are also sometimes used emphatically, as if the @-tag were simply an alternative name for someone, without the express purpose of gaining that user’s (or their followers’) attention. Consider the example that Bruns & Moe provide in 1.14:

(1.14) I support @BarackObama (p.20)

The authors acknowledge that the distinction between tagging type and non-tagging type @-tags would be difficult, if not impossible, for anyone other than the tweet’s author to make. Furthermore, such a non-tagging usage may always be accompanied by the slight hope that the @-tagged user does respond (e.g. a slight hope that Barack Obama actually replies to the tweet in 1.14), making it an even more nebulous designation.

The different types of non-tagging hashtag usages are more solidly attested in the literature than non-tagging @-tags. The first one I will address is possibly the least robust of the three, but is the most logical segue from the @-tags: emphatic non-tagging hashtag usage. Again, such a usage is akin to bolding, italicizing, or capitalizing the text, without the express purpose of invoking the technological affordance of tagging (Bruns & Burgess, 2011, p.5; Wikstrom, 2014, p.142). Scott (2015) terms this particular functionality of hashtags
“highlighting,” comparing it to contrastive stress, pointing, or “ostensive gazing” (p.14).

Wikstrom provides the following example of such tags from his data:

(1.15) @user Elric only has ONE sword. #Check #Mate (p.143)

The other two types of non-tagging hashtag usages are related, and can both be placed under the sub-heading of evaluative, a term taken from Page (2012). These usages have garnered quite a bit of attention in the literature, likely because they are inherently playful and creative, and because non-tagging usage in general is an emergent and non-original usage of hashtags.

The first of the evaluative usages is emotive stage directions such as #facepalm, #sigh, or #sobbing. Bruns & Burgess (2011) compare hashtags used this way to emoticons, noting that they “(often ironically) express the sender’s emotional or other responses...they are used to convey extratextual meaning, in a Twitter-specific style,” (p.5). Wikstrom (2014) likewise suggests that such tags are “meant to represent a face-to-face paralinguistic cue,” (p.140). He gives the following example of such a usage:

(1.16) Dont feel like walking...but ill make it #sigh (p.140)

---

20 Emphasis is also claimed to be a primary purpose of offline hashtags, especially of written ones. Caleffi (2015) found this usage to be prevalent in a sample of offline advertisements, postulating that “The message takes a prepackaged and condensed form that one can read as a whole, the preceding # symbol functioning as a pre-positioned exclamation mark, possibly aiming at producing catchy formulations,” (p.66). Likewise, Heyd & Puschmann (2017) concluded that offline hashtags in urban spaces are often simply “decorative” or “emblematic.” They are “visual evocations” of the practice of hashtagging. This analysis is supported by the fact that such tags often violate the typographic specifications necessary for a string of text to invoke the technological affordance of tagging. For example, the authors give the example of a sweatshirt featuring this tag: #I’m looking for trouble (p.61). Actual tagging type hashtags allow neither spaces nor apostrophes.
The second type of evaluative (and third type overall) non-tagging type hashtag usage is the metacommentary usage. This usage, and how it should be analyzed, has generated the most discussion in the literature. Caleffi (2015) describes this type as “a contextual aside to comment on, give more depth to, or somehow emphasize what has been said,” (p.49). Heyd & Puschmann (2017) focus on the non-tagging aspect of this type of hashtag, describing them as “metacommentary unlikely to be used as a search query by another microblogger,” (p.56). Wikstrom (2014) provides the following example of such a usage:

(1.17) Fo Realz, it is blizzarding. #AndiLikeIt (p.140)

Kunneman et al. (2015) focused on the paralinguistic aspect of these tags, finding that the presence of #sarcasm appended to a tweet reduced the likelihood that other textual markers of sarcasm would appear in the tweet. The authors hypothesized that such hashtags are “the digital extralinguistic equivalent of non-verbal expressions that people employ in live interaction,” (p.500).

Both Scott (2015) and Matley (2018) also focus on the ways in which meta-commentative hashtags aid interpretation of adjacent text. Scott suggests “that hashtags have been appropriated by users of Twitter to provide additional contextual information in an economical and stylistically unobtrusive way,” (p.9). That is, the brevity of Twitter posts combined with the absence of traditional non-verbal clues that is inherent to CMC, as well as the fact that a given tweet will be read by different people at different times and in different places, together result in a strong need for interpretive aid. Hashtags that are unlikely search terms instead serve as “a guide to the reader’s inferential processes,” (pp.12-13).²¹ Scott provides the

---

²¹ In at least one court case, the law agrees that hashtags can be meant to guide the reader’s inferential process. In a defamation case, “…the court...found that hashtags contained within [a]
example in 1.18, wherein the two hashtags appear to overtly state the feelings that the user is merely alluding to in the body of the tweet:

(1.18) I think all drs should be made to lie in a hospital bed wearing PJs & be stood over. See what it feels like. #vulnerability #powerbalance (p.13)

Matley (2018) took a speech act oriented approach to this type of hashtag, looking at the use of #brag on Instagram, and suggesting that it “functions as a meta-comment on the interpretation of the accompanying discourse...[and] can be read as an illocutionary force indicating device...that places the illocution of self-praise on record and makes the speech act of ‘bragging’ explicit,” (p.32). I disagree with this exact formulation, as will become clear in section 1.2.3.2, but for the moment it is enough to say that meta-commentative hashtags tend to direct or aid the interpretation of adjacent text, often by adding additional information.22

While the boundary among hashtag usages are blurry in general (due to the previously noted multifunctionality), the boundary between the two types of evaluative non-tagging type hashtags is especially muddy -- and often a given tag is better classified as simply “evaluative” than either emotive or meta-commentative. For example, #brag might be construed as a stage direction type tag, or as a meta-commentative tag, but either way it is evaluative and pragmatically relevant to the interpretation of the adjacent text.

____________________

22 Scott (2018) also found the meta-commentative function to be the primary usage of offline spoken hashtags.
1.2.2.2 Speech Act Status

The flow chart in figure 1.3 summarizes the breakdown of tags that I have outlined in sections 1.2.3.1.1 and 1.2.3.1.2. The gray boxes are potential terminal nodes. That is, a given hashtag usage could be classified (for example), as emphatic, evaluative, emotive, or meta-commentative, but not simply “non-tagging.”

Figure 1.3. Tag flow chart.

How exactly can we put these various usages of tags into a speech act theoretic framework? I will begin with hashtags, and with the one element that unarguably unifies all of them, the # symbol itself.

In short, I propose that # is best analyzed as an explicit IFID of exercitive force, like a technological version of the word “hereby.” When attached to a piece of text on Twitter, the hashtag symbol effects a change in that text, a change reliant on the affordances of technology. That text is now indexed to a given page of Twitter, the one that appears when a person clicks on the hashtag. In other words, the addition of # effects a change from plain text to hyperlink. A user may choose to put the hashtag symbol in front of a piece of text, knowing that it will effect
this change in the world, just as a speaker may choose to place the word “hereby” in their utterance (with the knowledge that it can likewise effect a change in the world).

The next question is: if the direct force of a hashtag is exercitive, what is the content, \( p \)? If we paraphrase the exercitive act of a hashtag like #WorldCup2016 into a sentence with “hereby,” it would be something like “I hereby make ‘WorldCup2016’ into a hyperlink.” Hence, the content of the direct act of a hashtag would be “Make \( X \) into a hyperlink,” where \( X \) is the string of characters appended to the symbol.

At this point, we may begin to bring in the various usages I’ve outlined above. The different sub-types, I propose, generally correspond to different indirect acts performed by the tags. All tagging type hashtags perform an assertive indirect act automatically, the content of which is something like “This tweet belongs on the page for \( X \),” where \( X \) stands in for a string of characters appended to the # symbol. Such a hashtag may also perform one or both of the other tagging type functions, outlined in table 1.4.

Table 1.4. Hashtag tagging sub-types.

<table>
<thead>
<tr>
<th>Tagging sub-type</th>
<th>Indirect act</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affiliating</td>
<td><em>I affiliate with the group associated with ( X ).</em></td>
</tr>
<tr>
<td>Participatory</td>
<td><em>This tweet is a contribution to ( X ).</em></td>
</tr>
</tbody>
</table>

As evident in table 1.4, the “exploitative” usage of tagging hashtags is left out. The fact that the user only means to exploit the practice of hashtagging for visibility is of course not intentionally conveyed, and therefore is not associated with a separate indirect act.

Let’s move on to non-tagging usages. Seeing as the same machinery is used to accomplish them (the # symbol), and the same changes are effected in a string of text by the
addition of the # symbol regardless of its meta-pragmatic function or likelihood as a search query, it must be the case that non-tagging usages also perform a direct effective act, with the same content, \( p \), as for tagging usages. These hashtags are the equivalent of mock performative statements, but ones that are secretly real performative statements, if anyone cared to follow up on them. Using a non-tagging type hashtag is like writing your friend a check for \( $0.01 \). Sure, it is a real check, that they could cash if they really wanted. But that was probably not the primary purpose behind the check.

Similar to how the tagging sub-types perform indirect acts in addition to their direct effective act, the non-tagging sub-types perform indirect acts as well. However, with the exception of the emphatic usage, I would posit that non-tagging usages are inherently more free-form than the tagging type indirect acts. Non-tagging usages are creative, and contextually-oriented. If they are meant to guide a reader’s interpretation of the adjacent text, then they must be flexible enough to convey a variety of indirect meanings. For this reason, I propose that the evaluative tags are not typified by performing an indirect assertive act, but rather a variety of potential indirect acts, of different forces.

I am making the decision to collapse the emotive and meta-commentative categories here because they often overlap anyway (e.g. #brag, as discussed in section 1.2.3.1.2), and are the only categories that do not perform an indirect assertive act automatically. Additionally, the emotive sub-type as I have defined it is confined to certain syntactic categories (items that could be construed as stage directions). Hence, if I wanted to look at only this category of hashtags in the post-annotation analysis, they would be easy to separate out.
Another argument for not only looking at the evaluative tag types together, but for allowing them to be annotated with a variety of indirect acts, is the fact that these types of tags tend to occur “after the main content of the tweet...rather than integrated into the main sentence,” (Scott, 2015, p.13). As I discussed in section 1.1.3, indirect speech acts are dependent on their associated direct acts. Therefore, it would make sense that the direct speech act of a tweet must be complete before one can begin to infer additional indirect acts from it.

The emphatic usage of hashtags, meanwhile, does not perform an indirect act, seeing as I would not consider any other bolded or italicized segment to perform an indirect act merely because it is bolded or italicized.

Turning to @-tags, we can apply the same pattern of logic. The @ symbol itself is an IFID marking exercitive force, which changes a string of text into a hyperlink when appended to it. The tagging type would have an assertive indirect act of something like “I am speaking to X user or X user’s followers.”\(^{23}\) The non-tagging type would not perform an indirect act (in line with the emphatic type hashtag usage).

Before concluding this section, I should point out that while an individual hashtag usage invokes or does not invoke the actual technological affordance of the hyperlink by definition (e.g. by definition a participatory usage invokes it, while an evaluative usage does not), an individual hashtag may be used in both ways at once. That is, just because a hashtag performs an evaluative function does not mean it does not also perform a tagging type usage. This is

\(^{23}\) This analysis is also fitting from the perspective of the literature on vocatives, which posits that they perform exercitive acts that grab or maintain a person’s attention. With @-tags, that exercitive act is encapsulated in the hyperlinking ability of the @-symbol itself. The indirect act I have proposed for tagging type @-tags simply indicates that this exercitive act is indeed the speaker’s intended outcome of using the @-tag (as opposed to non-tagging @-tags, where it is not). I will discuss this more in section 2.2.2.1.
particularly true for hashtag trends/games. For example, the hashtag #overlyhonestmethods is a long-standing hashtag trend on Twitter. The premise of this tag is to tweet a methodological choice made in a scientific study, explaining how that method was used or chosen for non-scientific reasons. Consider the example in figure 1.4, below:

"We used open source software, R, because it is free and we are poor" 😊
#overlyhonestmethods at #nca18

**Figure 1.4. Sample hashtag trend.**

The hashtag #overlyhonestmethods is indexing the tweet to a certain page (where I found it, by searching the hashtag), accomplishing participation in the game, and also making a meta-comment about the preceding text (something like *This is an overly honest explanation of our methodological choice*).

To summarize, as will be further discussed in chapter 2, all hashtags and @-tags in the corpus will be considered to perform exercitive direct acts automatically. Seeing as trying to identify emphatic usages of either type of tag would be highly speculative, I will not attempt to do so, and this distinction will be ignored by the annotation task. Thus, all @-tags will automatically be considered tagging type, performing the tagging type indirect act given in 1.5.

**Table 1.5. @-tag functionality.**

<table>
<thead>
<tr>
<th>@-tag function</th>
<th>Indirect act</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagging</td>
<td><em>I am speaking to X user or X user’s followers.</em></td>
</tr>
</tbody>
</table>
Hence, @-tags in the corpus will not require independent annotation to assess their usage.24

Hashtags, on the other hand, may perform any or all of the indirect acts given in table 1.6, corresponding with their possible functions. They will therefore receive independent annotation to assess their various usages. Note that, as defined in this section, a hashtag that performs an affiliating or participatory indirect act automatically performs a tagging indirect act. However, a hashtag that performs an evaluative indirect act does not necessarily not perform the tagging indirect act. How exactly the different tag usages will be presented to the annotators will be addressed in section 2.2.

Table 1.6. Hashtag functionality.

<table>
<thead>
<tr>
<th>Hashtag function</th>
<th>Indirect act</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tagging</td>
<td><em>This tweet belongs on the page for X.</em></td>
</tr>
<tr>
<td>Affiliating</td>
<td><em>I affiliate with the group associated with X.</em></td>
</tr>
<tr>
<td>Participatory</td>
<td><em>This tweet is a contribution to X.</em></td>
</tr>
<tr>
<td>Evaluative</td>
<td><em>[Various possibilities]</em></td>
</tr>
</tbody>
</table>

1.2.3 Quotations

1.2.3.1 Background

Quotations on Twitter pose some of the same problems as hashtags do. Namely, they blur -- or at least complicate -- the use/mention distinction. Consider the following tweet posted by the US Open Tennis official account:

\[\text{An @-tag may still incidentally receive independent annotation attention by virtue of the segmentation rules (e.g. if it comprises a direct address). This will be discussed in section 2.2.2.}\]
Figure 1.5. Tweet featuring quote with @-tag citation.

It could be that the quote in this tweet is merely stating words spoken by one person about another, and that the tweet is equivalent to saying, “Grigor Dimitrov said that there aren’t words or concepts to describe what Roger Federer is doing…” However, the quote might also be meaning to convey that the poster means to communicate the same thing that the quotee communicated. This is perhaps less likely when the tweet is coming from an official account like the US Open Tennis account (that would probably refrain from anything that could be construed as an opinion), but consider the following tweet from an apparently private citizen:

"The Presidency is not merely an administrative office. That's the least of it....It is pre-eminently a place of moral leadership."---Franklin D. Roosevelt, 1932

Figure 1.6. Tweet featuring quote with dash citation.

Again, the poster’s intention could just be to report that, “In 1932, Franklin D. Roosevelt said that the Presidency…” But the poster could also mean to convey that she is speaking these words too, and that she is asserting that “The Presidency is not merely an administrative office…”

Additionally, not all quotes appear with citations. For example, consider the following tweet consisting of only a quote with no citation or other text:
“How many seconds in eternity?”

Figure 1.7. Tweet featuring quote with no citation.

It is not even possible to restructure this tweet to the format of “[Person] asked…”, which might suggest that such an approach would be too reductive to begin with.

What to do about mixed quotation is also an important question. Consider the following tweet:

Excuse me, I do recall you copping off-white nike sneakers, talking about “I like them, I’m keeping them.”

Figure 1.8. Tweet featuring mixed quotation.

This is a mixed quotation because the quote is semantically incorporated into the rest of the tweet (beyond a mere citation of X said “…

Finally, there is the issue of recursive quotation. Consider the following tweet:

"My tombstone? I’m thinking something along the lines of, ‘Geez, he was just here a minute ago.’” George Carlin

Figure 1.9. Tweet featuring recursive quotation.

How the meaning of this tweet is structured and best analyzed is itself up for debate, let alone how we should place it into speech act theoretic terms.

With all of this in mind, the best approach to this problem as related to my study is to first explore the literature on this topic and land on one theory of the semantics of quotation. From there, we can better place these constructions into a speech act framework.
The body of literature surrounding quotation is quite large, but can generally be divided into a few distinct categories that solve the key problems that quotation poses in specific ways. I have already introduced one of those problems: the use/mention distinction and how it complicates our understanding of speaker meaning. There is another significant problem, however, that we might call the “opacity” problem.

The opacity problem is the fact that, in quotation, one cannot freely substitute co-referential terms while maintaining the same truth conditions. For example, while the noun phrases *Mark Twain* and *Samuel Clemens* have the same denotation (both phrases have the same real-world referent), the following (direct) quotations do not have the same truth values (unless John happens to have produced both utterances separately):

(1.19) John said “Twain wrote *Huck Finn*.“

(1.20) John said “Clemens wrote *Huck Finn*.”

As summarized by Botterell & Stainton (2005), the opacity problem pits two tenets of formal semantics against each other: 1) compositionality, and 2) innocence. Compositionality states that the semantic meaning of an utterance is a product of the semantic meaning of its component parts, and how those meanings should be combined (as dictated by syntax). Meanwhile, innocence “holds that the semantic value of an (unequivocal) expression E of L does not vary depending on what context E is embedded in,” (p.5). If we attempt to maintain both innocence and compositionality, we seem to have no way to explain the differing truth values of 1.19 and 1.20, seeing as the semantic meaning of the component referring noun phrases *Twain* and *Clemens* (i.e. the denotations) are one and the same.
The various theories of quotation account for opacity and the use/mention distinction in different ways. The two most popular types of approach (in terms of adherents) are known as the Demonstrative Theory and the Identity Theory (Saka, 2006, p.452). I will summarize the main points of both of these, in addition to the older (and generally rejected) Name Theory, before assessing how to incorporate quotation in Alston’s account of illocutionary acts.

1.2.3.1.1 The Name Theory

The Name Theory is the idea that quotations amount to “proper names” for quoted elements, and therefore do not have the same semantic values as those elements have without the quotation marks. Essentially, this theory proposes that all quotes are merely mentions. Davidson (2001) cites both Quine and Tarski as proposing some variation of this theory of quotation (pp.80-82).

Considering all quotes to be proper names does indeed solve the opacity problem. If “Twain” and “Clemens” are only names for expressions, with no relationship to the semantic values of their unquoted counterparts, then it would be expected that 1.19 and 1.20 would not have the same truth values.

But the Name Theory fails on the issue of the use/mention distinction. In short, many instances of quotation are clearly not only mentions. Cappelen & Lepore (1997) use mixed quotation to demonstrate this fact. Consider the following utterance:

(1.21) Alice said that life “is difficult to understand.” (p.438)

The quoted material here is being both mentioned and used. The speaker is conveying the fact that Alice said those exact words (i.e. used the proper name for that expression), but also that she made an assertion about the difficulty of understanding life. That is, the quoted material is
participating semantically in the rest of the utterance. The quoted material is a predication about “life,” a word outside the quotes in 1.21. If we change “life” to “death,” the utterance is now false (assuming that Alice’s original statement did not change), even though we haven’t changed any of the original quoted material. We have still used the same portion of her exact words, but changed the subject of them -- and that is enough to generate falsity.

(1.22) Alice said that death “is difficult to understand.”

Also, as Davidson points out, if the quoted words in 1.21 were not being used and did in fact consist of a proper name, the sentence should strike us as ungrammatical (p.81). Thus, the quoted material cannot only be mentioning.

1.2.3.1.2 The Demonstrative Theory

The Demonstrative Theory of quotation began with Davidson (1968, 2001), but has generated multiple off-shoots since then. Davidson’s account rests on the relationship between indirect quotation and pure quotation. He first proposes that the complementizer (“that”) appearing in indirect quotations is not actually a complementizer, but is instead a demonstrative pronoun (“that”). The pronoun is pointing at a token of an utterance (specifically, that which follows it), meaning that an indirect quotation is actually composed of two utterances. For example, under this theory the indirect counterpart to 1.19 is equivalent to 1.23:

(1.23) John said that. Twain wrote *Huck Finn.*

From there, Davidson proposes that quotation marks are not merely punctuation, but are actually referring expressions that function as demonstratives akin to the “that” in indirect

\[25\] Specifically, it would be equivalent to something like: *Alice said that life John.*
quotations (Botterell & Stainton, 2005, pp.5-6). Thus, pure quotations can be treated in the same way.

The appeal of the demonstrative approach is that it solves both the opacity problem and the use/mention distinction problem. It solves the opacity problem because the referent of the demonstrative changes when the exact wording of the quote changes. Thus, it is no longer problematic that two quotations can have different truth values even if they only differ in their usage of one or the other of two co-referenced terms.

As for the use/mention distinction, Davidson explains both the use and mention of quoted material by sinking the mentioning into the demonstrative (either the “that,” or the quotation marks). The demonstrative essentially copies the quoted material, so now there are two tokens of the quoted material, each linked to one of the functions (the demonstrative to the mentioning, the other to the using).

Despite the success of the Demonstrative Theory in solving some of the key issues of quotation, there are some strong challenges to it. First of all, there is significant evidence, both syntactic and semantic, against the idea that the “that” in indirect quotations is a demonstrative pronoun as opposed to a complementizer. On the syntactic side, the “that” of indirect quotations: 1) can easily be deleted, 2) cannot be replaced with other demonstrative pronouns, and 3) exhibits the phonological properties of complementizer “that” (ability to be reduced to “th’t”), as opposed to demonstrative “that” (Botterell & Stainton, 2005, pp.7-8).

26 Additionally, if we switch the order of the two elements that comprise an indirect quotation in Davidson’s view, the “that” can only be interpreted as a demonstrative pronoun and differs on all three of these points (e.g. Twain wrote *Huck Finn*. John said that.) (Botterell & Stainton, 2005, p.8).
On the semantic side, there is often a semantic dependence between the two parts of an indirect quotation that strongly suggests that the “that” is indeed a complementizer, and prevents the utterance from being split up in the way that Davidson proposes. Botterell & Stainton (2005) offer anaphora, negative polarity items, and VP ellipsis on this front (pp.8-10). For example, consider 1.24, where the negative polarity item is licensed by the matrix negation. Splitting 1.24 into two yields 1.25 and 1.26, a pair of sentences that do not together mean the same thing as 1.24 (and one of which may be ungrammatical).

(1.24) I didn’t say that there was any beer in the fridge.

(1.25) I didn’t say that.

(1.26) ??There was any beer in the fridge.

Saka (2006) also points out that mixed quotation and scare-quoting can both quote non-constituents. If the quoted material is not a constituent, then it cannot be replaced by something else (i.e. a demonstrative), and must be interpreted as integral to whatever syntactic unit it is a part of (pp.457-458).

1.2.3.1.3 The Identity Theory

The Identity Theory of quotation deals with both the opacity and use/mention problems differently from the Demonstrative Theory. Instead of diverting the mentioning function of a quote to an external element (the demonstrative), the Identity Theory proposes that a quoted element mentions itself (and therefore potentially performs both the mentioning and using functions at the same time). As to why non-quoted elements do not mention themselves, this is explained by proposing that a quoted element does not have the same semantic meaning as its unquoted counterpart. Washington (1992) cites both Frege and Searle on this point, quoting the
latter as writing “that in quotation ‘a word is uttered...but not in its normal use. The word itself is presented and then talked about, and that it is to be taken as presented rather than used conventionally to refer is indicated by the quotes’” (p.587). In other words, under this theory, all words are inherently polysemous, as they all inherently have at least two meanings (a standard meaning and a quotational meanings).

Returning to the opacity problem, and the *Twain* and *Clemens* example, the Identity Theory would explain the potentially differing truth values of 1.19 and 1.20 with the fact that, in this theory, two terms with co-referential standard (un-quoted) meanings do not necessarily have co-referential quotational meanings.

But the proposal that there is both a “standard” meaning for an expression as well as a “quotational” meaning for an expression is costly to a unified theory of semantics. As Botterell & Stainton (2005) put it, such a move “sacrifices” the tenet of innocence (p.21).

Potts’s (2007b) proposal is similar to the Identity Theory in some ways, as Potts also discusses a quotation as having two meanings, the “regular meaning” and the “speech report meaning” (i.e. the use meaning, and the mention meaning). He posits a grammar wherein a semantic quotation function takes any utterance and turns it into a linguistic object. This linguistic object can then participate in semantic activities in the same way that entities and propositions can (p.6). A direct quotation such as that in 1.19 yields both meanings because it yields two different propositions, represented as one object of a semantic “product type.” Basically, the quotation-taking meaning of “say” simultaneously outputs both the regular meaning as well as the speech report meaning, a result of its y-domain being a “product type” domain (p.11-12).
1.2.3.2 Speech Act Status

While it is important to consider the various theories of quotation before deciding how exactly it fits into a speech act framework, the most important takeaway from section 1.2.3.1 is actually a point of agreement between the Demonstrative and Identity theories. In short, they agree that quotations both mention and use the quoted material. That is, they agree that quotations themselves are neither just names, nor semantically empty.

Given this fact, it is initially tempting to propose that quotations should be considered to perform whatever illocutionary speech act(s) their non-quoted counterparts would perform. However, if we return to Alston’s definition of a direct illocutionary act, we can see an immediate problem with such an analysis. Alston states that the performance of a direct illocutionary act requires that the speaker take responsibility for certain conditions holding true. But, especially in the case of pure quotation, the speaker of a quote is often not taking responsibility for the same conditions that the quotee is. Consider the following exchange:

(1.27) **Dan**: I am happy.
**Steve, to John**: What did Dan say?
**John**: “I am happy.”

In 1.27, Dan is taking responsibility for the condition that he is, at the time of speaking, in possession of a certain psychological state. He is therefore performing an expressive illocutionary act. Meanwhile, John, in quoting Dan’s utterance exactly, is not performing an expressive illocutionary act. He is not taking responsibility for the condition that he (John) is in possession of a certain psychological state. He is not even taking responsibility for the fact that Dan is in possession of a certain psychological state. He is merely stating what it is that Dan said, thereby performing an assertive illocutionary act. If John’s utterance was instead...
performed with a rising intonation, and he was instead taking responsibility for laying a responsibility on the hearer to confirm or deny Dan’s exact words, he would be performing a directive illocutionary act (according to Alston). In both cases, this act of John’s is what Potts (2007b) calls the “speech report meaning” of a quotation, the meaning that encompasses the “mentioning” half of the use/mention distinction.

Meanwhile, in asserting or questioning that Dan said X, John could intend for any of a variety of additional acts to be inferred from his utterance. He could mean to convey that he shares the same psychological state as Dan, that he can’t believe Dan said that, etc. Such additional acts, if present, would of course be indirect, given that they are inferred. We might think of such acts as encompassing the “using” half of the use/mention distinction.

Thus, I propose that pure quotations perform a direct speech act automatically, an act that accomplishes the “mentioning” function, but may also perform other speech acts indirectly, acts that accomplish the “using” function. The equation of “mentioning” with direct acts and “using” with indirect acts is likely a bit reductive, but it is sufficient for my purposes and places pure quotation neatly into the annotation framework.

As for the other types of quotations laid out in the introduction to section 1.2.3.1, pure quotations without clear citations should be treated the same as pure quotations with citations. The speaker is still asserting or questioning that someone said X.

Instances of mixed quotation and indirect quotation are both more complicated, and draw attention to an issue that will be dealt with in section 2.2: how to segment utterances into speech act units. Hence, how exactly these types of quotation will be addressed will be revisited in section 2.2.
1.2.4 Likes and Retweets

1.2.4.1 Background

The practices of liking and retweeting are ultimately quite similar. As will be addressed in the following sections, the “like” button and the “retweet” button share similar non-linguistic functions. These functions will be summarized in section 1.2.4.1.1. Whether or not “like” and “retweet” perform linguistic functions, and therefore illocutionary acts, will be addressed in sections 1.2.4.1.2 and 1.2.4.1.3.

The Twitter “like” button used to be called the “favorite” button. Some of the sources regarding this button were written when it had its former name. In cases where I make reference to such a source, I use “like” instead of “favorite” to discuss their findings, for the sake of consistency and clarity. Likewise, “retweeting” on Twitter was an emergent practice, for which a dedicated button was added later on. Some of the sources referenced for this practice were written before the dedicated button existed, but this required no substitution of terms.

1.2.4.1.1 Posting Functions of Likes and Retweets

For the purposes of this section, “posting function” refers to the action that produces consistent, demonstrable, non-linguistic results when the “like” and “retweet” buttons are clicked, due to the technological affordances of the medium.

When the “like” button is clicked, several events occur: the “like” count of the tweet goes up, the button itself turns red in the liker’s user-interface, and the “like” count on the liker’s Twitter profile page goes up. Several additional events may occur, depending on the liker’s settings: the liker’s Twitter profile picture may appear beside the “like”/”retweet” counts of the tweet, other users may get notifications on their timeline regarding the liker’s action (e.g. John
\(\text{liked [tweet]}\), and the tweet that was liked may appear under the “like” count/tab on the liker’s profile page.

When the “retweet” button is clicked, the first thing that happens is that the retweeter has the option to retweet directly, or retweet with comment. If they choose to retweet directly, several events occur: the “retweet” count on the tweet goes up, the button itself turns green in the retweeter’s user-interface, the “tweet” count on the retweeter’s profile page goes up, the tweet is posted to the retweeter’s profile (with the retweeter’s name above it, e.g. \textit{John retweeted [tweet]}), and the tweet is posted to the timelines of the retweeter’s followers (with the retweeter’s name above it). As with the “like” button, depending on the retweeter’s settings, the retweeter’s profile picture may appear beside the retweet/like counts of the tweet.

If the retweeter chooses to retweet with comment, the tweet is inset and the retweeter types their comment above it before clicking “retweet” again. At that point, all of the events that occur with a direct retweet occur, except that instead of the original tweet appearing with \textit{[retweeter’s name]} retweeted above it, the original tweet appears inset with the retweeter’s comment above it.

Though the posting functions of these two buttons are not exactly the same (i.e. different sets of tangible results are produced when they are clicked), the role of the posting function in the performance of any potential speech acts is the same: it is merely the mode. “Posting” something is analogous to “speaking,” “signing,” or “writing” it. In Austin’s original terms, we might compare posting with the locutionary act by way of which any illocutionary acts are performed. Therefore it may be more accurate to refer to the posting function not as the
“non-linguistic” function of the “like” and ”retweet” buttons, but rather as their “non-illocutionary” function.

1.2.4.1.2 Linguistic Function of Likes

While the non-linguistic function of the “like” button is the posting function, any potential linguistic functionality of the button is more nebulous. Users report a wide variety of reasons for clicking the button.

Meier et al. (2014) performed a survey of 606 Twitter users, asking open-ended questions about the users’ motivations for liking a tweet. From the responses to those questions, the authors distilled a hierarchical taxonomy of 25 motivations for “liking” a tweet. Re-organized to be relevant to the present analysis, all of these motivations can be split into two groups: those that depend on the posting functionality (e.g. bookmarking for later, participation in a like-count competition), and those that express some sentiment or opinion by way of the like (e.g. agreement with or interest in the tweet content, personal relationship with or approval of the author, personal connection/relevance to the tweet content, etc.) (pp.5-7).

Gorrell & Bontcheva (2016) performed a similar study, and ended up with a taxonomy of user approaches to the Twitter “like” functionality that is similar to that of Meier et al. Again, the classifications in that taxonomy can be organized into those that depend on the posting functionality (bookmark and self-promotion), and those that depend on the linguistic functionality (thanks, conversational, and literal like) (p.19).

However, these self-reported motivations for clicking the “like” button cannot be considered direct illocutionary acts. If we return to Alston’s definition, we must again ask the question of what the liker can be understood to take responsibility for. Someone who has clicked
the “like” button is not taking responsibility for anything other than that click. They may have indeed intended to perform an illocutionary act by clicking the button (e.g. *I approve of this tweet*), but any such act would be perfectly deniable, as they could always appeal to one of the inherent results of the posting function of the “like” button.\(^{27}\) Therefore, any potential linguistic messages conveyed by the clicking of the “like” button, many of which are suggested by Meier et al.’s and Gorrell & Bontcheva’s taxonomies of motivations, must be indirectly performed.

1.2.4.1.3 Linguistic Function of Retweets

The linguistic functionality of the “retweet” button initially appears to be more clear than that of “like.” Retweets are quite naturally, and consistently, described as either direct or indirect quotations (Gruber, 2017, pp.1-3). As discussed in section 1.2.3, a direct (pure) quotation was determined to perform an assertive direct act automatically, as the only thing the speaker could be held responsible for was that the quotee did indeed produce that utterance. A retweet, similarly, might seem to hold the speaker responsible for the fact that the quotee (retweetee) did indeed tweet a given utterance.

Before the days of the dedicated “retweet” button, that might have been an appropriate analysis (seeing as the practice was generally to just copy and paste the retweetee’s tweet content and add some sort of citation). However, with the button, it is not exactly the same. A click of the “retweet” button actually reproduces an image of the initial tweet (on the retweeter’s profile and timeline) which is hyperlinked to the actual tweet. In spoken conversation, this would be equivalent to something like the following exchange:

---

\(^{27}\) For example: **Dan:** I saw that you “liked” John’s tweet. Does that mean you approve of it?  
**Steve:** No, I just wanted to save it under my “like” tab for later.
**Dan**: I am happy.

**Steve, to John**: What did Dan say?

**John pulls out a recording device and replays Dan’s exact utterance of “I am happy.”**

In other words, a click of the “retweet” button does not simply hold the retweeter responsible for the fact that the retweetee did indeed tweet a given utterance. The fact that the retweetee posted the initial tweet is a non-debatable fact, a condition that cannot fail. An infelicitous quotation can exist, which might receive a response of *X didn’t say that*. But with the dedicated “retweet” button, an infelicitous retweet resulting in a response of *X didn’t tweet that* does not exist.

As with the “like” button, however, a retweet could be used to perform any number of indirect acts (conveying agreement with the tweet, spreading news about a favorite person, increasing visibility for the retweetee, etc.). Such indirect acts, like those seen for the “like” button, would again be deniable, as the retweeter could always appeal to one of the inherent results of the posting function of the “retweet” button.

As for the “retweet with comment” option on Twitter, the comment of course may perform any number of direct or indirect acts (as the comment itself is essentially an additional tweet).

---

28 Technically, the most accurate analogy would be if John pulled out a time machine and took Steve back in time to the moment of Dan’s original utterance. A retweet does not make a copy of a tweet (as a recording device makes a copy of an utterance), but actually links to the original production (the original tweet).

29 This does not include situations wherein the person behind a given Twitter profile is not who the profile claims they are. In such a case, a response of *X didn’t tweet that* would be felicitous, if the speaker meant that the tweeter was misrepresenting their identity.

30 For example: **Dan**: I saw that you “retweeted” John’s tweet. Does that mean you approve of it? **Steve**: No, I just want to increase my overall tweet count.
1.2.4.2 Speech Act Status

In sum, the illocutionary possibilities of both the “like” and “retweet” buttons are limited to the indirect level. At this point, it is worth noting that I will not be asking annotators to annotate individual “likes” or “retweets” for indirect acts. Such would be a highly speculative process, even if I asked the annotators to annotate all likes and/or all retweets for a given tweet as a group. This discussion of the speech act status of these elements was done with the goal of thoroughness, to look at all of the basic elements of a tweet for potential speech act performance. Like and retweet counts will be recorded, however, for use in the post-annotation analysis of hypothesis 5, as will be discussed in section 2.4.

1.3 Summary: The Annotation Task So Far

In this section, I will briefly review what has been determined about the corpus and the annotation task so far regarding what information needs to be collected, and what choices the annotators should be given. This section will set up for chapter 2, wherein the form of the annotation task will be finalized, and guidelines will be laid out for both the annotation and segmentation of the data.

This project was begun with the intention to annotate a body of Twitter data for illocutionary acts. To that end, section 1.1 laid out the version of speech act theory which will underlie this project. In section 1.1.4.1, this version was used to define key terms such as “illocutionary act,” “direct,” and “indirect.” At this point, it is clear that the two primary pieces of information that the annotators should be asked to annotate are the direct act of a given segment and the (or a) indirect act of a given segment (if present).

31 For example: “This tweet received 5 likes. Do you think someone who liked this tweet meant to perform a speech act in doing so? If so, what type of act?”
Additionally, seeing as it would be interesting to see whether Clark’s (1979) findings are reproduced by this corpus, it is worth asking annotators about the “seriousness” of the direct acts they annotate, as well. Based on Clark’s findings, we would expect to find that a non-serious direct act was more likely to co-occur with an indirect act than not. In the event that an annotator classifies the direct act of a given segment as serious, but also classifies that segment as performing an indirect act, the annotator will be asked which of the acts is more important to the speaker’s overall message. Clark’s findings would expect that a serious direct act and an indirect act performed by the same segment are of equal importance to the speaker’s message. Whether or not Clark’s (1979) findings are reproduced by the results of the annotation will be discussed in the analysis of hypothesis 2, in section 4.2.2.

The questions to be asked of the annotators thus far are given in 1.29 through 133.

(1.29) What type of speech act is this segment performing when taken directly (whether literal or non-literal)?

(1.30) Do you think that the direct act is meant to be taken seriously?

(1.31) Is this segment performing any other speech act, indirectly?

(1.32) *If answer to 1.31 is yes* What type is it?

(1.33) *If answers to 1.30 and 1.31 are both yes* Which act do you think is more important to the speaker’s overall message? That is, which one seems to be the main purpose behind tweeting the segment?

How exactly the key terms of “directly/indirectly,” “seriously/non-seriously,” and “literal/non-literal” should be defined for non-linguist annotators will be discussed in section 2.1.3.

The taxonomy of act categories that annotators will be able to choose from when classifying the direct and indirect acts of a given segment will come mainly from Alston (2000),
but there are also issues of comparison and non-linguist accessibility to consider. These will be addressed in section 2.1, where the final taxonomy of acts and how they will be defined for the annotators will be decided.

Meanwhile, section 1.2 led to additional considerations, based on the speech act status of the technological affordances of the medium. In some cases, the speech act status of a given element suggests that it should in fact receive independent annotation. Both quotations and hashtags were determined to be capable of performing indirect acts. This being the case, in the case of pure quotation, the annotators will be asked the questions in 1.34 and 1.35. In the case of hashtags, the annotators will be asked the questions in 1.36 through 1.39, which should together cover the potential multifunctionality of hashtags.

(1.34) This segment is a direct quotation, and therefore performs a direct act automatically (an act that mentions the linguistic object of the quotation). Does it also perform a speech act indirectly?

(1.35) *If answer to 1.34 is yes* What type is it?

(1.36) Is this a hashtag that would actually be used? Does the speaker expect that the a Twitter user would ever actually click on or search for this tag?

(1.37) Do either of the following statement apply to this hashtag? You may select one, both, or neither. [Note: The X stands for the string of characters following the hash symbol.]
(a) The speaker affiliates with or supports the group associated with X.
(b) This tweet is a contribution to a meme or trend X.

32 This point constitutes a caveat of the present study. My original analysis of pure quotations was that they do not perform any direct act, not that they automatically perform a direct act of mentioning. That analysis determined the form of the annotation task for quotations, and annotators were told that quotations did not perform a direct act but could perform an indirect one. While in both cases annotators only assessed quotations for indirect acts, it may have affected their annotation to have been told that the quotation performed no act directly. Quotations were not a major focus of this study, and were not addressed by any of the 5 hypotheses, so the effect of this caveat on the analyses in chapter 4 is small to non-existent.
Is this hashtag meant to convey an emotive and/or meta-commentative meaning?

*If answer to 1.38 is yes* How would you best classify the speech act of the meaning?

In other cases, the speech act status of an element suggests that additional information should be noted on the part of the researcher. For emoticons/emoji, judgments of which of these elements are incorporated into adjacent text, and which unincorporated emoticons/emoji are lexical vs. tone-type, will be made by me during the segmentation. How these judgments will be made will be discussed further in section 2.2.

Emoticons/emoji are not the only elements for which incorporation into adjacent text is an issue. Quotations may be incorporated (so called “mixed quotation”), as can hashtags and @-tags. How exactly this issue will be addressed will be discussed in section 2.2, when guidelines for the segmentation of the corpus into speech act units will be laid out. Section 2.2 will conclude with the final version of the annotation task as it will be presented to the annotators.
CHAPTER 2: ANNOTATION INFORMATION AND GUIDELINES

2.0 Introduction

In this chapter, I will build on the annotation plan as developed in chapter 1 and summarized in section 1.3. In section 2.1, I will determine the exact taxonomy of speech act categories that the annotators will be able to choose from, how those categories will be defined for the annotators, and how certain key terms will be defined for non-linguist annotators.

Section 2.2 will address the issue of data segmentation, and what should be considered a speech act unit. In addition to establishing guidelines for segmenting the data, this section will make clear the need for a few more segment classifications. That is, we will find additional segment categories similar to pure quotation and hashtags, ones which will need to be independently annotated using a set of questions that is slightly different from the base set. This section will conclude with the final version of the annotation task.

Sections 2.3 and 2.4 will motivate the last two hypotheses, 4 and 5. Hypotheses 4 and 5 do not affect the form of the annotation task, but do require certain additional information be gathered.

2.1 Annotation Guidelines

In the course of finalizing the annotation task, including the exact format of the taxonomy and accessible definitions of key terms, there are a few basic requirements that it should strive to meet:

1) It should be theoretically motivated. That is, the information that the annotators are annotating, the options available for each variable, and the definitions of key terms
should be based on a coherent version of speech act theory. Identifying such a theory was the primary goal of section 1.1.

2) At the same time, hypothesis 3 seeks to compare the results of the present study with studies of speech acts in other media, possibly even in spoken conversation. For the results to be comparable, the information annotated for, and the options available for each variable, should be partially determined by the other studies with which I will be comparing results.

3) The annotation task itself should be intuitive and approachable for the average native speaker. A more intuitive and approachable task will generate more consistent results, while making it accessible to the average native speaker broadens the pool of potential annotators.

It is a bit of a balancing act to compose an annotation task that maximally respects all three of these requirements. The push-and-pull between one and two (theory vs. comparability) is especially difficult to navigate. However, I propose a preliminary concession to comparability over theory on this point. If I determine that category X is not sufficiently theoretically motivated, but every other study on the topic includes a category X, including category X in my annotation is better than not. Having comparable information but adding a disclaimer concerning its theoretical position is better than not having the information and making conjectures as to comparability.

In section 2.1.1, I will assess which illocutionary act categories should be included in the taxonomy provided to annotators, based on the theoretical background in chapter 1. In 2.1.2, I will move on to the issue of comparability, looking at what annotation/coding scheme studies of
speech acts in other media (including spoken conversation) have used. The conclusions of the theoretical motivation will then be re-assessed in the light of comparability. Finally, in section 2.1.3, the whole task will be assessed with respect to intuitiveness, approachability, and practicality, with the goal of making it accessible to the average native speaker.

2.1.1 Theoretical Motivation

Section 1.1 determined that the version of speech act theory which should underlie this project is Alston (2000). As such, Alston’s account has thus far been used to provide definitions of key terms in section 1.1.4.1, as well as to determine that certain technological affordances should receive independent annotation. Hence, I will now revisit Alston’s illocutionary act taxonomy as the starting point to determining the exact taxonomy of categories that should be provided to the annotators, and how those categories should be defined for them.

Alston defines illocutionary force by the socially entrenched rules that a speaker submits their utterance to -- rules which stipulate the necessary conditions “for permissible utterance only within a certain sphere of activity, only when [the speaker] is playing the appropriate ‘language-game’ [as determined by the speaker],” (p.62). Table 2.1 reproduces table 1.2, which defines the speech act categories in Alston’s taxonomy according to his proposed illocutionary rules.

[Note: U = the speaker; H = the hearer; S = the utterance or utterance surrogate; R = taking responsibility for X condition being satisfied. All rules in this table are reproduced verbatim.]
<table>
<thead>
<tr>
<th>Category</th>
<th>Associated I-Rules</th>
</tr>
</thead>
</table>
| **Assertive** | U asserted that \( p \) in uttering \( S \) iff:  
1. U R’d that \( p \).  
2. \( S \) explicitly presents the proposition that \( p \), or \( S \) is uttered as elliptical for a sentence that explicitly presents the proposition that \( p \).  
(p.120)                                                                                                                                                                                                                                                                                                                                                      |
| **Directive** | U D’d in uttering \( S \) (where ‘D’ is a term for some directive illocutionary act type, a purporting to be producing a certain kind of obligation on \( H \) to do D) iff in uttering \( S \), U R’d that:  
1. Conceptually necessary conditions for the existence of the obligation are satisfied. (These include such things as that it is possible for \( H \) to do D, that D has not already been done, etc.)  
2. Circumstances of the utterance of \( S \) are appropriate for the production of the obligation in question. (This includes the appropriate authority for orders, the right kind of interpersonal relationship for requests, etc.)  
3. By uttering \( S \), U lays on \( H \) a (stronger or weaker) obligation to do D.  
4. U utters \( S \) in order to get \( H \) to do D.  
(pp.102-103)                                                                                                                                                                                                                                                                                                                                                   |
| **Commissive** | U C’d in uttering \( S \) (where ‘C’ is a term for a commissive illocutionary act type, a purporting to produce an obligation on \( U \) to do D) iff in uttering \( S \), U R’d that:  
1. Conceptually necessary conditions for \( U \)’s being obliged to do D are satisfied.  
2. \( H \) has some interest in \( U \)’s doing D.  
3. U intends to do D.  
4. By uttering \( S \), U places herself under an obligation to do D.  
(p.97)                                                                                                                                                                                                                                                                                                                                                   |
| **Exercitive** | U O’d in uttering \( S \) (where ‘O’ is a term for purporting to be producing a particular conventional effect, E) = df. In uttering \( S \), U R’d that:  
1. Conceptually necessary conditions for \( E \).  
2. U has the authority to produce \( E \).  
3. Conditions are appropriate for the exercise of that authority.  
4. By uttering \( S \), U is bringing about \( E \).  
(p.93)                                                                                                                                                                                                                                                                                                                                                   |
| **Expressive** | U expressed a \( P \) (some psychological state) in uttering \( S \) iff in uttering \( S \), U R’d that U has a \( P \).  
(p.109)                                                                                                                                                                                                                                                                                                                                                   |
This taxonomy, in terms of the categories included, is quite similar to that proposed by many other accounts, across the various branches of the speech act family tree. There is nothing particularly unusual or special about his choice of categories. However, while Alston of course carefully motivates the definitions of these categories (based on his motivated definition of illocutionary act), he does not concisely motivate the categories themselves, seeming to include the categories that he does simply based on Austin’s original taxonomy.

As I stated in section 1.1.2.3, from a parsimony perspective, Alston’s whole proposal would be stronger with a unified dimension (or intersection of dimensions) wherein these five categories are aligned with the only five outputs that the dimension or intersection yields. For example, Searle & Vanderveken (1985) accomplish such a task with their concept of “direction-of-fit.” But Alston provides no such dimension.

This lack of unified motivating dimension is a caveat to the theoretical motivation of the annotation taxonomy. However, as emphasized in Chapter 1, the goal of this project is not to determine the most theoretically sound account of illocution. Seeking a unified motivating dimension would be a project unto itself. With that caveat noted, the five category taxonomy provided by Alston (2000) remains the theoretically motivated version of the taxonomy.

2.1.2 Comparability Motivation

With a theoretically motivated version of the task in hand, we may now move on to the issue of comparability. What sorts of taxonomies and coding schemes have other studies of speech acts (both in CMC and in spoken conversation) used? There are two primary motivations for investigating this question. The first is that it is worth asking why, theoretically or practically, studies similar to this one chose to follow a certain taxonomy. Their reasoning may
influence my own, and hence the formulation of the task. The second motivation is that hypothesis 3 seeks to compare the results of this study with studies of other media.

Based on the overview of these studies given in sections 2.1.2.1 and 2.1.2.2, I will determine whether any reasonable changes are called for with respect to the theoretically motivated version of the taxonomy.

2.1.2.1 Speech Act Studies in Computer-Mediated Communication

The various media in this category include: social websites like Facebook, Twitter, and online discussion forums; instant messaging applications; and email. I will begin with the social websites, seeing as they include my medium of study and are therefore inherently the most relevant of the group.

2.1.2.1.1 Twitter

While several studies have investigated one specific type of speech act on Twitter or discussed the pragmatics of the medium more broadly, only a few have attempted to classify general Twitter data into a taxonomy of speech acts. Zhang et al. (2011, 2012) and Vosoughi & Roy (2016) are the only current studies of this nature, all of which share a primary goal distinct from my own: constructing an automatic classifier for Twitter data. Zhang et al. (2011) and Vosoughi & Roy (2016) focus on building a program to automatically classify the data into speech act categories, while Zhang et al. (2012) build on their (2011) work to construct an

---

automatic topic summarizer for Twitter data. To this end, all three studies only categorized whole tweets, performing no segmentation of the tweet text, syntactically-based or otherwise.

Vosoughi & Roy (2016) construct their classifier using a semi-supervised method, starting with a set of manually annotated tweets that formed the basis for training the classifier. In their own words, they used Searle’s (1976) speech act taxonomy to “[establish] a list of six speech act categories that are commonly seen on Twitter: Assertion, Recommendation, Expression, Question, Request, and Miscellaneous,” (p.711). No other motivation is given for the selection of these particular categories. But we can see from the rest of their methodology that, in addition to using Searle’s taxonomy, Vosoughi & Roy subscribe to Searle’s theory of speech acts (that certain overt features can be interpreted as indicating the presence of certain illocutionary forces).

After the manual annotation using these (mutually-exclusive) categories, the authors analyzed the results, and referred to some outside sources, to determine (binary) semantic and syntactic predictors of particular acts. These cues were used to help train the classifier. The semantic features were of five main types: 1) opinion words (which “tend to signal certain speech acts such as expressions and recommendations”, p.712); 2) vulgar words (which tend to indicate an expressive speech act); 3) emoticons (which tend to signal expressiveness); 4) speech act verbs (which can indicate a speech act in any one of the categories); and 5) N-grams, because, aside from verbs, “there are certain phrases and non-verb words that can signal certain

---

34 An additional benefit to comparing results with studies aimed at building automatic classifiers for speech act information is that the various features that they use to train their classifiers may turn out, in the post-annotation analysis, to be good speech act category predictors in my own data. That is, the features which were found to be good predictors of certain speech acts in the comparison studies are potential candidates for overt illocutionary force indicators in my study. We will return to this possibility in sections 4.2.1 and 4.2.2.
speech acts,” (p.712). The features in the first four of those categories were identified with the help of outside resources (e.g. Wierzbicka (1987) for a list of “speech act verbs” in English, an online database of profanity, etc.). The N-grams were identified using only the dataset itself.

The syntactic features were also of five main types: 1) punctuation (? and ! only); 2) Twitter-specific characters (i.e. #, @, and RT); 3) abbreviations (“can signal informal speech which in turn can signal certain speech acts such as expression”, p.713); 4) dependency sub-trees (for reasons similar to N-grams); and 5) part-of-speech (interjections usually signal expressions, while adjectives “can signal expressions or recommendations”).

Zhang et al. (2011) used a semi-supervised method as well. They first annotated a body of tweets by hand, using Searle’s (1975) taxonomy, with some modifications because “commissives are rare [on Twitter] and declaratives are even rarer according to our observation of the experimental data” (p.86). Commissives and declarative were therefore lumped into a “miscellaneous” category. Thus, Zhang et al. ended up with a taxonomy identical to that of Vosoughi & Roy, but with “directive” broken up into only “question” and “suggestion,” as opposed to the three-way split of Vosoughi & Roy (i.e. “question,” “recommendation,” and “request”).

After the manual annotation, as with Vosoughi & Roy’s study, Zhang et al. analyzed the results to uncover overt (mostly binary) features that seemed to be reliable predictors of certain speech acts. The data had been pulled from specific trending topics on Twitter, and the authors pulled features from each topic separately. They then trained separate automatic classifiers using those features. After training, they tested the classifiers on the different datasets, to see: 1) how well this feature-based system worked in general, 2) the effect of “data granularity,” (i.e. whether
As for the actual cues, Zhang et al. include all of the feature categories that Vosoughi & Roy include, with the exception of dependency sub-trees and part-of-speech information. The largest difference here is that, although Zhang et al. did use outside references to identify opinion words, abbreviations, profanity, and emoticons, they did not use an outside list of “speech act verbs,” instead relying only on analysis of the manual annotation results for specific cue words and phrases that aren’t necessarily opinion words, profanity, emoticons, or abbreviations.

In their (2012) follow-up, Zhang et al. use the methodology described above to attempt topic summarization of Twitter data. In the course of this study, they assess which categories of features seemed to be the best predictors, finding that “cue words and phrases” (that is, N-grams pulled from the data itself) did better than the other categories (including opinion words, profanity, punctuation, etc.) (p.652).

2.1.2.1.2 Facebook

As of 2011, a user’s Facebook homepage looked quite similar to his Twitter homepage, with an ever-updating newsfeed of the most recent posts by the other users that he follows. Even with the huge difference in character limit (60,000+ vs. 140), Facebook is one of the social media sites most similar to Twitter. This makes Carr et al.’s (2012) study of speech acts in Facebook status updates a useful point of comparison for my own study. Additionally, Carr et al.’s goals are more like those of the present study than are the goals of the Twitter studies just covered. Instead of gearing their study towards the construction of an automatic classifier for

35 Actually, at the time of writing, Twitter has updated the character limit to 280. But all of the initial data for this study was collected during the reign of the 140 limit.
Facebook data, Carr et al. are more interested in analyzing the pragmatics of the medium, both quantitatively and qualitatively, using speech acts.

As with the Twitter studies, Carr et al. follow a Searlian taxonomy with a few modifications. The authors take Searle’s five-point taxonomy and make two modifications: 1) split “declarative” into “effective” and “verdictive,” following Clark (1996), and 2) add “quotation” as an entire category, following Nastri et al. (2006), though they clarify that those were not considered speech acts (p.184). They also note that they did not actually find any effective or verdictive acts in their data. Additionally, the authors coded the status messages for the presence of humor (i.e. was a given status intended to be funny).

The authors collected a total of 204 status messages from 46 different users over a 2-week period. The data was then segmented into “punctuation or propositional units,” after which the units were coded according to the (mutually exclusive) seven-point taxonomy of speech acts, as well as coded for the presence of humor. Two coders independently coded all the units, and their intercoder reliability for speech act classification was “moderate,” while intercoder reliability for the presence of humor was “substantial” (p.184).

2.1.2.1.3 Online Discussion Forums

Online forums exist for essentially every topic imaginable. Many such forums require one to be a member of a certain community before they may post, but this is often just a matter of setting up a free profile using one’s email address first. It is no surprise, given the abundance of this type of data, that studies of speech acts in this particular medium exist.

Online forums are similar to Twitter in one sense, seeing as it is usually the case that a user may respond to another’s post, creating a “thread” of conversation or discussion, as well as
“like” or re-post (i.e. re-tweet) the content. However, there is a significant difference between the two media in that discussion forums are typically geared towards a specific topic. Hence, there is some inherent restriction (either formally or informally) on what a person may post, which likely affects the speech acts used in posts.

The three studies I have found that analyze forum data in terms of speech acts share a goal similar to that of the Twitter studies: to construct an automatic classifier for the data, for the purpose of either speech act classification or topic identification (using speech acts). Jeong et al. (2009), Feng et al. (2006), and Qadir & Riloff (2011) all use a semi-supervised approach, with some slight variations in the exact methodologies and taxonomies.

Jeong et al. (2009) started with labeled data from the Switchboard-DAMSL and Meeting Recording Dialog Act databases of spoken dialogue, extracting phrase and dependency tree features from that data and then using those features to train a classifier. The classifier was then tested on new email and forum data, and those results were compared to the results of a manual annotation of that same data. This manual annotation was performed by two independent annotators, and inter-annotator agreement for the forum data was $\kappa = 0.73$.

As for the formulation of the annotation ask, the unit of annotation was the sentence, and the taxonomy that the authors used consisted of 12 categories, taken from the guidelines used in the Meeting Recording Dialog Act database (Dhillon et al., 2004). However, their taxonomy is actually a mixture of speech act categories (e.g. yes-no question, statement) and dialogue act categories (e.g. polite mechanism, accept response) (p.1252), making it overall less relevant to my study. We may be able to re-classify those categories into a more traditional speech act.
taxonomy, but I will discuss this option further when covering the MRDA database in the spoken conversation section.

Qadir & Riloff (2011) use a similar approach, training their speech act classifier using a variety of “lexical, syntactic, and semantic” features before testing it on the forum data from an online veterinary medicine message board, and comparing those results with the results of a manual annotation of the same data. However, the authors specifically note the differences between their methodology and that of Jeong et al. Instead of Jeong et al.’s dialogue act taxonomy, Qadir & Riloff use Searle’s original five-point taxonomy (with some name changes), though they drop “declarations” because “we virtually never saw declarative speech acts in our data set,” (p.750). Additionally, they use a much wider variety of features than did Jeong et al. (or any of the Twitter studies), with some outside resources providing some information in that area (as with the Twitter studies).

The authors divide their features into three main groups: *Lexical and Syntactic (LexSyn) features, Speech Act Clue Features*, and *Semantic Features*. The first group consists of 13 different subcategories, such as modal auxiliaries, infinitive clauses, “plan phrases” (e.g. *I’m going to*...etc.), sentence position in the message, and how many verbs are in the sentence. The second group has only two subcategories, both consisting of speech act word lists from external sources (Searle, 1976; and Wierzbicka, 1987). The last group is more domain-centered than the first two. That is, the Semantic Features group consisted of two subcategories whose contents were pulled from the data itself using bootstrapping algorithms. The first of these was a

---

36 Qadir & Riloff state that the first two groups, LexSyn and Speech Act Clues, are “domain-independent and should be useful for identifying speech act sentences across many domains,” (p.752).
semantic lexicon of the most common nouns in the corpus that were related specifically to the topic of veterinary medicine, and the second subcategory was a set of semantic class tags.

One unique taxonomic choice made by Qadir & Riloff was to differentiate sentences that perform speech acts from those that are merely “expository.” That is, on their definition, a sentence that is providing “factual” information (an expository one) does not perform a speech act, and is different from a sentence explicitly stating one’s belief in that information (an assertive act, or “representative” in their terminology). This is not simply a passing and insignificant designation, but integral to one of their main goals -- for the classifier to be able to separate the expository sentences from the ones that perform speech acts. While I disagree with the idea that sentences which provide “factual” information, however objective, do not perform speech acts, this designation of theirs does not preclude future comparison between their results and my own, as I would simply categorize those “factual” sentences as additional assertives.

Contrary to the previous two studies, Feng et al. (2006) were primarily aiming to identify the topic of a threaded forum discussion using a variety of information, including speech act analysis. However, though the authors say that they are following the Searle (1969) version of speech act theory, they coded the messages in their threads according to the speech act relations between them (p.211). That is, under a traditional interpretation of what speech acts are, the authors were actually coding something more like adjacency pairs or discourse/dialogue acts.37 This particular study is therefore much less relevant to my own, but it is worth taking a look at their taxonomy and methodology anyway.

37 Another study which I do not cover here, Ravi & Kim (2007), does the same thing, analyzing threaded discussion for speech act relations between messages.
The authors use a basic three-point taxonomy: request, provide information, or interpersonal relationship maintenance. Each of those three were then split into additional subcategories, including things like “command,” “elaborate,” and “simple answer.” They also coded each act according to its relative valence to the previous message (positive, negative, or neutral). The speech act annotations were done manually, and that information was placed into a weighted algorithm with the other features to attempt to automatically identify the topic of the thread. The result of that algorithm was then compared against a baseline system that randomly chose the message in the thread that was most important to topic identification.

2.1.2.1.4 Instant Messaging

Instant messaging is fundamentally different from the media already discussed, as it generally involves a more private interaction between fewer interlocutors. The two studies in this section investigate different aspects of instant messaging. Twitchell & Nunamaker (2004) examine actual IM conversations, looking to create “speech act profiles” for individual interlocutors. Nastri et al. (2006), on the other hand, investigate the speech acts used in the “away messages” that users put on their profiles for their friends to read when they are unavailable.

Twitchell & Nunamaker (2004) used a methodology similar to what we have seen for the other media, using labeled data from the Switchboard-DAMSL corpus to train a classifier in order to “obtain probabilities for each speech act type for a given sentence,” and for a given interlocutor (p.1713). Instead of the variety of potential features that other studies used for training, Twitchell & Nunamaker focused on n-gram language models to generate the probability of a given sentence belonging to each speech act type. The classifier was then tested on instant
messaging conversations, but without the back-up human annotator component that several other studies have used.\textsuperscript{38}

The taxonomy that Twitchell & Nunamaker use is that of the Switchboard-DAMSL corpus, and consists of 42 separate “speech acts,” which are really a combination of speech acts and dialogue acts, but which the authors also say can all be classified into one of Searle’s original five categories. They provide such a classification, noting that none of the 42 fall into the “declaration” category, and that three of the tags are best classified into a miscellaneous “other” category (they are Other, Third-Party Talk, and NonVerbal) (p.1717).

Meanwhile, the study done by Nastri et al. (2006) is a step in quite a different direction. The authors are not attempting to build an automatic classifier, but simply to analyze the pragmatics of the medium, similar to Carr et al. (2012) (and also to my own study). Additionally, even though the data analyzed in Nastri et al.’s study come from an instant messaging application, they are just the “away messages,” which are more akin to Facebook status updates than they are to actual instant message conversations. This study is therefore quite similar to Carr et al.’s study, a comparison bolstered by the fact that they use the exact same taxonomy (in fact, as mentioned above, Carr et al. took their taxonomy from Nastri et al.), and also code for the presence of humor and non-standard orthography in the messages, which were segmented into “punctuation or propositional units,” (p.1033).

Nastri et al. do explicitly acknowledge the limitations of Searle’s taxonomy, such as the fact that mutually exclusive classification “fails to account for the multifunctionality of language use.” But, they state that his taxonomy’s “widely accepted nomenclature and structuralist

\textsuperscript{38} Though the authors acknowledge that this would be a logical follow-up to their study (p.9).
approach provides a useful framework” for an analysis such as theirs (p.1030). Also, as with essentially every other study discussed thus far, the authors mention that effective/verdictive/declarative speech acts are unexpected in their data due to the lack of an “institutional component” to the medium (and indeed, only one effective and no verdictive acts were observed, p.1034).

The authors collected a total of 483 away messages from 44 users over the course of two weeks. Two raters then individually segmented and coded all messages, for which the inter-coder reliability was “high” ($\kappa = 0.90$), while the reliability for the coding of humor was “satisfactory” ($\kappa = 0.83$) (p.1034).

2.1.2.1.5 Email

Compared with the other media mentioned so far, there have been quite a few speech act analyses of email data. Jeong et al. (2009) has already been mentioned, in the section on discussion forums, so all I will add about them here is that the inter-annotator agreement for their email data was slightly better than for the forum data ($\kappa = 0.79$ vs. $\kappa = 0.73$).

All of the studies that I have found in this area are, as in the other media, primarily concerned with constructing and/or testing automatic classification methods for the medium. The team of Carvalho and Cohen have produced several publications in this area, beginning in 2004 with an attempt to classify speech acts in emails “according to an ontology of verbs...and nouns…, which jointly describe” the intention of the sender (Cohen et al., 2004, p.309). Significantly, this noun and verb ontology was built from an analysis of email data, as opposed to motivating a taxonomy theoretically and data-independently (as I have aimed to do for my own study). The authors provide two primary reasons for this decision:
First, we believe that it is more important for the ontology to reflect observed linguistic behavior than to reflect any abstract view of the space of possible speech acts. As a consequence, the taxonomy of verbs contains concepts that are atomic linguistically, but combine several illocutionary points…

Second, we believe that the taxonomy must reflect common non-linguistic uses of email, such as the use of email as a mechanism to deliver files. We have grouped this with the linguistically similar speech act of delivering information (p.2).

Hence, Cohen et al.’s taxonomy looks quite a bit different than the standard Searlian adaptation that we have often seen. There are five main verb categories: *other, remind, deliver, greet,* and *negotiate.* *Negotiate* is then divided into *initiate* (either *propose* or *request*), *amend,* and *conclude* (either *commit* or *refuse*). This verb-noun ontology was used to train and then test classifiers on several different email corpora.

While I appreciate the authors’ reasons for developing a taxonomy in the way that they did, I would point out that this is another case of a taxonomy that is made up of more “dialogue” acts than speech acts. Additionally, I do not think that developing a taxonomy in the way that they did would work well for actual speech acts, as it would only be descriptive as opposed to explanatory (whereas something like Searle’s direction-of-fit-based model is both).

Carvalho & Cohen (2005) continues the work of the (2004) paper, with the goal of adding an element of “context” to their classification system. To this end, the authors develop a set of ordered verb pairs from the verbs in their taxonomy, and add this predictive element to their classifier. Finally, in Carvalho & Cohen (2006), they add an n-gram feature extraction element to improve the model.

### 2.1.2.2 Speech Act Studies in Spoken Conversation

Some of the most well-known corpora of speech acts in spoken conversations have already been mentioned, seeing as several other authors have used or adapted their annotation
schemes: the Switchboard-DAMSL corpus (Jurafsky et al., 1997), and the Meeting Recorder Dialog Act database (Dhillon et al., 2004; Shriberg et al., 2004). As has already been commented on, the taxonomies of both of these studies are actually a combination of speech acts and dialogue acts (unsurprisingly, seeing as both corpora have “dialogue act” in the name). But Twitchell & Nunamaker (2004) provide a reclassification of the 42 Switchboard-DAMSL tags into Searle’s five-point taxonomy, and Dhillon et al. (2004) provide a general correlation between the Switchboard-DAMSL tags and their MRDA tags (pp.3-5). So perhaps comparison is not entirely ruled out.

Jurafsky et al. (1997) took 1,155 telephone conversations from the Switchboard database and manually labeled them using 220 different discourse tags which were then grouped into the final 42 categories. The unit of annotation was the “utterance,” which “roughly corresponds to a sentence.” They do not specify how many annotators tagged each conversation, but they say that “labeling agreement was 84%” (p.90). The authors then built an automatic classifier and tested it on these same data to see how well it could perform.

The classifier used three types of information: 1) Prosodic Information (e.g. “based on the literature we predicted that Yes-No-Questions would be detectable from their final F0 rise,”), 2) Words and Word Grammar (N-gram predictors -- a separate one for each tag trained only on the data for that tag), and 3) Discourse Grammar (“Pick the Dialog Act which is most likely given the surrounding DAs,” -- this element was also trained on the data). Significantly, despite the differences in medium, these features upon which the automatic classifier was based are not terribly different from those used in several of the automatic classifiers for the CMC data, with the obvious exception of “prosodic information” (p.91).
While the Meeting Recorder Dialog Act database is also a corpus of spoken conversations, they are significantly different from the Switchboard conversations as they are face-to-face and multiparty. For this reason, even though they start with the Switchboard-DAMSL tag set, they state that some of those tags do not apply to their data, and that they also needed to develop additional tags for multiparty-relevant actions such as “floor grabbing,” (Shriberg et al., 2004, p.98). The result is 11 “general” tags (plus an “unintelligible” tag), and 39 “specific” tags (a given utterance was annotated for both). Additionally, the main focus of the MRDA project was the manual annotation, as opposed to the building of an automatic classifier (although other studies followed-up with such a project).

The database consisted of 72 hours of speech, from 75 separate multiparty meetings, resulting in a total of over 180,000 dialogue acts. The annotators were responsible for marking three different types of information: 1) dialogue act boundaries (based on pauses, intonation, and units having “different discourse functions”), 2) dialogue act type, and 3) adjacency pairing (“correspondences between DAs”, p.97).

Ang et al. (2005), which includes two of the original developers of the MRDA corpus, take the next step and develop an automatic classifier for this data whose results were then compared to the hand annotations. The classifier was trained on “various lexical and prosodic features,” (p.1061). The authors provide few specifics in this area.

2.1.2.3 Summary

At this point, a version of the annotation task motivated by a distillation of comparable studies (and a desire to be able to compare results with those studies) suggests an obvious and familiar taxonomy: Searle & Vanderveken’s taxonomy, but with institutional acts (i.e.
“declaratives,” “declarations,” “effectives,” “verdictives,” etc.) left out. Cohen et al. (2004) is the only study mentioned above that uses a taxonomy which either does not make reference to Searle, or which other studies have not put into Searlian terms (e.g. Twitchell & Nunamaker (2004) re: the Switchboard-DAMSL tag set).

In short, if we were aiming only for comparison, the annotation task would consist of categorizing units into one of the following (mutually-exclusive) categories: assertive, commissive, directive, or expressive, potentially with an “other” or miscellaneous category to capture unintelligible units and the stray, but highly unlikely, “declarative.”

The main mismatch between the theoretically motivated version of the task and the comparability motivated version is that none of the studies discussed in sections 2.1.2.1 and 2.1.2.2 annotated for both direct and indirect acts. However, this is not as much of a problem as it might appear initially. I will continue with my plan to annotate for both direct and indirect acts. When I reach the point of comparison during the analysis for hypothesis 3, I will include a comparison of the results of these studies with the “highest-level of annotation” for my data. That is, if a unit is annotated for a direct and indirect act, only the indirect act will go into that comparison. This will be further discussed in section 4.2.3.

Additionally, none of the other studies that looked at Twitter segmented their data, instead annotating each tweet as performing only one speech act. When it comes to comparison, there is no real fix for this particular mismatch. This, too, will be further discussed in section 4.2.3.

As for the exact taxonomy of available acts, the comparability motivated version is essentially the same as the theoretically motivated one. The only exception is that the exercitive category is either not included or rebranded for “miscellaneous.”
However, the lack of institutional acts in the studies discussed in sections 2.1.2.1 and 2.1.2.2 does not mean that we should leave “exercitive” out. As a reminder, Alston views exercitive acts quite differently than Searle & Vanderveken view declaratives. Not finding actual institutional acts does not mean that there were no purported institutional acts (i.e. Alston’s exercitives). Additionally, it would be better to include a superfluous (but theoretically motivated) category that receives no annotations than to not include the category and miss out on interesting segments.

2.1.3 Accessibility Motivation

In this section, the whole task will be assessed with respect to intuitiveness, approachability, and practicality, with the goal of making it accessible to the average native speaker. This will include one change to the taxonomy, defining the illocutionary act categories in an approachable way, and defining key concepts like “direct/indirect,” “literal/non-literal,” and “serious/non-serious.”

2.1.3.1 Taxonomy

Regarding changes to the taxonomy, the main issue is the lack of a separate “interrogative” category. Under Alston’s taxonomy, “interrogative” is simply one type of “directive.” However, a pilot of the annotation task given to a set of local non-linguists suggested that this was a very un-intuitive classification for them (see section 3.1.3.1.2). Each of the three piloters independently raised the issue, explaining that putting questions in the same category as directives felt wrong, and it seemed like “questions” should be their own category.

Per this suggestion, I decided to include a separate “interrogative” category in the taxonomy. This decision does not disrupt comparison, as comparison with any study that does
not include interrogative as a separate category can be accomplished by collapsing the “interrogative” and “directive” categories together.

2.1.3.2 Category Definitions

The main issue with the definitions of the illocutionary act categories as given in table 2.1 is that they are completely unapproachable to someone new to the concept of illocutionary act. Also, they are quite long. Annotators need to be given a concise, approachable definition, that still respects Alston’s primary formulations. They do not need to be given long lists of necessary contextual conditions, as their native speaker intuition will recognize those conditions without prompting. In fact, annotators should only need to have the titular qualities of the acts specified (e.g. that “commissive” is about “committing”). To keep in line with Alston, however, it must be clear that these qualities are only purported (e.g. commissive acts only purport to commit the speaker to an action).

To that end, annotators will be given the following description of “speech act” (a more approachable term than illocutionary act) and the definitions of the categories as written in table 2.2, which will be accompanied by examples: The speech act of a segment is what the speaker accomplishes in saying the words. In this task, you will be categorizing these acts into one of six mutually exclusive categories based on what exactly the speaker purports to accomplish.
Table 2.2. Speech act category definitions.

<table>
<thead>
<tr>
<th>Speech Act Category</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assertive</td>
<td>The speaker has purported to <em>make an assertion</em>.*</td>
</tr>
<tr>
<td>Directive</td>
<td>The speaker has purported to <em>give a directive</em>.*</td>
</tr>
<tr>
<td>Interrogative</td>
<td>The speaker has purported to <em>ask a question</em>.*</td>
</tr>
<tr>
<td>Commissive</td>
<td>The speaker has purported to <em>commit to an action</em>.*</td>
</tr>
<tr>
<td>Expressive</td>
<td>The speaker has purported to <em>express a psychological state</em>.</td>
</tr>
<tr>
<td>Exercitive</td>
<td>The speaker has purported to <em>bring about a change in the real world</em> by way of their utterance.</td>
</tr>
</tbody>
</table>

2.1.3.3 Key Concept Definitions

As discussed in section 1.3, the primary questions to be asked of the annotators include certain key terms that need to be defined in an approachable and intuitive manner. These questions are given in 2.1 through 2.3, and the distinctions that need to be defined are bolded.

(2.1) What type of speech act is this segment performing when taken *directly* (whether *literal* or *non-literal*)?

(2.2) Do you think that the direct act is meant to be taken *seriously*?

(2.3) Is this segment performing any other speech act, *indirectly*?

In short, the three pairs of terms to be defined are “direct/indirect,” “literal/non-literal,” and “serious/non-serious.”

The most important of these is the direct/indirect pair. For this reason, it is not enough to simply tell the annotators that the direct act is the one that is more overt and on the surface, while the indirect act (if present) is inferred. The annotators should be given a simple test to perform
to help them judge the difference. For this, I will use a test of deniability.

As touched on in section 1.1.3, because indirect acts are inferred, they can be denied. If challenged on an indirect act, one can always deny it and appeal to the direct act of a segment. However, if challenged on the direct act, a denial is much more difficult. Consider the exchange in 2.4, where the speaker attempts to deny the indirect act.

(2.4)  
John: I couldn’t hear you. (Possible indirect act: Please repeat yourself.)
Steve: You want me to repeat myself?
John: I didn’t say that.

John’s denial is fine. He could go on to appeal to the direct act, saying something like, I was merely asserting my inability to hear you at this time. Now consider the exchange in 2.5, where John attempts to deny the direct act.

(2.5)  
John: I couldn’t hear you. (Possible indirect act: Please repeat yourself.)
Steve: You couldn’t hear me?
John: *I didn’t say that.

John’s denial is awkward and creates confusion. Even if Steve does not use John’s exact words and paraphrases the direct act, the denial is still awkward and confusing:

(2.6)  
John: I couldn’t hear you. (Possible indirect act: Please repeat yourself.)
Steve: You were unable to hear me?
John: *I didn’t say that.

The inability to deny a direct act is perhaps even more clear if it is the force of the act itself that is explicitly challenged. Consider the exchange in 2.7, where Mary asks a question (i.e. performs an interrogative direct act) that may also be performing an assertive indirect act.

(2.7)  
Mary: Why would I want a new bike? (Possible indirect act: I don’t want a new bike.)
Susan: Is that a serious question?
Mary: *I didn’t ask a question.

Mary did, in fact, ask a question. She could have responded that, no, it was not a serious
question (thereby most likely confirming the presence of an indirect act). But to actually deny that it was a question is an awkward response.

Thus, annotators will be provided with the definitions of “direct act” and “indirect act” given in 2.8 and 2.9.

(2.8) The direct speech act of a segment is the one that is on the surface, and undeniable in normal social circumstances. Denial of a direct act is awkward and creates confusion.

(2.9) The indirect speech act of a segment, if there is one, is beneath the surface. It is a speech act performed by way of another speech act (the direct act). Denial of an indirect act is acceptable because its presence is only inferred in the first place.

These definitions will be accompanied by examples, and a demonstration of the deniability test (see appendix 1).

The next pair of terms to be defined for the annotators is “literal/non-literal.” Annotators will not be asked to designate whether a certain direct act is literal or non-literal. But without any guidance to the contrary, annotators may confuse non-literal direct acts with indirect acts.

As discussed in section 1.1.2.3, Alston views non-literal sentences as performing the same direct acts as their literal counterparts. Hence, the use of idiomatic or metaphorical language should not confuse annotators into assuming that the segment is performing an indirect act (though, of course, it could be). Consider the exchange in 2.10, wherein John attempts to deny a non-literal direct act.

(2.10) John: I couldn’t catch that. (Possible indirect act: Please repeat yourself.)
Steve: You couldn’t hear me?
John: ?I didn’t say that.

John’s denial here is less awkward than the denial of the literal counterpart in 2.5, but more awkward than the denial of the indirect act in 2.4. In order to smooth over his response in 2.10,
he would have to go on to plead literality in some way, perhaps claiming that there was something that he was actually trying to physically catch with his hands. Alternatively, he could claim that there was a different idiomatic meaning that the hearer was unaware of.

In short, annotators will be warned against confusing non-literal and indirect, and given examples of non-literal direct acts.

The last pair of terms to be defined for the annotators is “serious/non-serious.” The main point of defining this delineation for the annotators is that they will be asked whether a given direct act is meant seriously. Also, as with non-literal, it seems prudent to clarify that non-serious does not mean indirect. While Alston does not explicitly reference seriousness or non-seriousness of an act, he mentions that “ironic” utterances should be treated the same as non-literal ones (see section 1.1.2.3). Hence, non-serious utterances should be considered to perform the same direct acts as their serious counterparts. Consider the exchange in 2.11.

(2.11) **Steve:** *shouts into John’s ear*
**John:** I couldn’t hear you.

John’s utterance here is clearly non-serious, but performs the same direct act as it does in 2.4 (wherein the same utterance appears to be taken seriously by his interlocutor). Again, annotators will be given examples of non-serious direct acts to help make this distinction clear.

2.1.4 Summary

At this point, the annotation task is almost finalized. The primary questions to be asked, the taxonomy of acts to be used, and the definitions of the speech act categories and other key concepts have all been determined. The final step to designing the annotation task is to determine the exact nature of the segments to be annotated. That is, what counts as a speech act unit? As we will see in section 2.2, answering this question will result in a few additional
segment types which will require their own sets of questions, slightly different from the base set that has been determined so far.\textsuperscript{39}

2.2 Segmentation Guidelines and Final Form of the Task

In this section, I will discuss the issue of what counts as a speech act unit, and develop guidelines for the segmentation of the data based on that information. These guidelines must include instructions on how to handle sentential fragments, embedded sentences, parentheticals, conjoined and/or coordinated elements, discourse particles, hashtags, and emojis/emoticons. Unlike the annotation guidelines, the segmentation guidelines do not need to be approachable to a non-linguist because I performed the segmentation myself.\textsuperscript{40}

First, I will review what guidelines traditional speech act theorists, as well as modern speech act annotators, have used when considering the boundary of a (direct) speech act.

2.2.1 Speech Act Units in Previous Literature

What counts as a “speech act unit” in the literature varies, depending on the theory, the author, and the goal of the research. Austin and Searle were not especially concerned with issues that would readily translate to data segmentation. As they were attempting to offer more generally applicable accounts of speech acts, the problems inherent to noisy data full of fragments and particles did not rank high on the list of topics these authors aimed to address. However, we can still garner from them some basic rules about what constitutes the boundaries of a speech act.

\textsuperscript{39} “Pure quotations” and “hashtags” are already two such segment types.

\textsuperscript{40} As will be explained in section 3.1.2, the initial plan was to use Amazon Mechanical Turk to accomplish the segmentation. However, poor results from that attempt encouraged me to perform the segmentation myself.
The early authors refer to the speech act “of an utterance” or “of a sentence,” without spending too much time explaining what counts as an utterance or sentence. Searle (1969) does state that “...the characteristic grammatical form of the illocutionary act is the complete sentence (it can be a one-word sentence)...” (p.25). Importantly, in all cases (including the one-word sentences), Searle only counts it as an illocutionary act if there is an illocutionary force applied to it. That is, you cannot simply perform an unembedded propositional act (refer and/or predicate) without also applying a force to it. What this suggests for my data is that even sentence fragments would indeed be considered to perform speech acts, as long as the information that would make the sentence complete is easily recoverable or discernable from context.\footnote{All of this, of course, makes the assumption that the form of the data was intentional. It is possible that a certain tweet could contain a sentence fragment that was not meant to be a fragment (and was in fact meant to be part of either the previous or subsequent sentence). As I have no way to determine whether any non-standard features of the data were intentional (i.e. whether or not they are typos), all I can do is acknowledge it as a caveat.}

Bach & Harnish (1979), who spend a little more time on fringe cases, agree:

...an utterance does not have to be grammatical to have been produced with an identifiable illocutionary intent...a speaker can produce a word, a phrase, or a dependent clause and thereby successfully perform an illocutionary act. (p.229)

The authors offer several samples of ungrammatical or simply fragmented utterances that nonetheless are entirely clear in terms of illocutionary act. For example, \textit{Lucerne two-ten low-fat milk} printed on a carton is clearly shorthand for the assertive act of \textit{This carton contains Lucerne two-ten low-fat milk} (p.230). As discussed in section 1.1.2.3, Alston (2000) is also of the belief that fragments and other “utterance surrogates” can perform direct illocutionary acts.
Searle also takes a stand on the status of restrictive relative clauses and complement clauses as regards speech acts. He states that clauses which begin with “that” cannot be considered complete sentences (and hence cannot be considered to perform speech acts by themselves), as they are “a characteristic form for explicitly isolating propositions” (p.29). As previously mentioned, Searle does not believe that merely speaking the words of a proposition amounts to performing a speech act. Searle gives the example of *The proposition that Sam smokes habitually is uninteresting*. He states that the proposition *Sam smokes habitually* is expressed in this utterance, but not asserted. Instead, it is “part of another proposition.” This other proposition is the content of the illocutionary act performed by the utterance, and *Sam smokes habitually* performs no separate speech act. As discussed in section 1.1.2.3, Alston provides the same reason for not interpreting presuppositions as performing speech acts (he does not consider them to be asserted).

In should be noted that this view would only seem to apply to restrictive relative clauses, as opposed to both restrictive and non-restrictive. First of all, non-restrictive relative clauses do not (stereotypically) begin with “that.” More importantly, while non-restrictive relative clauses (and the related phenomenon of parentheticals) are generally considered to be not-at-issue, it is not necessarily the case that they are presupposed. Neither Alston nor Searle directly addresses the question of whether not-at-issue content is presupposed, and hence whether it can perform a speech act. Abbott (2000) argues that non-restrictive relatives are presupposed (p.1431) and Asher (2000) argues that parentheticals can be analyzed as presuppositions (p.31). On the other end, Frazier et al. (2018) have argued that, despite the not-at-issue nature of non-restrictive relatives and parentheticals, these elements can perform speech acts separately from the main
Meanwhile, Esipova (2018) points out that contrastive focus can force “at-issue interpretations of typically not-at-issue content” (p.387).

Bach & Harnish (1979) appear to agree with Searle (and Alstron) about the status of complement clauses, noting that *I order you to leave* performs an order (a speech act of an explicit performative nature, what would elsewhere be termed “effective”), but *I state that I order you to leave* is only a statement and performs no order itself (p.207). Other explicit performatives in complement clauses seem to offer a challenge to this view, since something like *I regret to inform you of your dismissal* seems to successfully perform an effective speech act of informing (while also performing an expressive speech act of regretting). Bach & Harnish do not see this complication as a problem for their theory, however, as they view all explicit performatives as being indirect acts (p.209). For example, they argue that *I order you to leave* is just a statement, but performs the act of ordering indirectly because the hearer has to infer that it is the utterance itself that is issuing the order (and not, say, the pointing of a finger towards the door). Hence, *I regret to inform you of your dismissal* would be, directly, an expressive act -- but indirectly an effective act.

We have seen what the traditional speech act scholars, including Alston, had to say about segmentation and the status of the speech act unit in a theoretical capacity. With this knowledge providing some background coverage, let’s now take a look at how the authors of previous studies of speech acts in computer-mediated communication addressed the problem of data segmentation in their own corpora.

Seeing as the issue of “what counts as a speech act unit” was not the primary concern of these studies, it is perhaps unsurprising that many of them spend little to no time explaining their
rules for segmentation. Many studies simply refer to “utterances” or “sentences” as the unit of classification (not unlike the original speech act theorists themselves), including Asami et al. (2005), Gisladottir et al. (2015), Jeong et al. (2009), Stolcke et al. (2000), and Twitchell & Nunamaker (2004). Carr et al. (2012) explain a bit more in their study of speech acts in Facebook status messages. They state that they first coded the data for speech act units, which they define as “punctuation or propositional units” (pp.183-184). Despite this seemingly vague definition, the authors achieved high intercoder reliability for segmentation. Nastri et al. (2006), looking at speech acts in instant messenger away messages, used the same guidelines. Their example of what exactly “punctuation or propositional units” are is class now, then the gym, which they break into two speech acts, separated by punctuation (p.1033).

Meanwhile, Zhang et al. (2011) and (2012), and Vosoughi & Roy (2016) all investigated speech acts on Twitter specifically. However, as mentioned in section 2.1.2, none of them segmented the tweets into any smaller units. They classified each tweet, as a whole, using their taxonomy (p.88, p.652, and p.711, respectively).

Leech & Weisser (2003), with the goal of developing an automatic speech act recognition program for “task-oriented dialogues,” state the following with regard to their segmentation guidelines:

An utterance may contain a single speech-act unit, or more than one (if one or more of them are discourse-marker or yes or no tags, which may be considered peripheral to the utterance). The speech acts themselves are handled as attribute values of the C-unit, a grammatical unit which corresponds to a speech act in extent, and which is labelled by one of the syntactic class labels “decl” for declarative, “q-wh” for wh-question, etc. (p.443).
The authors do not go on to explain how to identify a C-unit. However, their position on discourse markers and yes/no tags - that these elements constitute separate speech act units - is highly relevant to Twitter data, which contains many of these elements.

2.2.2 Segmentation Guidelines

2.2.2.1 Motivating Segmentation Guidelines

Motivated by the previous literature, but necessarily extrapolating from it, the following sections will develop guidelines for segmentation by addressing the various possible and relevant syntactic phenomena.

2.2.2.1.1 Full Sentences

As a starting point, full sentences consisting of a single independent clause (with no embedded elements or parentheticals of any sort) will be considered separate speech act units. Full sentences with coordinated, subordinate, and/or parenthetical elements (i.e. any sentence that is not a simple, single independent clause) will be addressed, in turn, in the following paragraphs.

2.2.2.1.2 Sentential Fragments

Before moving on to the more complicated syntactic phenomena, it is crucial to address the status of sentential fragments. The literature agrees that fragments can perform illocutionary acts. However, it is not as simple as “one fragmented sentence = one speech act unit,” seeing as complex sentences can be fragmented. For example, consider Nastri et al.’s example class now, then the gym. Reconstructing this into a full sentence yields something like, I am going to class now, then I am going to the gym, a sentence with two coordinated clauses (and, I would say, two speech act units). Though Nastri et al. also concluded that this fragment contained two speech
act units, they did so based on how many “punctuation or propositional units” are in the fragment. I hesitate to rely on tweeters to always use punctuation to break up complex fragments into speech act units. Hence, I do not want the segmentation guidelines to use such a heuristic.

Instead, I will mentally reconstruct fragments (possibly the entire tweet) into full sentences before moving forward with the rest of the guidelines.

2.2.2.1.3 Conjoined and/or Coordinated Elements

Seeing as conjunction/coordination is essentially also a form of abbreviation, like fragmentation, we might be tempted to use the same heuristic for these elements. That is, “reconstruct” conjoined and coordinated elements -- undo any conjunction and coordination to yield separate full sentences. However, this is clearly too general of a rule. Conjunction and coordination can occur at various syntactic levels, and there is no real reason to suppose, prima facie, that a single speech act unit cannot contain conjunction or coordination at ANY level. I will use the $F(p)$ formalism to demonstrate the problem at hand.

Consider a simple coordination such as *I went to the store and to the gym*. Syntactically, we might think of this as a shortened way of saying the two full sentences *I went to the store* and *I went to the gym*. We could therefore consider *I went to the store and to the gym* as performing two separate speech acts. That is, $F(p)$ and $F(q)$. But, we could also consider the two propositions to form a unit together, to which the force applies once: $F(p \land q)$.

The best way to resolve this issue would be to sort out which kinds of elements (which syntactic levels) are capable of performing acts with different forces while coordinated. If a certain type of coordinated element can only ever perform the same type of act as the other

---

42 For example, even without the presence of the comma in *class now, then the gym*, I would still argue that two speech act units are present.
element(s) with which it is coordinated, that suggests that the force is indeed scoping over all of the elements (or rather, their associated elided propositions) together.

However, demonstrating with finality that a certain element cannot have a different illocutionary force than the other elements with which it is coordinated is difficult if not impossible.\textsuperscript{43} Hence, I propose that a good syntactically-based rule is to consider anything below an IP conjunct to be one segment. That is, only break a conjunct apart in the case of IP or above. For example, the following utterances should be considered to have one speech act unit each:

\textbf{(2.12)} Peter, Paul, and Mary wrote \textit{Puff, the Magic Dragon} in 1963. (coordinated DPs)

\textbf{(2.13)} The big, red, hairy dog is named Clifford. (coordinated adjectives)

\textbf{(2.14)} Jane shocked and confused her neighbors. (coordinated verbs)

\textbf{(2.15)} I have walked the dog and fed the cat. (coordinated VPs)

Meanwhile, the following utterances should be considered to have two speech act units each:

\textbf{(2.16)} I wanted to go to that new restaurant tonight, but it had no available openings. (coordinated IPs or above)

\textbf{(2.17)} We saw your lawn and demand that you remove the flamingo immediately. (coordinated IPs or above)

Differentiating a VP conjunct from an IP conjunct, by surface structure alone, will often be difficult, and sometimes impossible. When impossible, the default will be to separate the conjunct. However, before declaring that it is impossible to tell if a segment is a VP or IP conjunct, there are some potential clues that should be double-checked.

\textsuperscript{43} This would require “proving a negative.”
• Do the verbs have different tenses? If so, it is an IP conjunct.

• Is there a comma or other punctuation after the first half of a conjunct? If so, it is usually an IP conjunct. In the case of the coordination of three or more elements, this clue is no longer helpful.

• Do the verbs share an auxiliary? If so, it is a VP conjunct.

• Do the verbs have separate auxiliaries? If so, it is an IP conjunct (even if the auxiliaries are identical, e.g. both have).

If the verbs have the same tense, there is no informative punctuation, and there are no auxiliaries, then I will turn to the default of separating. In sum, in the cases where it is difficult to tell, I will separate the segments unless there is explicit evidence that it is a VP conjunct. If a coordinate conjunction is present, it will accompany the segment that follows it.

2.2.2.1.4 Complement Clauses

I will not consider complement clauses to be speech act units separate from their matrix clauses, following the arguments in section 2.2.1, especially that of Bach & Harnish. While they focused specifically on “explicit performatives” in complement clauses (e.g. I state that I order you to leave), I believe that their argument applies in a more general capacity.

For instance, consider the second half of the previous sentence:

(2.18) I believe that their argument applies in a more general capacity.

The complement clause that their argument applies in a more general capacity is the content of my belief. It would be redundant to say that two separate speech acts are occurring, one being a statement of I believe X, and the other being a statement of X. In stating X, it is clear that I
believe $X$. As Searle put it, clauses which begin with *that* are “a characteristic form for explicitly isolating propositions” (p.29).

As another example, consider:

(2.19) Do you know if Joe will be home by Christmas?

Even without the complementizer *if*, it is clear that *Joe will be home by Christmas* is the content of my (yes/no) question, and not a separate statement. In speaking 2.19, no one would think that the speaker had (directly) both asked a question and made a statement.\(^{44}\)

2.2.2.1.5 Subordinate Clauses and Restrictive Relative Clauses

Seeing as the propositions expressed by subordinate clauses and restrictive relative clauses are generally not-at-issue entailments, they present a distinct difficulty to speech act analysis. These elements seem direct in some ways, but presupposed or indirect in others. For example, consider the following utterance:

(2.20) John ate the rest of Mary’s sandwich after she left for class.

The content of the subordinate clause is *[Mary] left for class*. On one hand, we might say that the speaker is indeed asserting this proposition, committing to its truth and making a direct speech act. He could not cancel it with a following sentence such as *But Mary never left for class*. On the other hand, the proposition is not-at-issue. That is, the sentence is structured such that Mary’s leaving for class is being treated as common knowledge. Or, one might instead argue that the speaker does not necessarily think that Mary’s leaving is common knowledge, but

\(^{44}\) As Bach & Harnish argued for explicit performatives, I would argue that a complement clause might sometimes perform another speech act indirectly. In this example, by asking *Do you know if Joe will be home by Christmas?*, the speaker might be adding information to the common ground -- namely, that Joe is not currently at home -- depending on the context (i.e. depending on whether the hearer knew that Joe was not currently at home). For segmentation purposes, however, this is irrelevant.
rather knows that even if the hearer does not know about the leaving explicitly, they can easily accommodate the information. In either case, the proposition in the subordinate clause is not the main point of the utterance, even if it is new information to the hearer. But, in the case that it IS new information, it seems somehow more at-issue than if it were not.

Restrictive relative clauses act the same way. For example, consider the utterance:

(2.21) John ate the sandwich that I packed for him.

The speaker might be taking it for granted that it is common knowledge that he packed a sandwich for John, or he might not know whether the hearer knows but assume that the hearer can accommodate, or he might even know that the hearer does NOT know and be intentionally forcing the hearer to accommodate. In each case, the fact that the speaker packed a sandwich for John is not the main point of the utterance, though it seems to vary in terms of at-issue-ness.

However, if the subordinate clause is separated from the rest of the sentence by punctuation, it seems somehow more at-issue and less backgrounded, no matter where in the sentence it is placed: After Mary left for class, John ate the rest of her sandwich. The presence of punctuation causes the clause to seem more like a parenthetical addition. As we will see in the next section, parenthetical elements will have the option of receiving a direct speech act classification.

Hence, I will follow a few guidelines here. First of all, an element will be considered a subordinate clause based on the presence or absence of a subordinate conjunction. This is just to take a hard line, since “subordinate clause” can be ambiguous.

Next, to deal with the problem of “at-issue-ness,” I will use punctuation to determine whether or not a subordinate or restrictive relative clause is automatically backgrounded. All
subordinate clauses and restrictive relative clauses will be considered separate segments. However, un-punctuated subordinate clauses and the restrictive relatives (by definition un-punctuated) will automatically be considered backgrounded, and not available to receive a classification for direct speech act.45 In the next phase of data processing, the annotators will be given options to address the question of whether a given backgrounded segment is performing an indirect speech act.

Meanwhile, subordinate clauses that are separated from the main clause by punctuation should follow the same guidelines as parentheticals, which we will see in the next section.

Certain elements are syntactically ambiguous between a reduced restrictive relative clause and a verbal adjective phrase (e.g. I like the painting [?that is?] hanging in the hall). While the semantic difference between the two is negligible or perhaps even non-existent, and even though in either case the element would not be annotated for a direct speech act (because restrictive relatives are considered backgrounded automatically), syntactic classification has been the ultimate determinant for segmentation up to this point and so should it remain. These elements will be considered separate, backgrounded segments.

2.2.1.6 Non-restrictive Relative Clauses and Other Parenthetical Elements

Meanwhile, non-restrictive relative clauses, punctuated subordinate clauses, appositives, stage directions,46 and other parenthetical elements fall into a category together. This is because, while they might contribute the same type of semantic information as un-punctuated subordinate clauses, restrictive relatives, or even just attributive adjectives, they are purposefully separated

45 This guideline also respects Searle’s and Alston’s views that restrictive relative clauses are presupposed and cannot perform direct speech acts.
46 A stage direction is a type of unit that describes an action, as though the poster (or another salient referent) were performing that action as specified (e.g. How fun! *waggles eyebrows*).
from the main clause orthographically. These elements seem somehow less at-issue than the main clause but more at-issue than an un-punctuated subordinate clause or restrictive relative due to the purposeful use of commas, brackets, parentheses, or other punctuation marks to separate the element from the rest of the utterance.\footnote{Meanwhile, spoken language usually marks their separate status with a change in pitch or speed, or with pauses before and after the element.} Consider the following examples:

\begin{enumerate}
\item[(2.22)] My sister, \textbf{who is a college sophomore}, lives in California. (non-restrictive relative)
\item[(2.23)] Jessica, \textbf{a painter}, just sold one of her pieces for a great price. (appositive)
\end{enumerate}

Hence, as with subordinate and restrictive relative clauses, I will consider non-restrictive relatives and other parentheticals to be separate segments. I will also maintain the punctuation-based rule that I applied to the subordinate and restrictive relatives. This means that annotators will always be given the option of declaring that a non-restrictive relative or other parenthetical performs a direct act (as these elements are always separated from the main clause with punctuation). These hard guidelines mean that the segmentation can be categorical while deferring the onus of judgment to the annotation (which is already a task of judgment).

But, again, assigning non-restrictive relatives and other parentheticals a direct speech act will be optional, not mandatory. There are two main reasons for this.

First of all, it is too strong to declare that these elements perform a direct act by default.\footnote{It is worth noting that Frazier et al. (2018), which is cited in section 2.2 and does in fact state that these elements perform speech acts, was not published until after these segmentation guidelines were written (and the segmentation/annotation was performed).} After all, their level of “at-issue-ness” appears to be somewhere on a spectrum between backgrounded elements and a main clause. Which end of the spectrum they are closer to varies.
in a given instance, and is a matter of subjective judgment.\footnote{As Esipova (2018) noted, it is also a matter of other pragmatic factors such as contrastive focus.} Secondly, even though using punctuation as a guideline for making judgments categorical is useful, it might result in some erroneous segmenting, especially if there is erroneous or simply non-standard punctuation in the tweet. Hence, annotators will have the option of saying that an element of this type is actually backgrounded.

It should be mentioned that these rules only apply to parenthetical elements with propositional content. Some parenthetical elements, such as those in the following examples, do not actually offer an additional or explanatory proposition, but instead make a more pragmatic or stylistic contribution to the utterance:

\begin{enumerate}
\item \textbf{(2.24)} \textit{I do, in fact, know your mother.}
\item \textbf{(2.25)} \textit{Stephen, for example, is a terrible cook.}
\end{enumerate}

Parenthetical elements such as these should not be considered separate speech acts all on their own, as I will discuss further in the section on discourse particles.

As we can see, there are now two new segment types to be considered for the annotation (backgrounded, and potentially backgrounded). The various segment types and the types of acts that each will be annotated for will be summarized in section 2.2.3.

\textbf{2.2.2.1.7 Speech Act Units Interrupting Other Speech Act Units}

Parenthetical elements and subordinate clauses can interrupt other speech act units, creating a discontinuous unit. It is therefore important to have a guideline to keep the information organized in such a way that it is clear which is the interrupting unit, and where the interrupting unit was originally located. Interrupting segments will be given their own segment
number, but addressed separately by the annotators. In the interrupted segment, the annotators will be shown the number of the interrupting unit in brackets. They will still, of course, also be shown the entire tweet when annotating any part of it.

**2.26** John’s aunt, a **professional umpire**, got him tickets to the World Series.

1 John’s aunt, [2], got him tickets to the World Series

2 a professional umpire

**2.2.2.1.8 Recursion Guideline**

At this point, we need to develop a guideline to handle recursion. We have seen that subordinate clauses, relative clauses, and parenthetical elements should be considered segments separate from the main clause. However, those segments will not be further segmented. For example, consider the following example, which contains a parenthetical inside of a non-restrictive relative.

**2.27** My sister, **who is a college (or rather, university) sophomore**, lives in California.

A situation like 2.27 in my data would be considered to consist of two segments: *My sister...lives in California*, and *who is a college (or rather, university) sophomore*.

There are two reasons for this rule. First of all, subordinate clauses, relative clauses, and parenthetical elements are already either backgrounded or potentially backgrounded. Separating these elements into multiple backgrounded or potentially backgrounded segments on the chance that they perform separate direct or indirect acts would drain precious annotator attention while possibly providing very little additional information. The second reason to have this recursion guideline is simply to have a hard line to hold, as recursion is theoretically endless.
The recursion guideline also applies to complement clauses. Even though they are not separate segments, they are embedded within the matrix clause. The entire complement, even if it consists of two conjoined clauses, is the content of the matrix verb. Consider the following example.

(2.28) I believe that he is guilty and that he should be impeached.

The complement clause in 2.28 consists of two conjoined clauses. However, they together form the content of the belief. Embedded elements like complement clauses will not be further segmented.

2.2.2.1.9 Discourse Particles

We might be tempted at first to treat discourse particles like sentential fragments -- especially since, in CMC, they are purposefully typed in (as opposed to potentially unconscious verbal “tics” in spoken conversation.) However, their semantic status (i.e. whether they actually contribute semantic content to an utterance) is debatable. Leech & Weisser considered discourse particles to be separate speech act units in their classification, but they placed them into a speech act category labeled “dialogue control.” “Dialogue control” is not a speech act, as has been defined in this study. For example, consider the following utterance:

(2.29) I wanted one, too, you know.

It is not clear what speech act you know would be considered to perform directly here. It seems that you know contributes a more pragmatic, indirect meaning to the utterance in question, the exact form of which depends on the context. For example, if John says this utterance after Steve has taken both egg rolls from their takeout order, the you know would seem to indirectly
contribute some expressive meaning (the content of which would be something like *That was rude*).

Also, consider the same particle in a different utterance:

**(2.30)** Do you, you **know**, want to go get some coffee?

Given the context of asking someone on a date, the discourse marker *you know* in this case seems to be softening or hedging the offer, contributing indirect expressive meaning (the content of which might be something like *I don’t mean to be unduly forward…*). The fact that the contribution of the discourse particle is completely different in examples **2.29** and **2.30** supports the idea that its contribution is more pragmatic than semantic.

The literature supports the conclusion that discourse particles contribute (and constitute) pragmatic, rather than semantic, meaning. Schiffrit (1987) analyzed discourse particles as indexical elements that position an utterance within the local context. Aijmer (2002) points out that, within speech act theory, discourse particles have been described as modifiers of speech acts, or even as illocutionary force indicating devices (p.8). Finally, in her treatise on the semantics and pragmatics of discourse markers, Blakemore (2002) refers to them as “expressions which convey suggestions that are not part of the truth conditional content of the utterances that contain them.”

This being the case, I will not parse discourse particles as separate units by default. However, there is a fine line between discourse particle and sentential fragment. Some units such as *yeah* or *okay* could be either, depending on the context. For example:

**(2.31)** John: Hand me the menu.
    Steve: *Hands John the menu*
    John: Okay, what do you want to order? (particle)
(2.32) **Steve:** Want to get takeout?

**John:** Okay, what do you want to order? (fragment)

In 2.32, *Okay* is performing the direct act of answering a question, and is hence acting like a sentential fragment. As a discourse particle, it should be included in the larger segment. As a fragment, it should be considered a separate segment. The judgment of what is a fragment and what is a discourse particle will be made by me.

2.2.2.1.10 **Interjections**

Interjections should be considered separate segments. Certain elements such as “wtf” and “omg” may be ambiguous between interjection and discourse particle. As a rule of thumb, these elements should be considered discourse particles (and hence not segmented separately) unless they are separated from the rest of the utterance by punctuation (or fully spelled out).

2.2.2.1.11 **Hashtags and @-tags**

The general speech act status of hashtags and @-tags was already addressed in section 1.2.2. However, I have not addressed the issue of tag segmentation and incorporation.

When a user includes a hashtag in their tweet, they have two choices: incorporate the tag, or leave it unincorporated. An incorporated tag is one that is put in such a position that it contributes semantically to a larger unit. For example, given a conference hashtag such as #ACL2016, a user could incorporate it into a tweet like this:

(2.33) Excited to be here at #ACL2016! Looking forward to some great lectures

On the other hand, they could also choose to leave the tag unincorporated, wherein it is NOT occupying a position in a larger unit, such as:

(2.34) Excited to be here! Looking forward to some great lectures #ACL2016
The answer to the question of segmentation here is fairly obvious: incorporated tags should be segmented together with whatever speech act unit they are semantically incorporated into, while unincorporated tags should be considered segments all by themselves. However, all hashtags, regardless of incorporation, will receive independent annotation attention in order to address their potential to perform certain indirect acts discussed in section 1.2.2.

Turning to @-tags, the incorporation possibilities are the same as for hashtags, as show in 2.35 and 2.36 below:

(2.35) @John what are you up to today?

(2.36) What is @John up to today?

The conclusion from section 1.2.2 was that @-tags do not need to receive independent annotation attention because they are all either tagging type (performing an assertive indirect act that indicates the exercitive call/address was intentional), or non-tagging type (which is used only for emphasis), and asking the annotators to judge which of those types a given @-tag is seems highly speculative. Hence, all @-tags were to be assumed to be tagging type.

However, there is an additional non-linguistic function to @-tags. Specifically, they help to thread tweets together in cases where a given tweet is a reply to another one. Users do not add these @-tags themselves, they are a default that occurs when the user selects the “reply to tweet” option. Such @-tags may or may not be intentional vocatives, depending on the context. Consider the exchange in 2.37, wherein the content of Steve’s reply makes it clear that his comment is aimed at John, not Dan.

(2.37) John: what is everybody doing today?
    Dan: @John none of your business dude!
    Steve: @John @Dan what kind of question is that
In Steve’s tweet, @Dan appears because of the tweet’s position in the thread, and does not appear to be an intentional vocative usage.

Therefore, the annotation does actually need to include certain independent annotation attention for @-tags. Specifically, annotators need to be given the options to declare that a given unincorporated @-tag is intended as a vocative and/or functioning to thread tweets together. @-tags that are unambiguously semantically incorporated into adjacent text do not need to receive this independent annotation, as they cannot function to thread.

Hence, as with incorporated hashtags, incorporated @-tags will be segmented together with whatever speech act unit they are semantically incorporated into, while unincorporated @-tags will be considered segments all by themselves. Only unincorporated @-tags will receive independent annotation attention. The annotation plan for each segment type will be summarized in section 2.2.3.

2.2.2.1.12 Direct Address

The @-tag discussion is a convenient segue into the next topic, elements of direct address. In short, vocative acts can also be performed by non-@-tag units. These units are often set off from the rest of a tweet (or other text) by punctuation. However, I will consider instances of clear direct address to be separate segments even if punctuation is missing. For example, the bolded segment in 2.38 is an unambiguous instance of direct address, even without punctuation.

(2.38) Hey John what are you up to today

Certain elements may function as direct address in one context and discourse particle in another (e.g. “man,” “bitch,” etc.). Such an element will only be considered direct address if there is a contextually salient referent. A contextually salient referent will be considered present
if either: 1) the potential vocative occurs in a tweet with an @-tagged user who appears to be the referent, or 2) the potential vocative occurs in a reply tweet, wherein the original poster (or a subsequent replier) is the clear referent.

As for the type of act that these elements perform, the most common perspective in the literature is that vocatives perform the type of act that Alston would term “exercitive.” That is, they purport to bring about an event in the real world. Specifically, that event is either the capture of the addressee’s attention or the maintenance of that attention, corresponding to the two primary types of vocatives - calls and addresses (Zwicky, 1974). Sonnenhauser and Hanna (2013) describe calls and addresses as “vocative functions that in some way nominate the addressee (either by establishing contact or qualifying the addressee)” (p.14). Jansen (2013, in Sonnenhauser and Hanna’s volume) similarly describes the function of vocatives as “to get someone’s attention in order to begin or continue a conversation” (p.221). In speech act terms, Portner (2007) describes these functions of vocatives as “performatives” (in Alston’s terminology, exercitives).

Hence, instances of direct address will be considered to perform an exercitive direct act automatically (something like either I hereby claim your attention as the addressee or I hereby maintain your attention as the addressee), and annotators will not be asked to annotate them for direct acts. These segments will, however, be able receive an indirect classification if the annotator sees fit. This is, therefore, another segment type that will receive a slightly modified set of questions during the annotation.
2.2.2.1.13 Greetings, Goodbyes, and Thank Yous

As with instances of direct address, greetings, goodbyes, and thanks are often separated from the main body of a tweet (or other text) by punctuation, though not always. Nonetheless, I will treat them as separate segments. While my intuition is that these elements generally perform expressive direct acts automatically, their exact structure can vary enough that it seems prudent to have annotators assess these units for direct act category.

2.2.2.1.14 Emoji/Emoticons

Like tags, emoji and emoticons may be either semantically incorporated into adjacent text or left unincorporated. Also, as discussed in section 1.2.1, emoji and emoticons may be either “lexical” or “tone-type.” Lexical emoji/emoticons are meant to stand in for a specific word or phrase, while tone-type emoji/emoticons are meant to affect the interpretation of an adjacent segment. Clearly, emoji/emoticons that are semantically incorporated are automatically lexical. Such emoji/emoticons require no independent annotation attention, as they should be treated like any other word in the corpus.

However, unincorporated emoji/emoticons may be either lexical or tone-type. Unincorporated lexical emoji/emoticons will be annotated like any other regular segment in the corpus, but unincorporated tone-type emoji/emoticons will not. The reason for separating the tone-type emoji out, instead of presenting them as part of an adjacent segment, is that they do not always apply to only an adjacent segment. A tweet-final tone-type emoji can color the whole tweet, whereas a sentential adverb or other sentential modifier only modifies the segment(s) it is syntactically attached to. In general, I will be judging whether a given unincorporated

\footnote{Twitchell & Nunamaker (2004) agree, as they categorize all three of these segment types as “expressive” (p.5).}
emoji/emoticon is lexical or tone-type. In the case that a tweet consists of only emoji/emoticons, they will automatically be considered lexical.

2.2.2.1.15 Laughter Particles

While emoji and emoticons sometimes may function as contextualization cues (Gumperz, 1992) by informing the reader as to how to interpret a segment, laughter particles are always performing this function. In the same way that laughter in face-to-face conversation gives the hearer information on how to interpret a given segment, laughter particles in social media posts provide the reader with context (more so because they are intentionally written). Hence, elements such as haha, lol, and the like will be considered separate segments, treated like tone-type emoji. In the event that a laughter particle constitutes an entire tweet, it will be treated like a lexical emoji.

2.2.2.1.16 Quotations

The speech act status of pure quotations was discussed in section 1.2.3, the conclusion being that they automatically perform a speech act of mentioning, with the potential to perform an indirect act as well. This speech act is something like Someone said X, where the someone is the person given in the quote citation (if one is present) and X is the quoted utterance itself. This act may be, at least, either assertive or interrogative. As for the quote citation, it is always separated from the quote itself by punctuation or at least a new line, and a pure quotation may occur without a citation. Both of these facts suggest that the quote citation should be considered a separate segment.

The question now becomes whether the quote citation performs a direct act. Given that the quotation itself was decided to perform a direct act of mentioning (something like Someone
said X), I propose that the citation performs a similar direct speech act automatically, something like Y is that someone. Hence, this is another new segment type.

It should be noted that no internal segmenting of pure quotations will occur. Seeing as the quoted words are, in my analysis, essentially being understand as the content of a complement clause (i.e. Someone said X), they are covered by the recursion guideline that embedded elements should not be segmented. The recursion guideline also covers cases of recursive quotation, and helps illuminate the difficult case of mixed quotation. For example, consider the mixed quotation in 2.39:

(2.39) He said that he doesn’t want “to get involved.”

The quoted portion of 2.39 is participating semantically in the rest of the utterance. Hence, the quoted material will be considered incorporated in the same way that, say, a hashtag is incorporated. This means that the quoted material will not be considered a separate segment.

Finally, it has already been determined that complement clauses should not be considered segments separate from their matrix clauses. This covers cases of indirect quotation.

One last note about pure quotations. Only elements that are marked as being a quote through the presence of certain punctuation or a quote citation will be considered pure quotations. This is in order to be as consistent as possible. While there are tweets without punctuation or a citation that I may recognize as featuring a famous quote, there are undoubtedly other tweets that are also quoting something or someone (without punctuation or a citation) that I am not familiar with, and that I would leave out if I attempted to include quotes without punctuation or a citation.
2.2.2.2 Summary of Guidelines and Sample Segmented Tweets

The guidelines I’ve proposed in section 2.2.2.1 can be summarized with the following rules:

1) **Independent clauses** (with no embedded elements or parentheticals of any sort) should be considered separate speech act units.

2) **Sentential fragments** should be mentally reconstructed into full clauses before proceeding with the remainder of the guidelines.

3) For **conjoined or coordinated elements**, if they are IPs or above, they will be considered separate segments. If an element is ambiguous on the surface between a VP conjunct and an IP conjunct, the default will be separation.

4) **Complement clauses** will not be considered units separate from their main clause.

5) **Subordinate clauses, relative clauses (both restrictive and non-restrictive), and other (propositional) parenthetical elements** will be considered segments separate from the main clause.

6) **Complement clauses, subordinate clauses, relative clauses, parenthetical elements, and pure quotations** will not be further segmented.

7) **Discourse particles** will be considered part of whatever larger unit with which they are associated. Care will be taken to differentiate a discourse particle from a sentential fragment.

8) **Interjections** will be considered separate segments. Care will be taken to differentiate interjections from discourse particles.
9) **Incorporated hashtags and @-tags** will be considered part of the larger unit with which they are associated.

10) **Incorporated emojis and emoticons** will be considered part of the larger unit with which they are associated.

11) **Unincorporated hashtags, @-tags, emojis, and emoticons** will be considered separate segments.

12) **Vocatives and elements of direct address** will be considered separate segments.

13) **Greetings, goodbyes, and thanks** will be considered separate segments.

14) **Pure quotations and their citations** will each be considered separate segments.

15) **Laughter particles** will be considered separate segments.

For further clarification, I will now apply these rules to a random sample of picture-less and hyperlink-less tweets.

```plaintext
whos the loser who left a dislike on like all my vids like you have to be either rly jealous or bored like....
```

**Figure 2.1. Sample tweet #1.**

1  whos the loser

2  who left a dislike on like all my vids

3  like you have to be either rly jealous or bored like....
Figure 2.2. Sample tweet #2.

All the Ajax in the world ain’t gonna clean you’re dirty laundry

Figure 2.3. Sample tweet #3.

That awkward moment when the mom of the kids you babysit says love you guys and it’s just you & one of the boys lol #PartOfTheFamily #loveu2

Figure 2.4. Sample tweet #4.

Dame Baylor with the L it must suck to BU. 😂

Segments will be recorded verbatim, typos and all.

51
I really want to see @JamesTWmusic perform again. I fell in love with his voice & music when I saw him on the @ShawnMendes tour. Great music

Figure 2.5. Sample tweet #5.

1. I really want to see @JamesTWmusic perform again.
2. I fell in love with his voice & music
3. when I saw him on the @ShawnMendes tour.
4. Great music

im the 🐐 at dancing idc idc 😒😭

Figure 2.6. Sample tweet #6.

1: im the [goat] at dancing
2: idc
3: idc
4: [face with rolling eyes]
5: [loudly crying face]

2.2.3 Summary of Segment Types and Final Form of the Task

The final list of segment types, the annotation questions for each, and the answer options will all be summarized in this section. One of the new segment types requires one additional definition; “potentially backgrounded” requires that we define “backgrounded” for the annotators, seeing as they will be asked if the segment is indeed backgrounded.

52 “I don’t care.”
(2.40) A backgrounded segment is one that provides context or information that is necessary or helpful in understanding the main point of an utterance, but is not itself the main point. These segments do not perform a direct speech act. But they may perform a speech act indirectly.

Annotators will be given an example of a backgrounded segment that likely performs an indirect act.

At last, we have the final form of the task, including key category and term definitions for the annotators. In the next chapter, we will see how this plan was enacted and how the key definitions were used to construct an instructions and training module for the annotators.

The list of possible questions and their answer options is given in 2.41 to 2.55.

(2.41) What type of speech act is this segment performing when taken directly (whether literal or non-literal)?
- Assertive
- Directive
- Commissive
- Interrogative
- Expressive
- Exercitive

(2.42) Do you think that the direct act is meant to be taken seriously?
- Yes
- No

(2.43) Is this segment performing any other speech act, indirectly?53
- Yes
- No

(2.44) What type is [the indirect act]?
- Assertive

53 While it might seem like a research assistant’s ability to make accurate inferences about non-overt acts would be severely diminished when coding the Twitter speech act units of complete strangers, Riordan & Trichtinger (2017) actually found that strangers are no better than friends at accurately interpreting non-overt (specifically, sarcastic) acts in computer-mediated communication (specifically, in email). So while the research assistants will almost assuredly not always interpret the tweets exactly as the Tweeter intended, their interpretations are likely at least generally as accurate as those that would be made by a given Tweeter’s followers, the intended audience of those tweets.
Which act do you think is more important to the speaker’s overall message? That is, which one seems to be the main purpose behind tweeting the segment?
- Direct act
- Indirect act
They are equally important to the speaker’s message.

Is this segment backgrounded?
- Yes
- No

This segment is backgrounded, and therefore does not perform a speech act directly. Does it perform a speech act indirectly?
- Yes
- No

This segment is a direct address, and therefore performs an exercitive direct act by default (something like either “I hereby claim your attention as the addressee” or “I hereby maintain your attention as the addressee”). Does it also perform a speech act indirectly?
- Yes
- No

Which of the following statements apply to this @-tag? You may select either answer or both.
- This @-tag is functioning to thread tweets together.
- This @-tag is intentionally functioning as a direct address or reference, e.g. “I am addressing or referencing you, [@-tagged user].”

---

54 This question was phrased differently during the actual annotation, because my analysis of direct address changed after the annotation was complete. My previous analysis concluded that instances of direct address automatically perform assertive direct acts, not exercitive direct acts (as I have here). Both situations result in direct address segments being annotated for the same thing (indirect acts, if present), but the annotation results may have been different if the annotators had been told that these segments automatically perform exercitive acts (instead of assertive acts).

55 This sentence was phrased slightly differently during the actual annotation. The original version (what the annotators saw) did not include the word “intentionally.” The addition of this word is meant to more accurately reflect how my analysis of the speech act status of @-tags interacts with the speech act status of vocatives.
(2.50) Is this a hashtag that would actually be used? Does the speaker expect that a Twitter user would ever actually click on or search for this tag?
   - Yes
   - No

(2.51) Do either of the following statements apply to this hashtag? You may select one, both, or neither. [Note: The X stands for the string of characters following the hash symbol.]
   - The speaker affiliates with or supports the group affiliated with X.
   - This tweet is a contribution to a meme or trend X.

(2.52) Is this hashtag meant to convey an emotive or meta-commentative meaning?
   - Yes
   - No

(2.53) How would you best classify the speech act of that meaning?
   - Assertive
   - Directive
   - Commissive
   - Interrogative
   - Expressive
   - Exercitive

(2.54) This segment is a quotation, and therefore performs an assertive direct act automatically (something like “Someone said X,” where X is the content between quotation marks). Does it also perform a speech act indirectly?\footnote{This question was phrased differently during the actual annotation, because my analysis of quotation changed after the annotation was complete. My previous analysis concluded that quotations do not perform any direct acts, not that they automatically perform assertive direct acts (as I have here). Both situations result in quotations being annotated for the same thing (indirect acts, if present), but the annotation results may have been different if the annotators had been told that the quotation performs a certain assertive direct act automatically.}
   - Yes
   - No

(2.55) This segment is a quotation citation, and therefore performs an assertive direct act automatically (something like “[Cited person] said that”). Does it also perform a speech act indirectly?
   - Yes
   - No
The general outline of the annotation plan, summarized in terms of which questions will be asked for each segment type, is given below.

Basic non-backgrounded, non-tag segment that does not automatically perform a certain direct act

- 2.41
- 2.42
- 2.43
  - *If answer to 2.43 is yes* 2.44
  - *If answers to 2.42 and 2.43 are both yes* 2.45

Potentially backgrounded segment

- 2.46
  - *If answer to 2.46 is yes* 2.47
    - *if answer to 2.47 is yes* 2.44
  - *If answer to 2.46 is no* 2.41
    - 2.42
    - 2.43
      - *If answer to 2.43 is yes* 2.44
      - *If answers to 2.42 and 2.43 are both yes* 2.45

Backgrounded segment

- 2.47
  - *If answer to 2.47 is yes* 2.44
2.3 Guideline Revision and Motivation for Hypothesis 4

In the course of planning this study, I performed a small test pilot wherein I segmented and annotated 800 random tweets. The goal of this test pilot was to gain an idea of the segmentation and annotation task I was proposing, and assess what the major roadblocks to such a task might be.
Additionally, after an initial segmentation plan was developed and I segmented the data for the present study, the segmentation guidelines were revised to incorporate certain situations that I had not originally planned for. A second pass of the data was then done, to ensure that all of the data aligned with these revisions. The segmentation guidelines outlined above include those revisions.

During both the test pilot and the segmentation process, I repeatedly observed a feature of hashtags that has been observed in the literature as well (see section 1.2.2.1.1). Specifically, I noticed that hashtags which I would have classified as evaluative appeared to be unincorporated much more often than incorporated. This led to hypothesis 4:

(2.56) **Hypothesis 4:** The speech act profiles of unincorporated hashtags vs. the profiles of incorporated hashtags will be distinct, suggesting that "incorporated vs. unincorporated" is pragmatically meaningful.

Comparing the speech act profiles (and other attributes) of these two types of tags may help determine whether a hashtag’s incorporation status is meaningful as opposed to simply incidental, which in turn may inform our understanding of the communicative potential of the hashtag. In order to assess this hypothesis, whether a hashtag is incorporated or unincorporated was recorded by me during segmentation.

2.4 **Motivation for Hypothesis 5**

The last hypothesis is motivated by the general proposal that Twitter users are a linguistic community, as mentioned in section 1.1.4.3. As a linguistic community, we would expect to see users accommodate to a certain communicative style.

Communicative style is the particular manner or variety of language associated with a certain context. One way to characterize this context is through the lens of audience design,
wherein the presence of one audience will evoke one variety of language, and a different audience a different variety. In support of this view, Bell (1984) links the idea of style as audience design to the attested phenomenon of accommodation, wherein speakers converge with their addressee “on a number of levels such as speech rate, accent, content, and pausing” (pp.161-162). On Twitter, accommodation has been previously attested, for example, for lexical choice (Tamburrini et al., 2015; Danescu-Niculescu-Mizil et al., 2011).

In other words, because Twitter is a linguistic community and provides a specific communicative context, accommodation to certain stylistic norms should be observable in my corpus. Which speech act categories are common on Twitter is one potential aspect of what may typify communication on the platform. Other elements may include emoji use, part-of-speech frequencies, and presence or absence of punctuation. Together, such elements add up to the communicative style of Twitter.

In order to observe accommodation, we need a familiarity or convergence metric. This metric will determine how far along the path of accommodation a given user is. Observing phenomena over time is the usual familiarity metric. This is a possibility here, as a given user’s Twitter join date is available on their user profile. “Time since joining Twitter” is therefore one possible familiarity metric.

However, “time since joining Twitter” is not necessarily representative of how familiar a user is with the platform. A person that made a Twitter profile 5 years ago but rarely uses it is likely less familiar with its communicative norms than is a person who made a profile 2 years ago and uses it everyday. Hence, another possible familiarity metric is “total tweets sent,” which
is another piece of information available on a user’s Twitter profile. For the present study, I will look for accommodation both over time and over total tweets sent.

This all leads to hypothesis 5:

(2.57) **Hypothesis 5**: There is a distinct style to Twitter, and we will be able to see evidence for it in the annotation results.

In order to assess this hypothesis, user join date and total tweets sent was gathered during the data collection for all of the original tweet posters.

Observing accommodation to a set of stylistic norms is only the first step. These norms, in addition to conveying one’s membership in the community, can index values of the community. In other words, “stylistic choices are not random but constrained by overriding stylistic strategies that are conventionalized ways of serving identifiable universal human needs” and higher goals, such as politeness (Tannen, 2005, p.24). Hence, if accommodation is observed, the specific elements accommodated to may be able to tell us about the communicative goals and values of the Twitter community.

Previous studies have investigated accommodation on Twitter, as well as what higher goals and values the stylistic choices made may index. Marwick & boyd (2011) is one such study; the authors approached the investigation from the perspective of audience design. They asked how Twitter users construct utterances with their audience in mind when the potential audience for the utterance is hundreds (or thousands) of people from varying social groups and with varying relationships to the poster. The authors termed this problem the “context collapse” of social media. One general solution to this, on the part of the user, is to create “imagined audiences” for their tweets, styling their posts according to what that audience would expect/accept (p.117).
Marwick & boyd investigated what these imagined audiences are like, and how users might cater to them, by asking Twitter users about their strategies directly. Some users reported that their imagined audience was dependent on the tweet, since different topics would catch the interest of different people. Still others reported that they imagined some sort of “ideal” addressee, which is usually someone just like them. Many of the users actually rejected the idea that they tweeted with any audience in mind, claiming that they tweeted mostly for themselves. Marwick & boyd propose that this rejection is a way to distance themselves from the “personal branding” aspect of Twitter, instead positioning themselves as “authentic.” The irony, of course, is that authentic can itself be a brand (as we will see in a moment) (pp.119-120).

There are also the users that embrace the personal branding opportunities of Twitter. Their strategic “micro-celebrity” practices include trying to appeal to a wide variety of people by tweeting about a wide variety of topics. Another technique is using hashtags to direct a given tweet to a relevant audience. Yet another technique is to consciously perform authenticity (e.g. according to one strategic user, “transparency is so chic,” pp.121-122). In short, “...Twitter users negotiate multiple, overlapping audiences by strategically concealing information, targeting tweets to different audiences and attempting to portray both an authentic self and an interesting personality” (p.122).

Authenticity (or performed authenticity) as a reaction to the problem of context collapse makes sense. Presented with the difficulties of trying to consider every possible audience, aiming for authenticity is sort of like opting out (or, at least, pretending to opt out) of the issue instead of dealing with it. It’s like throwing your hands in the air and declaring “I gotta be me!” This practice is not just a characteristic of everyday personal accounts either. Studies have also
found behavior that resembles performed authenticity on Twitter by companies and political campaigns. Li & Wu (2018) found that corporations on Twitter attempt to form interpersonal connections with users through solidarity-building moves like sharing “self-ridicule,” expressing emotions, and even taking a public stance on “issues of public interests” on Twitter (pp.39-40). Similarly, Evans et al. (2014) analyzed the tweets of house congressional candidates’ campaign Twitter accounts in the months leading up to the 2012 election, coding them according to topic (e.g. attacking the opponent, discussion of issues, links to news stories, etc.). They found that, on average, 29% of the tweets coming from these accounts were “personal” in nature: “These tweets were about items not related to their campaigns, such as reflections on the September 11 attacks, pictures of their family and friends, and football games,” (p.457).

Hence, we see that authenticity appears to be one of the higher goals or values that may be indexed by stylistic choices on the platform. Whether and how this goal is realized in the corpus will be discussed in section 4.2.5.2.

An appropriate level of (in)formality is another potential goal of Twitter’s communicative style, though what exactly that level should be is not clear. The formality of a given medium is connected to its ideology and how users compare it to other media. This ideology determines “how casual, familiar, and acceptable [a user] would find any medium when used for a specific communicative task” (Gershon, 2010, p.283). But the ideologies of new media like Twitter are

---

57 Performed authenticity by politicians on social media has only escalated since 2012. In January 2019, Senator Elizabeth Warren, not long after announcing a 2020 presidential bid, live-streamed on Instagram from her home kitchen while casually drinking a beer (Holmes, 2019). Likewise, also in January 2019, Beto O’Rourke (another 2020 presidential candidate) live-streamed on Instagram while in the dentist’s chair (Rosenberg, 2019).
not yet fully developed. Like the media themselves, their ideologies are “emergent and continually evolving,” (Tannen, 2013, p.100).

Thus, it is unsurprising that a finite conception of formality for Twitter has remained elusive to previous studies. Scott (2015) proposes that the prevalence of hashtags on Twitter “facilitate[s] the use of an informal, casual style” in the face of a chaotic discourse context (p.8). On the other end of the spectrum, Hu et al. (2003) compared the style of Twitter with other forms of CMC, concluding that Twitter was “surprisingly more conversative, and less informal than SMS and online chat,” (p.244). Splitting the difference, Maity et al. (2016) concluded that Twitter style is “neither fully formal nor completely informal,” (p.1681). Section 4.2.5.2 will therefore also discuss whether a clear level of formality reveals itself during the analysis of hypothesis 5.
CHAPTER 3: CONSTRUCTING THE CORPUS

3.0 Introduction

Chapters 1 and 2 motivated the decisions made in the formation of segmentation and annotation guidelines for the corpus. In this chapter, I will address the remaining methodological questions, and as well as review some of the initial results and observations to be made from the complete annotated corpus. Chapter 4 will cover more in-depth analyses of the data, in an attempt to provide specific evaluations of the five hypotheses.

3.1 Methodology

Constructing the annotated corpus consisted of three main stages: 1) collecting the data; 2) segmenting the data according to the guidelines laid out in section 2.2.2; and 3) annotating the data according to the final form of the task laid out in section 2.2.3.

3.1.1 Data Collection

The data collection goal was 1000 Twitter conversations (i.e. threads). These were collected periodically over a two-month span from July to September 2016. In order for a thread to be included in the corpus, it needed to satisfy three requirements: 1) the original post and any reply posts must be in English; 2) the original post and any reply posts must not include any pictures or hyperlinks; and 3) the original post and any reply posts must not have been posted by a bot.

Requirement #1 was made because the theoretical literature used to motivate the segmentation and annotation guidelines was written about English, and this is therefore inherently a study of speech acts in English in particular. Including tweets in other languages would have required a different set of guidelines and annotators for each language.
Requirement #2 was made in an effort to reduce the amount of non-linguistic context needed to interpret the linguistic content of a tweet. Also, it would complicate the final format of the corpus if provisions needed to be made for more than simple text.

Requirement #3 was made because this is a study of human language. The goal is to learn about what actual humans tweet, not what pre-programmed bots do. How to identify bots is not necessarily a straightforward process, of course, especially in the cases where it is the goal of the programmer to make the bot appear human. What could be done, however, was to leave self-identifying bots out of the data. That is, any account that overtly stated it was a bot (e.g. in its handle, name, or user profile) was excluded.

In order to get as random of a sample of Twitter data as possible, the initial plan was to collect tweets directly from the Twitter livestream, as they came in. However, at the time of data collection, the user interface did not allow a user to view the entire livestream without a search term. To get around this, I used the advanced search feature, set the language to English, and input a disjunctive search term composed of the 40 most frequent words in the English language. The result was a modified English-only livestream.

For hypothesis 5, questions to be asked included “Which types of speech acts are most likely to receive replies?” and “Which types of speech acts are most likely to receive likes?”. Thus, all of the original posts (the initial tweet in each thread) needed to have been up for the same amount of time (seeing as a tweet which has been up for longer will have had more time to gather replies/likes/etc., which would distort likelihood calculations). Because of the

---

58 This is according to the British National Corpus, a corpus of English composed of over 100 million words from 4,124 files, 90% of which were from written data and 10% of which were from spoken data. (Leech et al., 2001).
fast-moving nature of Twitter, a simple 24 hours was decided on. A batch of 50-75 tweets was collected from the modified livestream, and then re-collected 24 hours later, along with any reply tweets. At that point, if any of the reply tweets violated requirements #1-3, the entire thread was thrown out.

Likelihood calculations such as those mentioned in the previous paragraph also required follower information about the original posters. A tweet by a user with more followers will appear in more Twitter feeds, will therefore be more likely to be noticed, and hence may be more likely to receive replies/likes/retweets, regardless of linguistic context. At the end of the 24 hours, the following user information (essentially, everything available) was gathered from the account profiles of all of the original posters: 1) number of followers, 2) number of accounts the user follows, 3) total number of tweets posted, and 4) when the user joined Twitter (month and year).

There are two main caveats to the user information. The first of these is that not all profiles list when the user joined Twitter. In these cases, a Twitter application was used to see when the user first tweeted, to get an approximation of when they joined. This is not a perfect fix, of course, seeing as when a user joined Twitter (when they made their account) is not necessarily when they became a participating user of Twitter. The second caveat is that if a user deletes a tweet, it affects their overall tweet count -- so that number is not entirely reliable.

---

59 The purpose of this application was simply to show you the first tweet of whichever account you searched for. However, this particular application appears to have been deleted sometime between Fall 2016 and Fall 2018.

60 Additionally, profiles only list the month and year that the user joined, not the day. To get an exact date, I defaulted to the 15th day of the given month. However, if a user’s first tweet took place before the 15th, the first of the month was used. If no first tweet was available, the 15th was used.
This is even more significant because applications exist to delete large batches of tweets, potentially altering that number a lot.

The entire data collection was performed manually. This is because, at the time of collection, there was not a way to use the Twitter API and an automatic scraper to collect entire tweet threads (at least not without inputting specific tweet IDs, which essentially made the process manual anyway). Despite this attention, however, only 11% of the original posts received replies. Thus, the final data collection yielded 1017 Twitter conversations, only 116 of which include reply posts. The final tweet count for the corpus is 1,310 tweets.

3.1.2 Data Segmentation

The data segmentation was also performed manually, using the guidelines laid out in section 2.2.2. The original plan was to use Amazon Mechanical Turk to complete the segmentation, but a pilot posted to Turk yielded poor results. Of the 14 complete responses, only 2 followed the instructions to any reasonable extent, and still not well enough to use their work. It seemed that the guidelines, as they had been designed, were not adequately accessible to a non-linguist worker. Additionally, several workers submitted bad faith responses, simply writing a random answer so that they were allowed to continue and eventually get paid. Hence, I decided to perform the segmentation myself.

The final segmentation yielded 3,679 segments. That number does not represent the total number of segments requiring annotation, however, as it includes incorporated @-tags,

---

61 The original goal of 1000 conversations was overshot in case any data had to be thrown out at some point.
62 For context, I asked a college-educated non-linguist friend of mine to attempt the segmentation task. After reading the guidelines, they candidly informed me that “relative clause” was meaningless to them, and that “clause” alone was not much better.
incorporated emojis, and tone-type emojis (none of which receive independent annotation attention, as per the annotation guidelines). Thus, excluding those three categories, the segmentation yielded 3,142 segments requiring some sort of annotation (i.e. 3,142 segments that received their own question block during the annotation task). Additional information about the segmentation yield is given in table 3.1, below.

Table 3.1. Total number of segments in the corpus.

<table>
<thead>
<tr>
<th>Segments requiring independent annotation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-backgrounded segments</td>
<td>2,114</td>
</tr>
<tr>
<td>Backgrounded segments</td>
<td>223</td>
</tr>
<tr>
<td>Possibly backgrounded segments</td>
<td>83</td>
</tr>
<tr>
<td>Direct addresses (not @-tags)</td>
<td>64</td>
</tr>
<tr>
<td>Hashtags</td>
<td>196</td>
</tr>
<tr>
<td>Unincorporated @-tags</td>
<td>393</td>
</tr>
<tr>
<td>Quotes</td>
<td>42</td>
</tr>
<tr>
<td>Quote citations</td>
<td>27</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,142</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segments not requiring independent annotation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone-type emojis</td>
<td>448</td>
</tr>
<tr>
<td>Incorporated emojis</td>
<td>27</td>
</tr>
<tr>
<td>Incorporated @-tags</td>
<td>62</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>537</strong></td>
</tr>
</tbody>
</table>

| Grand total number of segments                | 3,679 |

163
3.1.3 Data Annotation

The data annotation was completed using Amazon Mechanical Turk, according to the final form of the task laid out in section 2.2.3. The current section will be divided into three subsections: 1) preparing and piloting the task using Qualtrics and AMT; 2) problems/issues encountered during the course of the annotation; and 3) problems/issues that arose after the annotation was complete.

3.1.3.1 Preparing and Piloting the Task

3.1.3.1.1 Preparing the Task

The annotation task was built as a sequence of surveys, using Qualtrics survey software, and uploaded as a sequence of HITs (Human Intelligence Tasks) to Amazon Mechanical Turk. In total, 324 HITs were uploaded to AMT, each seeking completion by 3 distinct annotators. Each Qualtrics survey (each HIT) followed the same general outline:

- Welcome and disclaimer
- Demographic questions
- Instructions and sample annotated tweets
- 8-10 question blocks, corresponding to 8-10 segments to be annotated, and including 1-2 attention and quality control blocks
- Thank you and survey confirmation number

Each will be addressed in turn below.

---

63 The 3 annotators needed only be distinct from the other annotators on a given task, not from all annotators on all tasks.
Welcome and Disclaimer

The purpose of the welcome and disclaimer block was to briefly introduce the task while also informing the annotator under which circumstances they would not receive payment. Also, I wanted the annotators to know up front that the task would probably take them around 15 minutes, but that there were many more of the same type of task up on AMT. So, if they put in the effort to read and understand the instructions once (and passed the attention and quality control questions), they could take on several similar HITs, for the same pay, that would likely take them less time to complete. See figure 3.1, below, for the form and text of the block.
Welcome! Your task today will be to categorize segments of Twitter data into certain linguistic classes using your intuition as a native speaker of English. These classes are known as speech act categories, and you will be given instructions and examples before beginning the task.

You will be asked to annotate between 8 and 10 segments, and the training and task will take about 15 minutes to complete.

Your work may be rejected for the following reasons:

- You did not consent to participate.
- You are not at least 18 years old.
- You are not a native speaker of English.
- You fail to correctly answer one of the Attention and Quality Control questions (there will be one or two such questions dispersed throughout the task, which will be easy to answer if you have read and understood the instructions).

If you successfully complete this task and are interested in additional work of the same kind, there are several more tasks of the same nature (identical instructions but different data) available to do, from the same poster, for the same compensation. Moreover, if you choose to do such additional tasks, they will likely take you less time to complete (seeing as you will already be trained to complete them).

Please note: The data is uncensored, and may include profanity or other content which you find offensive.

By clicking the “Next” button below, you consent to participate in this task.

Figure 3.1. Annotation task welcome block.

Demographic Questions

The purpose of the demographic questions block was to insure that the annotators were over the age of 18 and native English speakers. Annotators that were under the age of 18 or not native English speakers were expelled from the HIT and not allowed to try again. Of course, seeing as this information was self-reported, it is possible that an annotator may not have
responded truthfully to these questions (so that they could continue with the task). As there is no way to know for sure, this fact remains a caveat. See figure 3.2, below, for the form and text of the demographics block.

![Image of form](image)

**Figure 3.2. Annotation task demographic block.**
Instructions and Sample Annotated Tweets

After the annotators confirmed their age and native language, they were given the actual instructions for the task, as well as a few sample annotated tweets. The instructions were modular, with a section for each of the following topics:

- Speech Act Categories
- Direct vs. Indirect Speech Acts
- Serious vs. Non-serious Direct Acts
- Literal vs. Non-literal Direct Acts
- Backgrounded Segments
- Final Notes
- Sample Annotated Tweets

The instructions were a combination of text and simple cartoon images, in order to make them more visually engaging.\(^{64}\) Because they were quite long, they needed to be available for reference during the task itself.\(^ {65}\) To achieve this, there was an optional (non-forced-answer) multiple choice question at the end of each of the 8-10 question blocks (i.e. at the bottom of each page). An annotator needed only select the topic of the desired instruction module(s) for it to appear below the question. (De-selecting the topic would make the module disappear again.)

Additionally, as a quick reference, some of the basic terms and phrases had hover-text definitions in the questions themselves. Terms/phrases with a hover-text definition available

\(^{64}\) The cartoons were a suggestion received in feedback from pilot annotators, as they found the text alone overwhelming.

\(^{65}\) This was also a recommendation from the pilot annotators.
appeared in blue. See table 3.2, below, for the blue terms/phrases as they appeared in one of their respective questions, and their quick-reference hover-text definitions/examples.66

**Table 3.2. Annotation task hover-text definitions.**

<table>
<thead>
<tr>
<th>Blue term/phrase</th>
<th>Hover-text</th>
</tr>
</thead>
<tbody>
<tr>
<td>What type of speech act is this segment performing when taken <strong>directly</strong> (whether literal or non-literal)?</td>
<td>The direct speech act of a segment is the one that is on the surface, and undeniable in normal social circumstances.</td>
</tr>
<tr>
<td>What type of speech act is this segment performing when taken <strong>directly</strong> (whether literal or non-literal)?</td>
<td>For example, ‘I didn’t hear you’ is literal, while ‘I didn’t catch that’ is non-literal. A non-literal act can still be meant seriously.</td>
</tr>
<tr>
<td>Do you think that the direct act is meant to be taken <strong>seriously</strong>?</td>
<td>For example, if I say ‘I didn’t hear you!’ after you’ve shouted into my ear, that is a non-serious (but direct) act.</td>
</tr>
<tr>
<td>Is this segment performing any other speech act, <strong>indirectly</strong>?</td>
<td>For example, in saying the words, ‘I didn’t hear you,’ I may be indirectly requesting that you repeat yourself, by way of asserting that I did not hear you.</td>
</tr>
<tr>
<td>This segment is <strong>backgrounded</strong>, and therefore does not perform a speech act directly. Does it perform a speech act indirectly?</td>
<td>For example, in the utterance ‘I’m gonna get pizza when I leave work,’ ‘when I leave work’ is providing background information for the main point of the utterance (that the speaker is going to get pizza).</td>
</tr>
<tr>
<td>Do either of the following statements apply to this hashtag?</td>
<td>For example, in the hashtag ‘#GeorgetownLinguistics,’ the X would stand for ‘GeorgetownLinguistics.’</td>
</tr>
<tr>
<td>You may select one, both, or neither. [Note: The X stands for the <strong>string of characters following the hash symbol.</strong>]</td>
<td>For example, ‘Touchdown!! #PhiladelphiaEagles,’ or ‘Call your representatives and demand action on gun control #MarchForOurLives.’</td>
</tr>
<tr>
<td>The speaker <strong>affiliates with or supports</strong> the group associated with X.</td>
<td>For example, ‘Touchdown!! #PhiladelphiaEagles,’ or ‘Call your representatives and demand action on gun control #MarchForOurLives.’</td>
</tr>
</tbody>
</table>

66 If a blue term/phrase happened to appear in more than one question, it received the same hover-text in each question.
Table 3.2 (Cont.)

<table>
<thead>
<tr>
<th>Blue term/phrase</th>
<th>Hover-text</th>
</tr>
</thead>
<tbody>
<tr>
<td>This tweet is a <strong>contribution</strong> to the meme or trend X.</td>
<td>These may be either silly or serious. For example, ‘I never got enough sleep #MyHighSchoolYearsInFiveWords,’ or ‘I was 17 and frightened #MeToo.’</td>
</tr>
<tr>
<td>Is this hashtag meant to convey an <strong>emotive and/or meta-commentative</strong> meaning?</td>
<td>For example, ‘What is she wearing?? #BurlapSacksAreAllTheRage,’ ‘Fo Realz, it is blizzarding. #AndiLikel,’ ‘Can’t believe it is finally happening #overjoyed,’ or ‘Dont feel like walking…but ill make it #sigh.’</td>
</tr>
</tbody>
</table>

See appendix 1 for the form and text of the instruction modules and sample annotated tweets. See figure 3.3, below, for the form and text of the instruction reference question that appeared at the bottom of each page.
8-10 Question Blocks, Including 1-2 Attention and Quality Control Blocks

After the annotator had read the instructions, they came to the body of the task: 8-10 question blocks, corresponding to 8-10 segments to be annotated. Each question block had the same visual format. At the top of the page, the annotator was shown an image of the Twitter thread in question. They were then asked various questions about one particular segment in that thread. In the case that a segment’s content was repeated within the thread, the individual instance in question was outlined in red in the image.
In each survey, 1-2 of the question blocks were made into attention and quality control blocks. These blocks had two primary purposes: 1) to prevent bad faith Turkers from indiscriminately selecting answers in order to simply reach the end and get paid; and 2) to prevent Turkers who do not actually understand the task from lowering the quality of the annotation. The annotators did not know which questions would be attention and quality control questions, though they were informed of their existence in the welcome and disclaimer. Failure to correctly answer an AQC resulted in expulsion from the HIT, with no option to try again.67

Deciding which segments to use as AQCs provided an initial difficulty, seeing as declaring that there is a “correct” annotation for a given segment seems counter to one goal of the annotation: to pinpoint potential predictors within the data based on how a majority of speakers interprets the undeniable speech act capacity of the segment (given the provided definitions). This leads to an important question: if I force the annotators to answer a certain way on certain questions, doesn’t that bias the results of the annotation towards my own particular interpretation?

In response: yes, it does. But I determined that the need to prevent exploitation by bad faith Turkers and pollution by inadequate ones was more important. That is, that an absence of AQCs would have lowered the quality of the annotation more than would the slight bias that their presence generated. Nonetheless, I followed a few guidelines in an attempt to lessen the “deck-stacking” effect inherent to designating certain responses as correct:

- Only the “direct force” question of a given block was ever made into an AQC.

The definition of “direct force” that the task is operating under declares that it is

---

67 This was achieved through the “prevent ballot box stuffing” option within Qualtrics, which insures that a given user can only attempt a given survey once.
an overt and undeniable aspect of a segment’s meaning, which should mean that it is more objective than, say, the indirect force.

- Bland segments that seemed as unambiguous as possible were preferred. In general, this meant choosing “full” (or almost full) clause segments with unambiguous syntax, that seemed to objectively perform a given act (and no others, at least not directly). Often, such segments were non-emotive and simply assertive (e.g. *I have an appointment today*), though not exclusively. See figure 3.4 for a direct force breakdown of the AQCs used.

- Which segments were used as AQCs was recorded, and the AQCs of HITs with high expulsion numbers during the annotation were reviewed.

![Figure 3.4. Percentage of attention and quality control questions of each speech act category.](image)

Also, although this is only retroactive support for the AQC guidelines, it is worth noting that the results of the annotation suggest the AQCs were usually well chosen. Figure 3.9, in the next section, breaks down how many respondents needed to attempt each HIT before 3 were
completely successful (i.e. before 3 passed the AQC). This figure indicates that for almost half of the HITs, the first three respondents to attempt the HIT were successful. And for more than half of the remaining HITs, it only took one more respondent to complete them (i.e. more than 75% of HITs were completed with 4 or fewer respondents). In short, the AQC ended up doing relatively little gatekeeping.

Meanwhile, the remaining majority of segments could receive different direct force annotations without expelling the annotators. This meant that we could potentially uncover ambiguous situations in the data. This will be addressed more in section 3.2.1.

The guidelines that were developed in chapter 2 were such that different types of segments required different sets of questions, as laid out in section 2.2.3.  

**Thank You and Survey Confirmation Number**

If an annotator successfully passed the attention and quality control question(s) and completed all other questions, they reached the thank you and confirmation page. See figure 3.5, below, for the form and text of this page.

---

68 To make the survey construction process efficient, and to insure consistency among the surveys, I began by making a template question block for each type of segment. Each template block could then be imported as needed from the Qualtrics block library, and revised to feature the segment in question.

69 When a HIT was submitted as completed, I cross-checked the confirmation number input by the Turker with the confirmation number designated by Qualtrics for that HIT. As long as the numbers matched, I accepted the HIT (and the Turker got paid). The user interface of AMT is such that a Turker who didn’t successfully reach the end of the survey could still submit the HIT with a random number of their choosing. I would know and be able to reject the HIT, however, seeing as the number would not match the confirmation number given by Qualtrics.
Figure 3.5. Annotation task thank you block.

Failure Message

In the event that an annotator was expelled from the task, they were always shown the same failure message. This message explained that they would not be allowed to continue the task for one of the reasons that had been mentioned in the initial welcome and disclaimer. See figure 3.6, below, for the form and text of the failure message.

Figure 3.6. Annotation task failure message.
The final logistical piece of preparing the task was deciding on Turker qualifications and payment. Turker qualifications are requirements within AMT that a Turker must meet in order to even view and attempt the HIT. For this study, three qualifications were designated: 1) the Turker must be located within the US (according to IP address); 2) the Turker must have an approval rating of at least 95%; and 3) the Turker can be an AMT “master.”

The first qualification was chosen in order to immediately reduce the likelihood that a given user was not a native English speaker. The second qualification meant that over 95% of the HITs that the user had submitted in the past were approved. This was meant to reduce the chance that bad faith Turkers would attempt to exploit the HIT. The final qualification was also meant to filter out bad faith Turkers, as well as promote the chance of higher-quality results, as AMT Masters are high-performing Turkers designated by AMT itself. From the AMT website: “Mechanical Turk has built technology which analyzes Worker performance, identifies high performing Workers, and monitors their performances over time. Workers who have demonstrated excellence across a wide range of HITs are awarded the Masters Qualification. Masters must continue to pass our statistical monitoring to retain the Mechanical Turk Masters Qualification.”70 The annotators did not have to be Masters to attempt the HITs in this study, but they could be.

A Turker who successfully completed a HIT received $1.00 for that work. Initially, this was a poor rate of payment ($4/hr), seeing as a given annotator’s first task would likely take about 15 minutes. However, as explained on the welcome and disclaimer page, they only needed

70 https://www.mturk.com/help#what_are_masters
to read the instructions once (and refer back to tricky items as necessary). So, the more tasks the annotators completed, the better their rate of pay became. And indeed, several annotators who went on to complete many of the HITs. In each case, their rate of completion (and therefore their rate of pay) increased as they completed more and more HITs. For example, the worker who completed the most HITs (319) in the corpus took \(\sim 18\) minutes to complete their first one (the pilot) and \(\sim 10\) minutes to complete their second one (#2). But by the time they reached the very last HITs posted, the worker was only taking 2-4 minutes to complete a HIT.\(^71\)

3.1.3.1.2 Piloting the Task

Once a potential version of the task was ready, three local non-linguists were recruited to attempt the task and provide feedback. Their consensus was that the instructions needed to be constantly available for reference, and that they were a bit dense and difficult to get through at first. Per that advice, the instructions were made available at the bottom of each page (as described above), and cartoon images were added to break up the text. These changes received positive feedback from the local piloters.

Additionally, as previewed in section 2.1.3, the local piloters were very troubled by the fact that questions should be considered “directives,” each raising the issue independently. They were then given a revised taxonomy that included a separate “interrogative” category. This change received positive feedback from the local piloters.

A pilot on AMT itself followed, which yielded positive results (i.e. 2 of the Turkers who attempted it managed to pass the AQCs and complete the task, making reasonable annotation

\(^71\) In addition to the costs of paying the Turkers, AMT charges a fee for uploading each HIT, as well as fees for making certain requirements of the Turkers. Opening up the HITs to AMT Masters costs an extra 5% of the reward you are paying the Turker, and the general upload fee is 20% of the reward. So, each individual completed and approved HIT response cost $1.25.
choices). Following this result, the first 30 HITs were uploaded to AMT. These tasks were successfully completed by three distinct annotators, so 30 HITs were uploaded every day until the entire corpus was uploaded.

3.1.3.2 Issues Encountered During Annotation

The annotation itself was generally uneventful, although there were a few minor issues. The first was that many of the HITs were not fully completed within the initial 7 day period of posting, and had to be re-posted or extended.\textsuperscript{72} In most cases, re-posting or extending the HIT was sufficient to gather the final annotation(s) needed for it. In a few cases, successive re-postings and extensions did not result in completion. After these unfinished HITs were left up for several weeks with no change, they were removed from Turk and local non-linguists were recruited to complete the final annotations. In total, 9 of the 324 HITs were incomplete when removed from Turk. Four of these nine HITs, were still in need of two more successful annotations when removed from Turk. The remaining five HITs were only in need of one more successful annotation.

Why it was that these particular 9 HITs remained incomplete even after so long on Turk is not entirely clear, though I suspect at least one of the following reasons: 1) one or several of the “primary” annotators (i.e. the Turkers that invested the time to learn the instructions and then did many of my HITs) failed the AQC\textsuperscript{s} on those HIT\textsuperscript{s}, and with so few of my HIT\textsuperscript{s} left to complete, there was not much incentive for Turkers new to the task to invest the time to learn it; 2) the HIT\textsuperscript{s} had already exhausted the store of potential annotators on Turk that fit the necessary qualifications (i.e. every possible annotator that could try those tasks had tried them); and/or 3)

\textsuperscript{72} Re-posting/extending does not incur additional costs.
the AQCs for those HITs were poorly chosen. In the case of the four HITs that still needed two annotators when removed from Turk, the AQCs appeared to be the main problem. These cases will be further discussed in section 3.2.1.

The other minor issue was one slightly problematic Turker. In all cases, when I noticed that a Turker had completed many of the HITs, I did a few spot-checks of their work (opened the results of their annotations at random to see how they were doing). In one case, one such Turker seemed to be doing a somewhat sub-par job. While they passed the AQCs (they had to, in order to complete the HIT), it looked like they were not annotating many indirect acts, and I was concerned that perhaps they were not making their best effort.

My cautious feelings about this Turker were furthered when they emailed me to tell me they had accidentally submitted a HIT with the wrong code, because they had several of my HITs open at once and had lost track of which window corresponded to which HIT. I felt that this demonstrated a lack of attention which was exactly the sort of thing that the AQCs were meant to prevent against. Thus, I blocked this particular Turker from completing any more of my HITs. Nonetheless, they completed 75 of the 324 HITs before I took this action.

Given my actions, it would be reasonable to think that this Turker’s data should be excluded. However, in viewing the final corpus, it would appear that this particular Turker was not (as I had thought at the time) an outlier in terms of indirect act annotation. Consider figure 3.7, below, which shows the percentage of segments with potential indirect acts that each Turker coded as having no indirect act. The Turker in question is highlighted in blue.
Figure 3.7. Percentage of segments with potential indirect acts that each Turker coded as having no indirect act.

While this figure does show that the Turker in question did code a large percentage of potential indirect acts as non-existent (95%, to be exact), this does not make the Turker an outlier in the data, especially if we compare them to only the most frequent turkers. Consider figure 3.8:
Figure 3.8. Segments with no indirect acts by most frequent Turkers.

Clearly, the Turker in question was not, in fact, an outlier by this metric. I would hazard that my original reason for blocking the Turker was a lack of knowledge that their responses were representative of what would actually turn out to be the norm (coupled with my bias against a worker who told me that they had several tasks open at once). Thus, there should be no problems with including this Turker’s annotations in the corpus.

3.1.3.3 Issues Encountered Post-Annotation

After the annotation was complete, a few more issues arose. The first of these was occasional errors in the surveys themselves. For 12 of the 3,142 segments to be annotated, I imported the appropriate question block template but failed to alter it to apply to the segment in question. When I discovered this while recording the annotation responses, I decided to
consolidate those 12 segments into one new make-up survey, and have my locally recruited annotators complete it. In this way, the data did not have to be excluded.

The second issue was that, in one case, a HIT was accepted when it should have been rejected. Specifically, the Turker was expelled from the task but made up a confirmation code and input it into AMT. I failed to catch the mismatch between this fake code and the code generated by Qualtrics, and the Turker was paid. I discovered the error after all of the HITs had been removed from Turk, so this HIT was added to the list of HITs to be completed by locally recruited annotators.

The third issue was essentially the reverse of the second. According to Turk, only one person successfully completed HIT #137. That is, only one person submitted a confirmation code in order to receive payment. However, Qualtrics shows that two people actually successfully completed the survey. One of two explanations seems likely: 1) for some reason, a Turker forgot to submit their confirmation code, or perhaps lost the code; 2) both responses were by the same Turker, who realized after the fact that they had completed that one before, and decided not to submit their confirmation code the second time. While it was generally not possible for a Turker to take the same survey twice (due to specifications within Qualtrics), it would have been possible if they used a different IP address. In fact, it happened once (to be discussed next). At any rate, seeing as they did not submit their work to Turk, the second “Qualtrics complete” response was not considered a complete and successful annotation.

The fourth issue, as just mentioned, was that in one case a Turker successfully took the survey twice. How this came to pass is unclear, though the likely explanation is that the Turker
simply used two different computers. The “prevent ballot box stuffing” option in Qualtrics relies on IP address. If the Turker used a computer in a different location, or did the survey via mobile device, they would have been able to take it again. Thankfully, this was the only instance of this particular problem, and this Turker’s second response was disregarded.

The fifth issue, which was mentioned briefly above, was that the AQCs for some of the HITs turned out to be poorly chosen upon review. These four HITs had the highest numbers of attempted responses, while also being the only HITs that still needed two more successful annotations when they were removed from Turk. When the locally recruited annotators were also expelled from these HITs, the AQCs were reviewed and these HITs were ultimately excluded from the final corpus. More detail on this will be provided in the next section.

The final issue, which was the most prevalent and least preventable, was that many segments did not receive a unanimous annotation by all three annotators. Moreover, in some cases, a segment actually received three distinct annotations, and hence did not even receive a \( \frac{2}{3} \) majority annotation. This situation will be addressed in the next section.

3.2 Initial Results and Observations

3.2.1 Exclusions and Ambiguities in the Data

This section will briefly discuss the HITs that were ultimately excluded from the corpus, and why. It will also discuss the responses that did not receive unanimous annotation, what that might mean, and how this situation should be addressed in terms of analysis.

First of all, the four HITs that still needed an additional two annotators when they were removed from Turk were also the four HITs with the highest response attempt counts. According

\[73\text{ I do not think this was done in bad faith. This Turker successfully completed many of the HITs, and I suspect simply did not realize that they had already completed this one.}\]
to Qualtrics, 21, 16, 16, and 24 annotators attempted these HITs, respectively -- the highest numbers of respondents for all 324 of the HITs with the exception of the pilot. Consider figure 3.9, below, which shows how many respondents were needed to achieve three complete and successful annotations for the other (complete) 320 HITs (i.e. the four problematic HITs are excluded from this figure).

![Figure 3.9. Percentage of human intelligence tasks with X number of respondents. X is the number of attempts the task required to achieve 3 successful annotations.](image)

For half of the HITs, it only required three attempts to achieve the three complete and successful annotations. That is, half the time, no one dropped out or was expelled. As we move counter-clockwise around the figure, the number of respondents (attempts) slowly increases. The five unlabeled bars in the figure above represent 9 (~1%), 10 (~1%), 11 (~1%), 15 (<1%), and 49 (<1%) respondents, respectively.

If we include the incomplete HITs, and look at the 10 HITs with the highest numbers of respondents, we get figure 3.10, below. Not only do the four incomplete HITs have the highest numbers of respondents (with the exception of the pilot), but -- again -- they are incomplete.
Thus, to get their three complete and successful annotations, they would require even more respondents.

**Figure 3.10. Number of respondents for 10 human intelligence tasks that received 10 or more respondents.** Bars in red represent tasks that failed to receive 3 successful annotations even after local recruitment.

In short, the relatively high number of respondents that these four HITs received suggested that something may have been amiss within the HITs. After revisiting the attention and quality control questions for these four, I concluded that, indeed, they were likely at least part of the problem.

Upon revision, the AQC's for these HITs were generally ambiguous in some way which had gone unnoticed on the first pass. In most cases, several of the annotators expelled from a
given AQC had all made the same annotation choice, suggesting a robust ambiguity (as opposed
to a fluke of bad faith annotation). Additional details, and the exact AQC's in question, can be
found in appendix 1.

These four HITs (and their conversations) were therefore excluded from the final corpus.

Fortunately, the exclusion of this data does not put the corpus below the 1000 conversation
goal. These four HITs comprise 11 conversations in total. Subtracted from the 1017
cussions that were posted on Turk, 1006 conversations still remain. Table 3.3, below,
updates the segment counts to reflect this exclusion.

---

74 I decided not to exclude the data from the pilot because, even though it received 49
respondents, 35 of those dropped out before even attempting the first question. Of the remaining
14, 2 more dropped out later, which leaves 9 to be expelled by the two (reasonable) AQC's
(leaving the 3 successful complete annotations). As the very first HIT to be posted, and likely
the very first HIT that all of those Turkers had tried, I think the higher expulsion rate is
understandable without needing to exclude the data.
Table 3.3. Updated total number of segments in the corpus.

<table>
<thead>
<tr>
<th>Segments requiring independent annotation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-backgrounded segments</td>
<td>2,092</td>
</tr>
<tr>
<td>Backgrounded segments</td>
<td>219</td>
</tr>
<tr>
<td>Possibly backgrounded segments</td>
<td>83</td>
</tr>
<tr>
<td>Direct addresses (not @-tags)</td>
<td>63</td>
</tr>
<tr>
<td>Hashtags</td>
<td>193</td>
</tr>
<tr>
<td>Unincorporated @-tags</td>
<td>388</td>
</tr>
<tr>
<td>Quotes</td>
<td>41</td>
</tr>
<tr>
<td>Quote citations</td>
<td>26</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>3,105</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Segments not requiring independent annotation</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Tone-type emojis</td>
<td>441</td>
</tr>
<tr>
<td>Incorporated emojis</td>
<td>24</td>
</tr>
<tr>
<td>Incorporated @-tags</td>
<td>61</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>526</strong></td>
</tr>
</tbody>
</table>

**Grand total number of segments**

The other issue to be addressed with regard to the final response counts is that of disagreement in the annotation. Many segments received \(\frac{2}{3}\) majority on a given metric, and some received three completely different annotations. In total, 42 segments received three different annotations for direct act category, 90 segments received three different annotations for indirect act presence or category, and 53 segments received three different annotations in some other category.
In fact, a recurring phenomenon was that 2 of the 3 annotators would classify a segment as having X direct force, while the third annotator would classify the segment as having Y direct force, with X indirect force. There were even a few instances of a sort of complete complement: ⅔ said X direct force and Y indirect force, and the other ⅓ said Y direct force and X indirect force.

A segment’s receiving different annotations (whether with a 2-1 majority or a complete 1-1-1 split) can really only be due to one of the following reasons: 1) the “mistake” possibility -- an annotator (or more than one) was not paying attention, didn’t really understand the task or the categories, or didn’t follow the instructions in some other way; or 2) the “ambiguity” possibility -- the segment really is ambiguous between the two or three different annotations.

Regardless of the reason for the lack of consensus, there are three main ways to deal with it when it comes to analysis.

The first way is to take the majority annotation for each question, and analyze that as the corpus. In such a case, questions with three-way split answers would either need to be left out or made to have a majority. Of course, the downside to the majority way of analysis is that it sacrifices subtlety for categoricalness. We lose all of the information about what might actually be ambiguous.

The second way is to analyze the whole corpus “as is,” with all three annotations. Each segment would just appear three times in the corpus. This method preserves the subtleties of ambiguity in the corpus, but also the mistakes.

The third way, and the path to which I will be adhering, is to do both: to create and analyze a “majority-annotation” version of the corpus, but also analyze the whole corpus “as is,”
when relevant. As will become clear in the next chapter, analyzing both versions of the corpus may be one way to address data with marginal agreement scores.

3.2.2 The Majority-Annotation Version of the Corpus

A few general guidelines were followed while making the majority-annotation version of the corpus. First of all, in reviewing the three-way splits for indirect force, if one of the three annotations was “None,” that answer was eliminated first. This was done to minimize my influence on the annotation. If I were to choose “None” in such a case, I would have been disregarding the majority answer to the question of whether the segment performed an indirect act. Likewise in the case that a three-way split for direct force included a “Backgrounded” annotation (that answer was eliminated before choosing between the remaining two).

In the case of three-way splits with no “None” answers, there was usually at least one of the answers that seems to have been a mistake. This answer was eliminated first before choosing which of the other two seemed more accurate.

Even with the guidelines in place, in many cases a given segment was still ambiguous between at least two of the answers, either due to an actual theoretical ambiguity between the categories or due to a dearth of syntactic or semantic information. For example, segment #1058, *fuck it*. yielded three different direct force annotations: directive, expressive, and assertive. Assertive was considered the mistake answer and eliminated, but the choice between directive and expressive was difficult. In my dialect, the phrase “fuck it” can be used as a dismissive (which I would code as expressive), equivalent to “whatever.” However, there is also the phrase “fuck X,” meaning “disregard X,” which is an instruction (a directive). Additionally, “fuck it”

75 The literal interpretation of “fuck it” seemed unlikely here and was immediately eliminated as a possibility, but would have been a directive.
can be used as a sort of rallying cry, equivalent to “YOLO”\textsuperscript{76} or “carpe diem”, which could be either expressive or directive depending on who is meant to be emboldened (oneself vs. another).\textsuperscript{77}

In a few rare cases, a segment was actually ambiguous among all three direct act annotations (at least in my dialect). For example, segment #695, \textit{Like maybe I just wanna chill and go on dates?} was annotated as commissive, assertive, and interrogative. The interrogative interpretation is of course suggested by the question mark, but it is possible that the poster only meant for the question mark to indicate rising intonation on an assertive (in a textual representation of uptalk). Meanwhile, there is also the ambiguity of what counts as commissive versus assertive. Is the poster simply asserting a desire, or indicating a commitment to the fulfillment of that desire?\textsuperscript{78}

In short, it is worth noting that the majority-annotation version of the corpus is certainly reductive in some cases.

3.2.3 Initial Results and Observations

This section will provide some basic data and observations to be made from the results, and discuss what this information may suggest about the methodology itself.

Let’s begin with the direct force breakdown. For the full corpus, for non-backgrounded non-hashtag segments, the majority (64.5\%) were annotated as assertive. The second most

\textsuperscript{76} “You Only Live Once”
\textsuperscript{77} I ultimately coded this segment as expressive, because the dismissive interpretation was enforced by the rest of the tweet, which was simply: \textit{nah,}.
\textsuperscript{78} I ultimately coded this segment as interrogative, because the preceding segment was also interrogative, and the ambiguous segment seemed to merely be a re-wording of the first question (as suggested by the initial “Like”): \textit{Why is it so hard to find a guy who doesn’t just want you for your body or sex?}
frequent category was expressive, with 16.3%. On the other end, the least frequent category was exercitive, with 0.25%. A full breakdown for the full corpus can be seen below, in figure 3.11. It is worth noting that these numbers (and the figure below) include those “possibly backgrounded” segments that were determined by the annotators to not be backgrounded after all. 17.3% of “possibly backgrounded” segments were determined to not be backgrounded (and therefore to have a direct force).

Figure 3.11. Direct force for non-backgrounded non-hashtag segments. Figure does not include automatic direct acts.

For the majority-annotation corpus, the breakdown of direct force is almost identical. For non-backgrounded non-hashtag segments, the majority (66.6%) were annotated as assertive. The second most frequent category was expressive, with 15.0%. On the other end, the least frequent category was exercitive, with 0.1%. A full breakdown for the majority-annotation corpus can be seen below, in figure 3.12. It is worth noting that these numbers (and the figure below) include
those “possibly backgrounded” segments that were determined by the annotators to not be backgrounded after all. 13.7% of “possibly backgrounded” segments were determined to not be backgrounded (and therefore to have a direct force).

![Figure 3.12. Direct force for non-backgrounded non-hashtag segments in majority annotation version of the corpus. Figure does not include automatic direct acts.](image)

The fact that the exercitive category received the fewest annotations is not surprising. As mentioned in the section 2.1.2.1, several other studies of speech acts in CMC have found the same (few or even no exercitives). What is surprising, though, is that the annotators found any. More plainly, I believe that most of the exercitive annotations in the full corpus are mistakes, either in the sense that an annotator accidentally selected the wrong category or that they simply did not fully understand the category’s description. For example, segment #1100, *FISH AND CHIPS*, received two annotations of assertive, and one of exercitive. This segment was a reply to a tweet reading: *Hunt standing out in the middle with one member of staff, no hoodie or umbrella. No idea what he is thinking.* Thus, in context, “fish and chips” seems to be a
proposed answer to what “Hunt” is thinking. But proposing what Hunt is thinking does not make it so (does not put a thought into his head).

The idea that most exercitive annotations in the full corpus were erroneous is further supported by the fact that, of the 16, only 2 of them overlapped. That is, only one segment (out of over 2,000 possible) received a majority (2 out of 3) annotation of “Exercitive” for direct act. I will admit, though, that “Exercitive” is a reasonable possibility in this particular case. The segment in question is **$5.95 Chef Special**, from conversation #590:

![Figure 3.13. Conversation #590.](image)

As an annotator, I would have said that this segment is assertive, simply asserting what the prices of the day are. However, one could instead view the tweet as *setting* the prices, making the tweet function almost like a coupon (especially if we somehow knew that the person tweeting was a person with the power to set prices).

Additionally, in making the majority-annotation version of the corpus, I did ultimately code one of the three-way ambiguous segments as exercitive. The segment was #3351, *Glory be*
to His Holy Name. This segment was coded as expressive, assertive, and of course exercitive. In my view, the expressive interpretation comes from the fact that statements of religious praise are common interjections (e.g. “Thank God!” or “Holy Mother!”), and the annotator may have interpreted this segment as something of the like. Meanwhile, the difference between the assertive and exercitive interpretations is very slight. In both cases, the segment means that praise is being given to a religious figure. The difference is where the glory comes from. If I see a man praying, I may point at him and assert, “Glory be to God,” meaning that the man in question is giving glory to God. However, if I said the same sentence in the course of a prayer, the glory would be coming from me, via the words in question.79

Regardless of these few exceptions, the fact that 13 of the 16 exercitive annotations were likely erroneous is important to note here for one (perhaps obvious) reason: it shows that there are likely other mistakes in the data. If an annotator misunderstood the exercitive category, they may have misunderstood other categories. Similarly, if they accidentally selected exercitive in place of something else, that may have happened elsewhere.

These exercitive examples also highlight real instances of non-erroneous ambiguity shifting the majority annotation. For example, in the tweet shown above in figure 3.13, why would only $5.95 Chef Special receive majority exercitive annotations while $8.95 Steak or Chicken Fajitas and $6.95 Boneless Wings both received unanimous assertive annotations? If an annotator viewed the tweet as setting prices, one would think they would view it as setting all the prices listed, not just one. Similarly, if they viewed the tweet as asserting prices, one would

79 It is for this reason that I ultimately coded this segment as exercitive. It seemed, from the preceding segments in the tweet, that this segment was part of a prayer: Jesus Christ, the Author of Life.
think that to be the consistent interpretation. So were these exercitive annotations simply fortuitous mistakes?

For better or worse, I believe not. The issue here was different annotators. While my original intention in constructing the HITs was to ensure that annotators always addressed the entire tweet, to make sure the interpretations were consistent and to avoid this exact problem, the segment $5.95 Chef Special happened to be one of the 12 segments for which I erred in HIT construction. That is, it was part of the make-up HIT that the locally-recruited annotators completed. Thus, the different available interpretations of this tweet are not only clear but were both strong enough that a separate set of annotators resulted in a separate majority annotation.

While the mistakes mentioned above will hopefully be mostly mitigated by analysis of the majority-annotation corpus, they also may suggest an upper limit to the abilities of crowd-sourced speech act annotation. A brief introduction to speech acts might be sufficient for categories that more directly line up with one of the traditional “sentence types” as taught in grammar school (i.e. assertive, interrogative, directive), but insufficient for less familiar concepts (like exercitive).

Moving on to some indirect act results, among the non-backgrounded non-hashtag segments in the full corpus, 83.9% were determined to perform no indirect act. We can see a breakdown of indirect force below, in figure 3.14. Figure 3.15, directly following it, shows the results less the “None” category.
Figure 3.14. Indirect act response totals. Colors represent direct act of the segment.

Figure 3.15. Indirect act response totals except “none.” Colors represent direct act of the segment.
The results for the majority-annotation corpus were similar, though the number of “None” responses was even higher, and the exercitive and interrogative categories completely fell out. 93.05% of non-backgrounded non-hashtag segments in the majority-annotation version were determined to perform no indirect act. We can see a breakdown of indirect force below, in figure 5.19. Figure 3.16, directly following it, shows the results less the “None” category.

![Figure 3.16. Indirect act response totals in majority annotation version of the corpus. Colors represent direct act of the segment.](image)
Figure 3.17. Indirect act response totals except “none” in majority annotation version of the corpus. Colors represent direct act of the segment.

Most relevant to the assessment of the methodology, perhaps, is the fact that there was so much disagreement in the indirect act annotations. While not necessarily surprising, as I warned the annotators in the instructions that a segment might perform more than one indirect act (and that they should code for what they believe is the most salient one), this result again suggests the upper limit of the methodology. And in particular, of attempting to annotate indirect acts in this manner -- not so much because of potential mistakes, but because indirect acts are inferred. This means they are inherently more subjective than direct acts, and therefore more likely to indicate true ambiguity or polysemy in the segment. While the results of this annotation may hopefully provide a glimpse into some ambiguities (with the differing responses), an annotation focusing on only indirect acts would be able to dedicate more attention to the different possibilities.
Another result that showcases the problems of annotating for direct and indirect acts at the same time is the “complete complement” cases mentioned before, wherein 2 of 3 annotators said X direct force and Y indirect force, and the last one said Y direct force and X indirect force. These cases seem to show agreement in which acts are being performed, but disagreement in how they are being performed. Again, this could be due to mistake or misunderstanding on the part of the annotator. However, they could instead be due to actual ambiguity (and perhaps a spectrum) in the direct/indirect dimension. An annotation that asked participants to annotate indirect acts on segments that had already been annotated for direct acts would likely result in less disagreement on this dimension, though of course it would also thereby lose any hints it could provide about a possible direct/indirect spectrum.

Finally, figures 3.18 through 3.21 show the results for two of the hashtag-specific questions. The first set of graphs, for the full and majority annotation version respectively, address: “Does this tag perform an emotive or meta-commentative act? If so, how would you best qualify that act?” The second set address the “tagging” vs. “non-tagging” hashtag issue. This is the question of whether or not a given hashtag would actually be used (whether someone would click on or search for it).
Figure 3.18. Responses to “Does this hashtag perform an emotive or meta-commentative act? If so, what kind?”

Figure 3.19. Responses to “Does this hashtag perform an emotive or meta-commentative act? If so, what kind?” for the majority annotation version of the corpus.
Figure 3.20. Responses to “Is this a hashtag that would actually be used?”

Figure 3.21. Responses to “Is this a hashtag that would actually be used?” for the majority annotation version of the corpus.
This second set of graphs, of whether or not a hashtag would actually be used, shows another result of the annotation that I found surprising: how many hashtags were annotated as functional tags. In perusing the results of this particular category, I came across several instances in which I disagreed with the annotator (i.e. tags that I would have personally coded as non-tagging, but that they did not).

This result suggests a possible mistake in the task design. The instructions did not really prepare the annotators for the hashtag questions, with the exception of telling them that there would be some, and that they would be similar to the questions for the other segments. Additionally, in this particular question, they were simply asked: “Is this a hashtag that would actually be used? Does the speaker expect that a Twitter user would ever actually click on or search for this tag?” They were not given any examples of tags that might not “actually be used.”

In short, this question, and the categories it invoked, may have been inadequately explained. It is also possible, of course, that the annotators understood the question perfectly, and that they simply have a different idea of which tags a user would actually search for on Twitter. Seeing as there is no way of knowing which of these is the case, this result will simply be accompanied by the caveat that it may not actually be answering the question it was intended to ask. This question will be revisited in the next chapter, in the discussion of the inter-annotator agreement scores.

The next chapter will enter the primary analysis of the results, exploring what answers the final annotated corpus might provide to the hypotheses laid out in chapters 1 and 2.
CHAPTER 4: ANALYSIS AND DISCUSSION OF THE CORPUS

4.0 Introduction

The goal of this chapter is to address the five hypotheses laid out in sections 1.1.4.2, 1.1.4.3, 2.3, and 2.4 through analysis of the complete annotated corpus. Section 4.2 will provide the analysis for each hypothesis in turn, along with discussion. First, however, I will provide and discuss the results of the inter-annotator agreement calculations, and how those results might inform all other analyses of the corpus.

4.1 Inter-annotator Agreement

Inter-annotator agreement calculations can be helpful in assessing the quality and reliability of the annotation results. If all three annotators agree on the answer to a given question, that would seem to support a conclusion that: 1) both the task and the question were clear and accessible, and 2) the chosen answer is the “correct” (or best) one. As mentioned in Chapter 3, it is of course true that there is not always necessarily one best answer (i.e. that some segments may actually be ambiguous in a given dimension, whether due to a dearth of semantic/syntactic information in the tweet or to an actual theoretical ambiguity). A significant amount of any such ambiguity would preclude the strict interpretation of agreement scores as necessarily indicative of annotation quality.

However, having examined all three-way ties in the annotation in the course of whittling the results down to the majority-annotation version, I can say with certainty that the instances in which a segment was truly triply-ambiguous in any dimension were exceedingly rare. That is, when I was making the final majority-annotation choices for the three-way ties, there was almost always one option that was clearly a mistake (whether due to clumsiness or misunderstanding),
and easily ruled out. This means that the vast majority of the time, if there was true ambiguity, it would have only been between two options -- meaning that there should still be 2 annotators in agreement on the segment. In short, despite the presence of some true ambiguity in the data, it should have only rarely led to complete disagreement on a segment.

Another potential complication in calculating inter-annotator agreement comes from the choice of test. As a baseline, I have used Fleiss’s kappa, a variation on Cohen’s kappa that is specifically designed to assess agreement among three or more judges. However, kappa scores assess agreement above what would be expected by mere chance. Several of the questions that the annotators were asked in the course of this study were binary, with only two possible answers (e.g. “Does this segment perform an indirect act?”). This means that the chance agreement for these questions has the potential to be quite high, especially in the case that the responses are not equally distributed between the two possibilities. With observed agreement held constant, a higher chance agreement will result in a lower kappa score.

For these reasons, for each question, I will provide percent chance perfect agreement, percent agreement, and percent perfect agreement calculations in addition to calculations of Fleiss’s kappa. Percent chance perfect agreement is calculated as the sum of the products of the fractions of each type of response by each annotator, as seen in figure 4.1:
Figure 4.1. Calculation for percent chance perfect agreement.

Percent agreement is calculated as the sum of the fraction of annotator pairs that agreed for each observation over the total number of observations, as seen in figure 4.2:

\[
\text{Percent agreement} = \sum_{n}^{N} \left( \frac{\text{number of annotator pairs that agreed}}{\text{total number of annotator pairs}} \right) \times 100
\]

Figure 4.2. Calculation for percent agreement.

Percent perfect agreement is simply the number of observations with all identical answers over the total number of observations, as seen in figure 4.3:

\[
\text{Percent perfect agreement} = \left( \frac{\text{number of observations with perfect agreement}}{\text{total number of observations}} \right) \times 100
\]

Figure 4.3. Calculation for percent perfect agreement.
The combination of these tools should give us an idea of the reliability of the data, despite the presence of non-erroneous ambiguity and the difficulties of assessing agreement among three annotators in a binary dimension.

### 4.1.1 Direct and Indirect Acts

The following section gives the kappa scores for questions applied to regular text segments in the corpus. Scores for questions asked about @-tags and hashtags will be addressed in sections 4.1.2 and 4.1.3, respectively.

#### Table 4.1. Agreement scores for “Is this segment backgrounded?”

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is this segment backgrounded?</td>
<td>2</td>
<td>0.129</td>
<td>56.4%</td>
<td>75.1%</td>
<td>62.7%</td>
</tr>
</tbody>
</table>

The question of *Is this segment backgrounded?* was only asked of segments coded as “possibly backgrounded” during the segmentation phase, so the agreement for that question was only calculated using those 83 segments. In this case, the chance perfect agreement calculation is over 50%, due to a skew of the responses towards “Yes,”\(^{80}\) which contributes to a fairly low kappa score.

---

\(^{80}\) Equal distribution between “Yes” and “No” answers for all three annotators would result in a percent chance perfect agreement of 25%.
Table 4.2. Agreement scores for direct act category.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>What type of speech act is this segment performing when taken directly (whether literal or non-literal)</em>?</td>
<td>6</td>
<td>0.522</td>
<td>30.0%</td>
<td>75.4%</td>
<td>64.3%</td>
</tr>
</tbody>
</table>

The direct force values were calculated using only segments that were not @-tags or hashtags, were not backgrounded or possibly backgrounded, and were not segments used for attention and quality control questions during the annotation (which necessitated a certain answer). Segments with an automatic direct act categorization were direct address segments (automatic exercitive), quote citations (automatic assertive), and quotations (automatic assertive). This left a total of 1640 segments.

If we collapse the directive and interrogative columns, as Alston (2000) would have us do, the kappa score is slightly lower: 0.518. If we collapse directive, interrogative, and commissive, as Searle & Vanderveken (1985) would have us do, the kappa score is still lower: 0.514. If we include segments that were classified as “possibly backgrounded” during the segmentation (as an annotator’s assessment that a possibly backgrounded segment was indeed backgrounded is still a judgment of the direct act status of that segment), there are 7 possible categories,\(^{81}\) and the kappa score is slightly higher: 0.55.

\(^{81}\) Assertive, Commissive, Directive, Interrogative, Expressive, Exercitive, and Backgrounded (indicating that this was a “possibly backgrounded” segment that the annotator judged as backgrounded).
While there is no definitive agreement on what constitutes an acceptable kappa score, all of these direct force scores are moderate at best. However, as noted in the introduction to this section, there is some non-erroneous ambiguity in the data. Additionally, I would reiterate that segments used as attention and quality control questions during the annotation were excluded from these numbers. That is, 452 segments that I judged to be clearly and unambiguously of a certain direct act category were left out of these calculations. Furthermore, as mentioned in section 3.2.1, half of those AQC's received perfect agreement with no interference (i.e. only three attempted responses were necessary to yield three complete responses). In short: several easy segments with perfect agreement were left out of these numbers because the annotators did not technically have a choice of annotation (even though, half of the time, that fact did not matter).

**Table 4.3. Agreement scores for direct act seriousness.**

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you think the direct act is meant to be taken seriously?</td>
<td>2</td>
<td>0.175</td>
<td>51.51%</td>
<td>73.3%</td>
<td>60.0%</td>
</tr>
</tbody>
</table>

The values for the question of whether a given direct act was meant seriously were calculated using all of the segments used for the direct act values, along with any “possibly backgrounded” segments wherein all annotators judged the segment as not backgrounded. In addition, segments used as attention and quality control segments were left in, as the question of *Do you think the direct act is meant to be taken seriously?* was not itself the AQC question. This resulted in a total of 2093 segments.
As with the backgrounded question, the percent chance perfect agreement score here is over 50%, due to a skew of the annotators’ responses towards “Yes,” which contributes to a fairly low kappa score.

Table 4.4. Agreement scores for indirect act presence.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Is this segment performing any other speech act, indirectly?</em></td>
<td>2</td>
<td>0.0212</td>
<td>52.1%</td>
<td>70.0%</td>
<td>55.0%</td>
</tr>
</tbody>
</table>

The values for the question of whether a segment performs an act indirectly were calculated using all annotated non-@-tag, non-hashtag segments, a total of 2524. The chance perfect agreement is again over 50%, this time due to a skew of answers towards “No,” but the kappa score is especially low here.

Table 4.5. Agreement scores for indirect act category.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>What type is [the indirect act]?</em></td>
<td>6</td>
<td>0.421</td>
<td>14.3%</td>
<td>65.2%</td>
<td>56.5%</td>
</tr>
</tbody>
</table>

The values for the question of *What type is [the indirect speech act]?* were calculated using all non-@-tag, non-hashtag segments for which there was perfect agreement that there was an indirect act. Out of a possible 2524, there are only 23 in which there is perfect agreement that there is indeed an indirect act. Among those 23 there was low-to-moderate agreement. This is
one category for which a lower kappa score would be expected, however, as indirect acts are much more ambiguous than direct acts. Annotators were told that a segment may perform more than one indirect act and that they should code for the one they consider most salient.

If we instead calculate these scores using all annotated non-@-tag, non-hashtag segments (all 2524), there are 7 possible answers (including “None”) and the kappa score is much lower: 0.0248, partially due to the previously mentioned skew of answers towards “None” which results in a chance perfect agreement of 52%.

Table 4.6. Agreement scores for act importance to overall message.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Which act do you think is more important to the speaker’s overall message?</td>
<td>4</td>
<td>-0.227</td>
<td>13.2%</td>
<td>31.2%</td>
<td>0.14%</td>
</tr>
</tbody>
</table>

The values for the question of which act was more important were calculated using all segments for which at least one of the annotators was asked the question. That is, if at least one annotator: 1) had indicated that the direct act was meant seriously, and 2) had indicated that there was an indirect act present. If an annotator had not done either of those things, they were not asked the question. Given the low agreement for both the question of seriousness and the question of indirect act presence, it is perhaps unsurprising that this question yielded especially low agreement.
Only 4 segments received unanimous agreement that the direct act was meant seriously and that an indirect act was present. If we instead calculate the scores for this question using only those 4 segments, the kappa score is higher, but still very low: -0.029.

4.1.1.1 Summary

The full table of all the scores in this section, table 4.7, is given below. Of the three binary questions, *Is this segment performing any other speech act, indirectly?* easily has the lowest agreement (comparable percent chance perfect agreement, but lowest kappa score). This is perhaps unsurprising, given the theoretical ambiguity of indirect acts, but likely also reflects the fact that annotators ultimately received very little training in the topic. They were given examples, and resources for reference during the task itself, but no feedback of any kind. This is in contrast to the direct act assignment question, which gave feedback in the form of expulsion from the task (with no re-admittance) for a wrong answer to an attention and quality control question. Annotators expelled from one task could learn from their mistake for future tasks (if they so chose).  

Regardless of the reason for the poor agreement, however, the *Is...indirectly?* scores may negatively impact any conclusions we find for some of the hypotheses, as will be discussed further in section 4.1.4.

---

82 Technically, annotators did not know for certain that it was their direct act answer that caused their expulsion. All questions for a given segment appeared on the same page of the task, and it was only once the annotator attempted to move on from the segment that they were expelled. However, it is possible that the annotators (especially repeat annotators) deduced that it was the direct act question that caused their expulsion, seeing as the “serious” question began with *Do you think....?*, and they were explicitly told that indirect acts are more ambiguous than direct acts.
Another point of interest from table 4.7 is that even though the degree of skew for the answers for all three binary questions is comparable, which answer dominated was not always the same. In order, they were “Yes,” “Yes,” and “No.” It is tempting to conclude that these reflect the default answers to these questions (i.e. that most segments that seem backgrounded are, that most segments on Twitter are serious, and that most segments do not perform an indirect act). However, there is also a more pessimistic explanation. In the case of the “backgrounded” and “indirect” questions, the dominating answer happens to be the one that did not lead to more questions. If an annotator said that a segment was backgrounded, they were then asked whether it performed an indirect act. If an annotator said that a segment was NOT backgrounded, they were first asked to classify the segment’s direct act before moving on to whether it also performed an indirect act. Likewise for the question of whether or not the segment performs an indirect act (a “Yes” requires classifying the act).

Meanwhile, in the case of the “serious” question, neither answer led (immediately) to more questions. This could mean that, in this case, the dominating answer is indeed the true default answer (i.e. that most segments on Twitter are meant seriously). However, this could also be a result of the fact that “Yes” was listed as the first answer.

---

83 A “Yes” on this question combined with a “Yes” on the “indirect” question would together result in a new question: Which act do you think is more important to the speaker’s overall message? That is, which one seems to be the main purpose behind tweeting the segment? However, this question did not appear until after the annotator had indicated that there was indeed an indirect act.
Table 4.7. All agreement scores from section 4.1.1.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Is this segment backgrounded?</em></td>
<td>2</td>
<td>0.129</td>
<td>56.4%</td>
<td>75.1%</td>
<td>62.7%</td>
</tr>
<tr>
<td><em>What type of speech act is this segment performing when taken directly (whether literal or non-literal)?</em></td>
<td>6</td>
<td>0.522</td>
<td>30.0%</td>
<td>75.4%</td>
<td>64.3%</td>
</tr>
<tr>
<td><em>Do you think the direct act is meant to be taken seriously?</em></td>
<td>2</td>
<td>0.175</td>
<td>51.51%</td>
<td>73.3%</td>
<td>60.0%</td>
</tr>
<tr>
<td><em>Is this segment performing any other speech act, indirectly?</em></td>
<td>2</td>
<td>0.0212</td>
<td>52.1%</td>
<td>70.0%</td>
<td>55.0%</td>
</tr>
<tr>
<td><em>What type is [the indirect act]?</em></td>
<td>6</td>
<td>0.421</td>
<td>14.3%</td>
<td>65.2%</td>
<td>56.5%</td>
</tr>
<tr>
<td><em>Which act do you think is more important to the speaker’s overall message.</em></td>
<td>4</td>
<td>-0.227</td>
<td>13.2%</td>
<td>31.2%</td>
<td>0.14%</td>
</tr>
</tbody>
</table>

4.1.2 @-tags

The calculations in the section were performed using only the 388 segments that were designated as non-incorporated @-tags during the segmentation, or a subset thereof.
Table 4.8. Agreement scores for @-tag threading.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Does the following statement apply to this @-tag?] “This @-tag is functioning to thread tweets together.”</td>
<td>2</td>
<td>-0.191</td>
<td>31.9%</td>
<td>51.5%</td>
<td>27.3%</td>
</tr>
</tbody>
</table>

The scores for whether an @-tag was functioning to thread tweets together are particularly low. There are a few causes for this. One is that the ratio of “Yes” to “No” was quite different for the three annotator columns, with annotator 1 favoring “Yes”, and the other two favoring “No”. The splits, respectively, were: 222/166, 11/377, and 93/295. Thus, the annotator 2 column pulled down the “Yes” chance agreement, while annotator 1 pulled down the “No” chance agreement. The other cause for the low score, however, is that the distributions of the responses were also quite different -- worse than chance.

These particular scores are quite unexpected, as this was probably the most objective question that the annotators were asked. Whether or not an @-tag is functioning to thread tweets together is simply a matter of whether or not the @-tag: 1) appears at the beginning of a reply post, and 2) comprises the handle of the previous tweeter (or an account mentioned by the previous tweeter). The fact that this score is so low may suggest that the question was not clear, due to an incorrect assumption on my part as to the transparency of the term “threading,” with

---

84 As a reminder, the three annotator columns do not in fact represent only three unique annotators. Each column is an assortment of several different annotators, which is acceptable for Fleiss kappa calculations.
regard to tweets. The other possibility is that annotators did understand the question, but did not really pay attention to its correct answer.

Table 4.9. Agreement scores for @-tag direct address functionality.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Does the following statement apply to this @-tag?] “This @-tag is intentionally functioning as a direct address or reference.”(^{85})</td>
<td>2</td>
<td>0.252</td>
<td>85.8%</td>
<td>93.1%</td>
<td>89.7%</td>
</tr>
</tbody>
</table>

The question of whether an @-tag was intentionally functioning as a direct address was significantly less problematic than the “threading” question. There is a huge skew of responses towards “Yes,” which results in an extremely high chance perfect agreement score, but in this case the actual perfect agreement is yet higher.

These results are not especially surprising. As discussed in section 1.2.2.1, direct address is considered the primary function of @-tags in the literature. It is therefore no surprise that the “Yes”/”No” ratios for the three annotator columns here were: 346/42, 386/2, and 376/12.

4.1.2.1 Summary

The table of all the scores in this section, table 4.10, can be seen below. Put together, the main conclusion to be drawn from these scores is that @-tags appear to be just as uniform as expected, in terms of function. That is, the vast majority of @-tags do function as direct address.

\(^{85}\) As explained in section 2.2.3, the original form of this question (the form that the annotators saw) did not include the word “intentionally.” It is possible that the addition of this word would have affected the outcome of the annotation.
Meanwhile, the results of the threading question suggest that the annotators’ answers to that question in the corpus should be replaced with the results of the objective algorithm described above. Any @-tag that appears at the beginning of a reply post, and comprises the handle of the previous tweeter (or an account mentioned by the previous tweeter) will be labelled as “threading.”

Table 4.10. All agreement scores from section 4.1.2.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Does the following statement apply to this @-tag?] “This @-tag is functioning to thread tweets together.”</td>
<td>2</td>
<td>-0.191</td>
<td>31.9%</td>
<td>51.5%</td>
<td>27.3%</td>
</tr>
<tr>
<td>[Does the following statement apply to this @-tag?] “This @-tag is intentionally functioning as a direct address or reference.”</td>
<td>2</td>
<td>0.252</td>
<td>85.8%</td>
<td>93.1%</td>
<td>89.7%</td>
</tr>
</tbody>
</table>

4.1.3 Hashtags

All 193 hashtag segments were included in these calculations, regardless of incorporation.

---

86 This is as opposed to being threading or “emphatic” (i.e. non-tagging, see section 1.2.2.2).
87 As explained in section 2.2.3, the original form of this question (the form that the annotators saw) did not include the word “intentionally.” It is possible that the addition of this word would have affected the outcome of the annotation.
Table 4.11. Agreement scores for hashtag usage.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is this a hashtag that would actually be used?</td>
<td>2</td>
<td>0.0761</td>
<td>82.0%</td>
<td>88.9%</td>
<td>83.4%</td>
</tr>
</tbody>
</table>

Continuing a theme from the @-tag results, the scores for the question of whether a hashtag would actually be used are heavily affected by a skew in the response ratio. This time, the dominating response was “Yes,” leading to a chance perfect agreement of over 80%, which contributed to a low kappa score.

Table 4.12. Agreement scores for hashtag affiliation functionality.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Does the following statement apply to this hashtag? ] “The speaker affiliates with or supports the group associated with X.”</td>
<td>2</td>
<td>0.349</td>
<td>27.3%</td>
<td>68.6%</td>
<td>52.8%</td>
</tr>
</tbody>
</table>

The responses for the question of whether a hashtag signals affiliation did not exhibit a significant skew, resulting in a chance perfect agreement score close to 25% (the score if all three annotators agreed on a response that is not likely to occur by chance).

---

88 The annotators were given further explanation that “used” meant: Does the speaker expect that a Twitter user would ever actually click on or search for this tag?

89 Annotators were also told: The X stands for the string of characters following the hash symbol.
annotators’ “Yes”/“No” ratios were 1:1). This fairly even distribution, and a perfect agreement score almost twice that of chance, contributed to low-to-moderate scores for this question.

Table 4.13. Agreement scores for hashtag contribution functionality.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Does the following statement apply to this hashtag?] “This tweet is a contribution to the meme or trend X.”</td>
<td>2</td>
<td>0.221</td>
<td>25.5%</td>
<td>61.3%</td>
<td>42.0%</td>
</tr>
</tbody>
</table>

As with the question of affiliation, the question of whether a hashtag was a contribution to a meme or trend had almost 1:1 “Yes”/“No” response ratios for all three annotators. However, the kappa score is quite a bit lower, suggesting that whether a hashtag was contributing to a meme or trend was more difficult to judge than was affiliation. This is perhaps unsurprising, seeing as the data was collected about 18 months before the annotation took place. The memes/trends popular on Twitter had of course turned over several hundred times in that span. But the low agreement also likely points to an ambiguity between what is a “topic/category” and what is a “trend.” For example, the hashtag #JokesForLinguists following a language-related pun could be viewed as simply a categorization, or it could be seen as a contribution to a trend, in the event that the tweeter saw the hashtag trending and wanted to join in. In short, a trend may only be a trend in context -- and these tweets were all judged far away from their original context, making this a difficult question.
Table 4.14. Agreement scores for hashtag evaluative functionality.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is this hashtag meant to convey an emotive and/or meta-commentative meaning?</td>
<td>2</td>
<td>0.0824</td>
<td>39.5%</td>
<td>63.7%</td>
<td>45.6%</td>
</tr>
</tbody>
</table>

For the question of whether the hashtag performed an emotive and/or meta-commentative meaning, there was a skew in the responses towards “No,” but not as drastic of one as for many of the other binary categories. Even with a relatively low chance perfect agreement, however, the actual perfect agreement is barely higher (not quite 50%), meaning that the kappa score is again quite low.

These results suggest that this was another difficult and/or poorly-defined question, or that annotators were not paying close enough attention. Annotators were given examples of possible emotive/meta-commentative hashtags for reference, but it is certainly possible that these did not provide sufficient guidance.

Table 4.15. Agreement scores for evaluative act category.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss’s kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>How would you best classify the speech act of [the emotive and/or meta-commentative meaning]?</td>
<td>7</td>
<td>0.0713</td>
<td>38.2%</td>
<td>59.6%</td>
<td>43.0%</td>
</tr>
</tbody>
</table>
The scores for emotive/meta-commentative act type are quite similar to those from the previous question, because they were calculated on the same set of segments (with “None” as a response option), as only 7 out of a possible 193 hashtags were unanimously annotated as even having an emotive/meta-commentative act.

Essentially, this low score is a continuation of the problems from the previous question. If we calculate kappa based on just the 7 segments for which there was agreement that such an act existed, it is quite a bit higher, but still only low-to-moderate: 0.318.

4.1.3.1 Summary

The full table of all the scores in this section, 4.16, can be seen below. Put together, one initial conclusion that may be drawn is that hashtags are not nearly as uniform in function as are @-tags (some skew in the binary answers, but not much). This is not surprising, given the literature on tags presented in section 1.2.2, but it is worth noting the confirmation of that expectation.

However, the results for the binary hashtag questions are not surprising, seeing as I analyzed and presented the different proposed hashtag uses as indirect acts that may or may not be present. Thus, it makes sense that they generated agreement scores similar to those for the non-hashtag indirect presence question. I do not believe that this poor agreement reflects disagreement regarding the existence of the phenomenon in general, just disagreement about specific instances. For example, I do not believe that “non-tagging” hashtags do not exist simply because agreement over which ones were “non-tagging” was poor. But whether or not a specific hashtag is “non-tagging” is inferred, leaving plenty of room for disagreement.
Nonetheless, the overall low kappa scores here could negatively impact any conclusions we may make regarding some of the hypotheses. This will be detailed more fully in the following section.

Table 4.16. All agreement scores from section 4.1.3.

<table>
<thead>
<tr>
<th>Question</th>
<th># of possible choices</th>
<th>Fleiss's kappa value</th>
<th>Percent chance perfect agreement</th>
<th>Percent agreement</th>
<th>Percent perfect agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Is this a hashtag that would actually be used?</em></td>
<td>2</td>
<td>0.0761</td>
<td>82.0%</td>
<td>88.9%</td>
<td>83.4%</td>
</tr>
<tr>
<td><em>[Does the following statement apply to this hashtag?]</em> “The speaker affiliates with or supports the group associated with X.”*90</td>
<td>2</td>
<td>0.349</td>
<td>27.3%</td>
<td>68.6%</td>
<td>52.8%</td>
</tr>
<tr>
<td><em>[Does the following statement apply to this hashtag?]</em> “This tweet is a contribution to the meme or trend X.”</td>
<td>2</td>
<td>0.221</td>
<td>25.5%</td>
<td>61.3%</td>
<td>42.0%</td>
</tr>
<tr>
<td><em>Is this hashtag meant to convey an emotive and/or meta-commentative meaning?</em></td>
<td>2</td>
<td>0.0824</td>
<td>39.5%</td>
<td>63.7%</td>
<td>45.6%</td>
</tr>
<tr>
<td><em>How would you best classify the speech act of [the emotive and/or meta-commentative meaning]?</em></td>
<td>7</td>
<td>0.0713</td>
<td>38.2%</td>
<td>59.6%</td>
<td>43.0%</td>
</tr>
</tbody>
</table>

4.1.4 Influence of the Agreement Results on the Analyses

In order to thoroughly detail how the agreement results may affect the analyses of the various hypotheses, I will look at each hypothesis in turn.

90 Annotators were also told: *The X stands for the string of characters following the hash symbol.*
**Hypothesis 1:** Certain overt elements available to Twitter users function as direct illocutionary force indicating devices (IFIDs, from Searle & Vanderveken (1985)).

The most important annotation question for the assessment of hypothesis 1 is the direct act category, which (as has been mentioned) had the best agreement of all the questions. Hypothesis 1 will be assessed using both the full corpus and the majority-annotation corpus, in order to allow for potential ambiguity. As far as the potential IFIDS (i.e. the explanatory variables), they are mostly objective and were not annotated for. The two exceptions are the “serious yes/no” and “indirect act choice” variables. The fact that these categories had low agreement scores will not affect the analysis of hypothesis 1 directly, though it may mean that these two variables prove to be insignificant, causing them to drop out during the model selection phase.

**Hypothesis 2:** Certain overt elements available to Twitter users signal the presence of an indirect act.

The annotation question most relevant for the assessment of hypothesis 2 is indirect act presence. Seeing as this question yielded quite poor agreement, the search for overt elements that signal such a presence is generally compromised. However, the poor agreement itself supports a related conclusion: that such overt elements do not, in fact, exist (at least not as systematic signals that function consistently regardless of context). For the sake of completion, I will perform the analysis as planned, but I will add the caveat that the results must be viewed with a skeptical lens.

**Hypothesis 3:** We will see a speech act landscape for Twitter that is distinct from that of other social media and certain spoken genres.
In general, this hypothesis will be explored by comparing the direct act distributions to what has been found for other media. Seeing as the direct act question produced a moderate kappa score, the best of all the questions, this hypothesis will be perhaps the least affected by the agreement results. Nonetheless, it will be assessed with the full corpus in addition to the majority-annotation corpus, in the hope of accounting for some of the previously mentioned ambiguity.

**Hypothesis 4:** *The speech act profiles of unincorporated hashtags vs. the profiles of incorporated hashtags will be distinct, suggesting that "incorporated vs. unincorporated" is pragmatically meaningful.*

This hypothesis is perhaps the most affected by the agreement results, as the hashtag agreement scores were fairly low across the board (possible exception for the question of affiliation). The matter of whether a hashtag was incorporated or unincorporated is unaffected, as that was a matter for the segmentation phase and was not an annotation question. However, all of the categories intended for comparison to incorporation status were annotation questions.

Instead of completely abandoning the original assessment plan, I will perform the analysis as planned (comparing the annotation categories with incorporation, with caveats noted), but will also look at incorporation vs. some objective categories (e.g. position in the tweet, adjacency of emoji, length of hashtag).

**Hypothesis 5:** *There is a distinct style to Twitter, and we will be able to see evidence for it in the annotation results.*

The original analysis plan for this hypothesis was to look at the annotation categories, in addition to several simply factual categories (e.g. number of emojis in the tweet), over time,
noting the ones that change significantly. Such an analysis is still possible, but categories with low agreement scores will be noted, as will any categories that may interact with a low-agreement category.

4.2 Analysis and Discussion by Hypothesis

4.2.1 Analysis and Discussion of Hypothesis 1

**Hypothesis 1:** Certain overt elements available to Twitter users function as direct illocutionary force indicating devices (IFIDs, from Searle & Vanderveken (1985)). Candidate elements include lexical choice, capitalization presence and/or choice, punctuation presence and/or choice, emoji presence and/or choice, tag presence and/or choice, unit position within the tweet, and status as an original post vs. reply post.

This hypothesis will be assessed through logistic regression and random forest techniques, with direct act category as the response variable. The list of explanatory variables, the potential IFIDs, was mostly composed of items pulled from previous studies whose goals were to build speech act classifiers (some for written media, some for spoken). The list, and any explanations for the inclusion of each variable, will be given in the analysis section. Any variables that are shown by the model to be significant for classification will then be addressed in the discussion section. The discussion section will review any relevant variable-specific hypotheses and attempt to propose some potential explanations for the associations observed, before stepping back and assessing the significance of the results as regards speech act interpretation and classification.
4.2.1.1 Analysis

Before addressing the form and results of the analysis, it is crucial to review the various explanatory variables being tested, any motivations for their inclusion, and any interactions among them. A wide variety of variables were tested, due to the vague nature of Searle & Vanderveken’s original characterization of IFIDs. In terms of concrete examples of what might be an IFID, the authors merely said that they appeared “in the syntax of actual languages in a variety of ways, e.g. mood, punctuation, word-order, intonation, contour, and stress,” (Searle & Vanderveken, 1985, pp.1-2). Both lexical and contextual variables are given there, and of course no specific correlations with exact speech act categories are mentioned. For this reason, I have included a variety of both lexical and contextual variables, mentioning which other studies included them when relevant.

I have also provided some variable-specific hypotheses, regarding which category (or categories) we might expect to find correlates with each variable. Some of these hypotheses come from the study from which I took a given variable, and some are proposed here. There are two types of variable-specific hypotheses to be made here: 1) those that are grounded, in the sense that they make sense from a perspective of reasoning (e.g. expressive acts may have more emotion words because emoting is a type of expressing); and 2) those that are not necessarily grounded, but are instead an example of an established convention (e.g. interrogatives will end with a question mark because that is the conventional purpose of the question mark). The outcomes of these variable-specific hypotheses, and whether each outcome is expected with regard to reason and/or convention, will be assessed in section 4.2.1.2.1 for those variables that survive model selection.
Several variables include “fraction” in their name. All of these variables, except one, were obtained by dividing the pure count of the phenomenon by the number of tokens in the segment. For example, “Emotion word fraction” refers to the number of high emotion/sentiment words in the segment over the total number of tokens in the segment. The exception is “Emoji in tweet fraction,” which refers to the total number of emoji in the encompassing tweet divided by the total number of tokens (words and emoji) in the encompassing tweet.91

I have also included two variables that are not actually overt, but are instead inferred. Specifically, whether the segment was annotated as “serious” or not, and the indirect act choice of the segment (including “none” as an option). I have included them primarily because if they correlate with direct act classification, this could support a conclusion that inferred information interacts with classification (alongside, potentially, overt information).

One final note should be made about the response variable, direct act category. As detailed in section 2.2.3, some types of segments were judged to perform certain direct acts automatically. For example, hashtags were judged to automatically perform exercitive direct acts, and quote citations were judged to automatically perform assertive direct acts. For this analysis, only segments that performed non-automatic direct acts were included (i.e. only segments whose direct act category was determined by annotation). This is because of the fundamental difference between automatic and non-automatic acts.

Automatic acts were designated only for cases of absolute certainty, as determined by the presence of certain characters (e.g. #, @, quotation marks) that classified the segment into a very

91 Fractions were used instead of pure counts because count features conflate length, which was a separate feature. That is, a longer segment has more opportunities to include a given category. Fractions adjust for this fact.
specific syntactic category (e.g. hashtag, direct address, quote citation). Those classifications already tell us what the communicative function of the segment is, and the automatic direct act is just the umbrella category for that function (e.g. exercitive is the umbrella category for direct address). Meanwhile, the communicative function of a segment with a general syntactic category such as “clause” or “fragment” is not determined by that classification alone. Those are the segments whose direct act category was determined in the annotation, and whose potential IFIDs we are seeking to find.

4.2.1.1 Explanatory Variables

**Serious Yes/No:** This variable indicates whether the direct speech act was annotated as being meant “seriously” or not. I expect the various categories to exhibit approximately equal amounts of “seriousness,” as jokes come in all flavors. As a result, I wouldn’t be surprised if this variable dropped out in the model selection phase.

**Indirect act choice:** This variable indicates the indirect act choice for the segment, including “None” as an option. In terms of variable-specific hypotheses, I expect expressive indirect acts to be the most common, seeing as acts which would be awkward to say directly are often relegated to the indirect level (and acts of opinion, which are under the umbrella of expressive, have a high potential for awkwardness), but in theory any type of direct act should be able to perform such an expressive indirect act. Therefore I expect that the associations for expressive indirect acts will be even across direct act category.

**Emotion word fraction:** This variable indicates the fraction of high emotion/sentiment words in the segment. Vosoughi & Roy (2016) hypothesize that expressives and directives may
have a higher than average number of high emotion/sentiment words (p.712). Zhang et al. (2011) also include this feature, but do not provide a specific reasoning for this inclusion (p.87).

The list of high emotion/sentiment words was pulled from the lexicon of Mohammad & Turney (2010), EmoLex. EmoLex includes 14,182 English words and their (binary) association with eight emotions and the two sentiment polarities -- positive and negative. For the purposes of this study, only words with either a positive or negative association in EmoLex were used: a total of 5,622 words.92

**Vulgar word fraction:** This variable indicates the fraction of “vulgar” words in the segment. Vosoughi & Roy hypothesize that expressives likely include a higher than average number of vulgar words (p.712). Zhang et al. also include this feature, but do not provide a specific reasoning for this inclusion (p.88).

What counts as “vulgar” is often a matter of personal opinion. Even the FCC does not provide an established list of profane words that are not allowed on network television, resorting to the platitude of “I know it when I see it.”93 In the absence of an official list of profanity, I resorted to the list that Vosoughi & Roy and Zhang et al. used. Both sets of authors pulled their vulgar word lists from the website noswearing.com, a site set up by programmers to combat the lack-of-official-profanity-list problem. They first compiled as many “bad words” as they could think of, and now accept user submissions which they add to the list as long as the word is in

---

92 The actual number of EmoLex terms with a positive or negative association was 5,636. However, as will be mentioned shortly, words that appeared in both the emotion word list and the vulgar word list were removed from the former.

wide (profane) usage and does not have a common usage that is a “real word” (i.e. “non-profane,” I presume).  

The noswearing.com list consists of 343 words (as of November 2018). A few (14) of these words were also in the emotion word list. Words in both were removed from the emotion word list.

**Modal fraction:** This variable indicates the fraction of modal auxiliaries in the segment, a feature included by Qadir & Riloff (2011). The authors do not indicate which speech act category they expect to find correlates with the pure presence of modals, but I expect any or all of the following correlations, based on the person of the subject: commissive (first person subject, e.g. “I will/might/may”), directive (second person subject, e.g. “you must/should/could”), or assertive (third person subject, e.g. “he will/would/can”).

The list used by Qadir & Riloff consisted of: may, must, shall, will, might, should, would, and could. To this I added can and ought.

**Initial modal yes/no:** This variable indicates whether the first token of the segment is a modal auxiliary, and is a feature included by Qadir & Riloff (2011). They hypothesize that it will correlate with interrogative and directive acts (p.751).

**Personal pronoun fractions:** Three separate personal pronoun variables were included, one for each person.

---

94 [https://www.noswearing.com/about.php](https://www.noswearing.com/about.php) While the noswearing.com list mostly consists of various permutations of the common four-letter words, several entries are questionable and do not actually follow all of their rules. For example, “gay” and “lesbian” were included in the list. This means that either the list-makers accidentally violated their rule of making sure the common usage of an entry isn’t a “real word,” or they believe the common usages of “gay” and “lesbian” are themselves profane. These particular examples did not matter for my corpus, as neither “gay” nor “lesbian” occurred in it. But they indicate a caveat that the noswearing.com list is not without the same “I know it when I see it” vulnerability to bias as is the FCC.
Qadir & Riloff include these features, but do not indicate which speech act categories they expect to correlate with them (p.4). I included them on the conjecture that **First** will possibly correlate with commissive, and **Second** with directive. I have no specific hypothesis regarding **Third**, but included it for the sake of completeness.

**Wh-word fraction**: This variable indicates the fraction of wh-words in the segment, a feature included by Qadir & Riloff (p.751). They do not state the reason for inclusion, but we may hypothesize that higher wh-word fractions will be seen in interrogatives, and in expressives (e.g. “What a night!”). The list used by Qadir & Riloff consisted of: *who, what, where, when, why, which, and how.* To this I added *whom*.

**Speech act verb fractions**: 39 separate speech act verb variables were included, one for each word on Wierzbicka’s (1987) speech act verb list that appeared at least once (as a verb) in the corpus (pp.33-35). The hypothesis here is that the different verbs will correlate with different direct act categories, depending on which category is the most appropriate super-category for the act denoted by the verb.⁹⁶

Vosoughi & Roy (p.712) and Qadir & Riloff (p.751) both include these same verbs in some capacity. The latter simply includes one feature that checks for the presence of ANY of the Wierzbicka verbs, while Vosoughi & Roy do as I have done (separate features for separate

---

⁹⁵ Technically, the list they provide in Qadir & Riloff (2011) omits *why* and mentions *what* twice. I have assumed this was a typo.

⁹⁶ Some examples on the list are joke, name, ask, and tell.
verbs). To that end, I have used their methodology, including the tools they used. After using Owoputi et al. (2013)’s Twitter part-of-speech tagger to tokenize and POS-tag the tweets,97 I used the Porter Stemming algorithm to stem all tokens tagged as verbs.98 I then stemmed the 229 Wierzbicka verbs before cross-checking them with the stemmed verbs from each segment. Count features for Wierzbicka verbs that did not appear in the corpus at all were left out.

**Segment length:** This variable indicates the length of the segment, in number of tokens. Because of interjections, I expect the expressive category to positively correlate with short segment length. I also expect the assertive category to positively correlate with short segment length, as verbless fragments are often assumed to be copular with a contextually-salient or expletive subject (e.g. “Late today” = “I am late today,” or “Raining” = “It is raining”). Of course, if the fragment is an adjective expressing a psychological state (e.g. “Happy” = “I am happy”), the contextually-salient subject and copula would result in an expressive act, another reason for expressive acts to correlate with short segments. Even if the adjective is not related to a psychological state, it may be communicating the speaker’s emotions about another subject (e.g. “Disgusting” = “[contextually-salient subject] is disgusting”), which would still be expressive.

**Part-of-speech fractions:** 20 separate part-of-speech variables were included, one for each type of tag output by Owoputi et al.’s tagger.99 Vosoughi & Roy did likewise, but only for the adjective, interjection, and verb tags. They hypothesized that a higher fraction for the first

---

97 Available at [http://www.cs.cmu.edu/~ark/TweetNLP/#parser_paper](http://www.cs.cmu.edu/~ark/TweetNLP/#parser_paper).
98 Specifically, I used the Porter Stemming algorithm coded for R by Mohit Makkar, available at [https://tartarus.org/martin/PorterStemmer/](https://tartarus.org/martin/PorterStemmer/).
99 Features using the @-tag and hashtag counts made by the tagger were left out, as I had already included other @-tag and hashtag features. Likewise for the personal pronoun counts.
two and a lower count for verb tags would all correlate with expressive acts (p. 713). For the same reasons that I expect short segments to correlate with assertive and expressive (above), I expect verbless segments to correlate with assertive and expressive. I would also add a conjecture that higher adverb fractions will correlate with expressive acts, as well. While I do not have specific hypotheses for the various other part-of-speech tags, they were simple enough to include, and may end up exhibiting associations with certain direct act categories.

**Tweet length:** This variable indicates the length of the tweet that a given segment is a part of, in number of segments, including incorporated segments. I expect assertive segments to appear in longer tweets, as acts of explanation are generally assertive. An explanation is a natural follow-up to really any other type of act (e.g. explaining why you have expressed something, explaining why you are asking a certain question, explaining why you are committing yourself to an action).

**Segment position:** This variable indicates the position of the segment within its respective tweet. The options are: entire (1 out of 1 segment), beginning (1 out of >1 segments), end (n out of n segments, where n ≠ 1), and middle (everything else).

Qadir & Riloff include this feature, hypothesizing (for their online forum data) that expressives often occur at the beginning or end of a post, and directives tend to occur near the end (p.752). Due to the explanatory capability of assertives just mentioned, I expect to see them occur often in the middle or end of a post (after whatever segment they are meant to explain).

**Original/reply status:** This variable indicates whether a segment is part of an original post or a reply post. I expect that directive acts may be common in reply posts, seeing as a
common reason to reply to someone is that you have advice regarding a concern that they have voiced.

**Standard capitalization Yes/No:** This variable indicates whether a given segment exhibits capitalization consistent with standard writing conventions. For the purposes of this feature, “standard” was defined as: 1) the first letter of every proper noun is capitalized, 2) any word at the beginning of a tweet (not necessarily at the beginning of a segment) is capitalized, and 3) any word coming after a period, exclamation point, or question mark is capitalized. A segment that violated any of these rules, or contained any word with all letters capitalized, was coded as exhibiting non-standard capitalization.

I expect non-standard capitalization to correlate with expressive acts, seeing as “all caps” capitalization is sometimes used in modern media to represent focus stress or “volume.”

**Punctuation:** This variable is categorical, giving the final character of the segment out of period, comma, exclamation point, question mark, colon, semi-colon, or “other” (anything else). Vosoughi & Roy (p.713) and Zhang et al. (p.88) both include final character variables, but only as two binary variables checking for the presence of exclamation point and question mark, which they of course hypothesize to correlate with expressives and interrogatives, respectively, based on standard writing conventions.

**Hashtag features:** Three variables were included that related to the presence and/or distribution of hashtags:

- **Hashtag in tweet Yes/No:** This variable indicates whether the tweet of which a given segment is a part contains a hashtag anywhere (“Yes”) or not (“No”).
• **Incorporated hashtag Yes/No:** This variable indicates whether the segment contains an incorporated hashtag.

• **Non-incorporated hashtag adjacent Yes/No:** This variable indicates whether a non-incorporated hashtag is adjacent to (either before or after) the segment.

Vosoughi & Roy (p.713) and Zhang et al. (p.88) both include features to check for the presence and distribution of “Twitter-specific” characters, including #, though neither provide very clear hypotheses about the exact connection between these variables and speech act category.

**@-tag features:** Three variables were included that related to the presence and/or distribution of @-tags. They are identical to the hashtag variables, but for @-tags: **@-tag in tweet Yes/No, Incorporated @-tag Yes/No,** and **Non-incorporated @-tag adjacent Yes/No.** The “Twitter-specific” character features included by Vosoughi & Roy and Zhang et al. also assessed the presence and distribution of @. Seeing as non-incorporated @-tags were classified as direct address segments during the segmentation phase, I expect the **Non-incorporated @-tag adjacent** variable to correlate with directives and/or interrogatives (e.g. “@John, come pick me up,” or “@John, what’s up?”).

**Emoji⁹⁰ features:** Three variables were included that related to the presence and/or distribution of emoji. They were similar to, but not entirely identical to, the variables for hashtags and @-tags:

---

⁹⁰ “Emoji” here is used as a cover term for both emoji and emoticons.
• **Emoji in tweet fraction**: This variable gives the total number of emoji present in the encompassing tweet over the total number of tokens in the encompassing tweet.

• **Incorporated emoji Yes/No**: This variable indicates whether or not a given segment includes an incorporated (lexical) emoji.

• **Tone-type emoji adjacent Yes/No**: This variable indicates whether the segment is adjacent to (before or after) a tone-type emoji.\(^{101}\)

Vosoughi & Roy (p.712) and Zhang et al. (p.88) both include one binary feature that checks for the presence of any member of a long list of emoticons (but no emoji). V&R include the hypothesis that the presence of emoticons will correlate with expressive acts. For the variables presented here, I would hypothesize perhaps that higher fractions of emoji might correlate with expressives, as might adjacency to a tone-type emoji.

**Bigram features**: Approximately 30 additional features were included that each indicated the presence or absence of a certain bigram. The bigrams consisted of the two word sequences from each direct act category with the highest tf-idf scores in the training portion of the corpus (the top 5 for each category were included, in most cases).

Vosoughi & Roy, Zhang et al., Qadir & Riloff, Twitchell & Nunamaker (2004), Carvalho & Cohen (2006), and Jurafsky et al. (1997) all included similar n-gram features in some way, usually pulled from the data itself (as done here), but sometimes pulled from another corpus. The hypothesis here is of course that if a bigram from the training portion correlates with the

\(^{101}\) Tone-type emoji are non-incorporated and do not appear to stand in for a specific word or phrase. These are in contrast to lexical emoji, which appear to stand in for a word or phrase (though they do not have to be incorporated).
same speech act category in the test portion, it is a potential IFID for that category, at least in the context of Twitter. The dearth of exercitive acts in the corpus means that the bigrams for that category are basically taken from one of two tweets, and are therefore quite unlikely to actually be meaningful for categorizing unseen data, but were included for the sake of completeness.

**Interaction terms**: The following interactions terms were included in the logistic regression analysis (random forest analysis does not necessitate declaration of interaction terms):

- For hashtags, @-tags, and emoji: two terms for each set of features. In each case, the “[element] in tweet” feature is crossed once with the “incorporated” feature, and once with the “adjacent” feature, seeing as the presence of an incorporated or adjacent element automatically increases the “[element] in tweet” fraction variable for that segment
- One term for the interaction of tweet length and segment position, seeing as a segment with a segment position of “entire” will automatically have a tweet length of one segment
- One term for the interaction of serious yes/no and indirect act choice

The last term was included because of Clark’s (1979) hypothesis that seriousness correlates with the presence or absence of an indirect act.\(^{102}\)

---

\(^{102}\) This potential interaction will also be discussed further in section 4.2.2.
4.2.1.1.2 Model Selection

All together, the analysis tested 122 variables. These were too many variables for them all to be meaningful, especially for the amount of data in the corpus, so model selection was a crucial aspect of the analysis.

For the logistic regression, stepwise feature dropping was used to whittle the 122 features down to only those features whose absence caused a significant change in the model. A change was considered significant if it resulted in a p-value of <0.05 in the analysis of variance (ANOVA) between the two models. A significant change meant that the feature was kept, while an insignificant change meant that the feature remained dropped.

The random forest analysis required a different model selection method. The random forest output includes the mean decrease in accuracy (MDA) of the model if the feature were to be left out. The lower the value for MDA, the less important the feature. The random forest output also includes the mean decrease in Gini coefficient (MDG) for each feature. A higher mean decrease in Gini coefficient means that a feature contributes more to the purity of the tree branches. That is, a feature with a higher MDG used as a decision node splits the data into more homogenous groupings.

To arrive at a list of only the important variables in the random forest model, both MDA and MDG were used. First, any feature with an MDA value of zero (or less) was eliminated. After that, features were eliminated in a stepwise fashion by establishing a threshold MDG value. Features with MDG values below the threshold were eliminated. If this elimination led to an increase in model accuracy, those features remained eliminated, and the threshold was raised.

103 The exact number of explanatory variables changes slightly for the different data subsets, due to different numbers of bigrams used, as evidenced by table 4.17.
The threshold was gradually raised until the increase in threshold led to a decrease in model accuracy.

One final step of model selection was that four separate analyses were performed for both the logistic regression and random forest techniques, one for each of four different subsets of the corpus data. These four subsets are a result of crossing two meta-variables: 1) whether the majority-annotation version (one row per segment) or the full version (three rows per segment, one for each annotator) of the corpus was used, and 2) whether attention and quality control questions were included or excluded. If the majority-annotation version of the corpus yields better results than does the full version, this may suggest that, on the whole, the differing annotations represent more erroneous noise than non-erroneous ambiguity. Meanwhile, the inclusion of the attention and quality control questions should yield better results than their exclusion, as they simply provide more data. But if their exclusion yields better results than does their inclusion, this may suggest that the “correct” answers to the attention and quality control questions may have been poorly chosen.

The training/test split of the data was kept consistent across the four different subsets of data. A random 90% of the majority-annotation version, with AQCs included, was selected first. The unique segment IDs from that random 90% were then used to select the 90% of the full version that was used for training as well. Finally, AQCs were filtered out of both to get the data subsets without AQCs. Likewise for the test portions of the data.

The same set of variables was used for each of the four data subsets, with the exception of the bigrams (and any associated interaction terms). Tf-idf was calculated for each training portion of the four subsets of data, and usually the 5 bigrams in each category with the highest
tf-idf were used. In the event of a tie, more or less than 5 bigrams were used. If including the tie would keep the number of bigrams closer to 5, the tie was included. If excluding the tie would do that, the tie was excluded.

Table 4.17 provides the lists of bigrams used in each category for each of the four subsets.

**Table 4.17. Bigrams for direct act analysis.**

<table>
<thead>
<tr>
<th>Data Subset</th>
<th>Assertive</th>
<th>Directive</th>
<th>Commissive</th>
<th>Interrogative</th>
<th>Expressive</th>
<th>Exercitive</th>
</tr>
</thead>
<tbody>
<tr>
<td>MajAnno, No AQCs (#1)</td>
<td>is a; i just; in a; is the; i have</td>
<td>come on; be safe; gotta stop; let’s go; you must</td>
<td>i will; i’m going; im just; i’m gonna; watch it</td>
<td>is it; am i; fuck is; it just; just me</td>
<td>i hate; i love; i’m so; thank you; i miss</td>
<td>$5.95 chef; be to; chef special; glory be; his holy</td>
</tr>
<tr>
<td>Full, No AQCs (#2)</td>
<td>the most; the same; is the; i have; one of</td>
<td>be safe; gotta stop; get the; out of</td>
<td>i’m going; i’m gonna; i will; watch it; going to</td>
<td>is it; am i; fuck is; it just; just me; or is</td>
<td>i love; i miss; thank you; i’m so; i hate</td>
<td>$5.95 chef; chef special</td>
</tr>
<tr>
<td>MajAnno, AQCs (#3)</td>
<td>is a; in a; the same; i have; i just</td>
<td>come on; be safe; give me; gotta stop; just give</td>
<td>i will; i’m going; i’m gonna; im just; watch it</td>
<td>is it; are you; can i; how can; did you</td>
<td>i hate; i love; i miss; thank you; i’m so; i hate</td>
<td>$5.95 chef; be to; chef special; glory be; his holy</td>
</tr>
<tr>
<td>Full, AQCs (#4)</td>
<td>the same; the most; i have; been a; it’s been</td>
<td>be safe; give me; gotta stop; just give; just stop; lmk if; my mama</td>
<td>i will; i’m gonna; i’m going; watch it; gonna go</td>
<td>is it; are you; can i; how can; did you; is there; it just</td>
<td>i love; i miss; i’m so; thank you; so excited</td>
<td>$5.95 chef; chef special</td>
</tr>
</tbody>
</table>
4.2.1.1.3 Regression Analyses

The multinomial logistic regression function `multinom` from the R package `nnet` was used for these analyses. For all four data subsets, after model selection, no variables were marked significant by the final model. But many of the variables appeared in the final model of more than one regression, as summarized in table 4.18. This means that, for more than one of the data subsets, the elimination of each of these variables caused a significant change in the model.

Table 4.18. Variables that appeared in the final model of more than one regression.

<table>
<thead>
<tr>
<th>Number of final models</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 of 4</td>
<td>Modal fraction, adjective fraction, vulgar word fraction, preposition fraction, “be safe” (directive bigram), “i will” (commissive bigram), “i’m gonna” (commissive bigram), “i hate” (expressive bigram)</td>
</tr>
<tr>
<td>2 of 4</td>
<td>“Give” fraction, verb fraction, common noun fraction, adverb fraction, original/reply status, coordinate conjunction fraction, emoji in tweet fraction, “i’m going” (commissive bigram), “am i” (interrogative bigram), “is it” (interrogative bigram), “thank you” (expressive bigram), “$5.95 chef” (exercitive bigram)</td>
</tr>
</tbody>
</table>

Although the point of hypothesis #1 was not to build a classifier but rather to uncover potential IFIDs (illocutionary force indicating devices), how well the model predicts the direct act category of unseen data may give us an idea of whether (or to what extent) we have succeeded on that front. The overall classification accuracy of each of the final models on the
test portion of its data subset is provided in table 4.19. This number indicates what percentage of segments were classified accurately by the model, across all categories. Below it, table 4.20 gives the category-specific accuracies of each model.

**Table 4.19. Classification accuracy of final regression models for test portions of data.**

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MajAnno, No AQC (#1)</td>
<td>77.3%</td>
</tr>
<tr>
<td>Full, No AQC (#2)</td>
<td>72.8%</td>
</tr>
<tr>
<td>MajAnno, AQC (#3)</td>
<td>77.2%</td>
</tr>
<tr>
<td>Full, AQC (#4)</td>
<td>75.1%</td>
</tr>
</tbody>
</table>

**Table 4.20. Category-specific classification accuracy of final regression models for test portions of data.**

<table>
<thead>
<tr>
<th>Data subset</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>I</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>96.6%</td>
<td>0%</td>
<td>22.2%</td>
<td>N/A</td>
<td>83.3%</td>
<td>21.7%</td>
</tr>
<tr>
<td>#2</td>
<td>96.4%</td>
<td>4.3%</td>
<td>14.7%</td>
<td>0%</td>
<td>52.6%</td>
<td>17.8%</td>
</tr>
<tr>
<td>#3</td>
<td>95.8%</td>
<td>0%</td>
<td>6.7%</td>
<td>N/A</td>
<td>70.0%</td>
<td>40.0%</td>
</tr>
<tr>
<td>#4</td>
<td>94.9%</td>
<td>8.7%</td>
<td>9.6%</td>
<td>0%</td>
<td>75.4%</td>
<td>34.2%</td>
</tr>
</tbody>
</table>

If we look at individual categories, all four models performed best on the assertive category, each classifying ~95-96% of assertive segments correctly. This is an unsurprising result, as the majority of segments in both the training and test data are assertive.
In all four cases, the model performed second-best on the interrogative category, classifying between 53% and 83% of interrogative segments correctly.

In three of four cases, the model performed third-best on the expressive category, classifying between 17% and 40% of expressive segments correctly. In the fourth case, the expressive category was surpassed by the directive category, which was otherwise in fourth place. Among the four models, between 7% and 22% of directives were classified accurately.

In terms of which model performed best overall, the model trained on data subset #1 (majority-annotation, no AQC)s) performed the best on the unseen data by the metric of overall classification accuracy, and also by the metric of category-specific accuracies, as it produced the highest category-specific accuracies for three of the categories: assertive, directive, and interrogative.

4.2.1.1.4 Random Forest Analyses

The function randomForest from the R package randomForest was used for these analyses. As with the regression analyses, the same set of variables was used for each of the four random forest analyses, with the exception of the bigrams. For all four data subsets, after model selection, no variables were marked significant by the final model. But several variables appeared with high mean decrease in accuracy values in multiple final models. Table 4.21 provides the variables that appeared in the top 5 variables (by MDA rank) for more than one final model.
Table 4.21. Variables that appeared in the top 5 for more than one random forest final model.

<table>
<thead>
<tr>
<th>Number of final models</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 of 4</td>
<td>Punctuation</td>
</tr>
<tr>
<td>3 of 4</td>
<td>(none)</td>
</tr>
<tr>
<td>2 of 4</td>
<td>Interjection fraction, tweet length, wh-word fraction, verb fraction, “i love” (expressive bigram), “i hate” (expressive bigram)</td>
</tr>
</tbody>
</table>

The overall classification accuracy of each of the final models on the test portion of its data subset is provided in table 4.22. Below it, table 4.23 gives the category-specific accuracies of each model.

Table 4.22. Classification accuracy of final random forest models for test portions of data.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Classification accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MajAnno, No AQC's (#1)</td>
<td>77.9%</td>
</tr>
<tr>
<td>Full, No AQC's (#2)</td>
<td>70.1%</td>
</tr>
<tr>
<td>MajAnno, AQC's (#3)</td>
<td>74.9%</td>
</tr>
<tr>
<td>Full, AQC's (#4)</td>
<td>70.5%</td>
</tr>
</tbody>
</table>
Table 4.23. Category-specific classification accuracy of final random forest models for test portions of data.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>A</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>I</th>
<th>X</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>100%</td>
<td>0%</td>
<td>11.1%</td>
<td>N/A</td>
<td>0%</td>
<td>34.8%</td>
</tr>
<tr>
<td>#2</td>
<td>99.1%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>5.3%</td>
<td>9.6%</td>
</tr>
<tr>
<td>#3</td>
<td>100%</td>
<td>0%</td>
<td>13.3%</td>
<td>N/A</td>
<td>10.0%</td>
<td>40.0%</td>
</tr>
<tr>
<td>#4</td>
<td>98.5%</td>
<td>0%</td>
<td>5.8%</td>
<td>0%</td>
<td>31.1%</td>
<td>18.9%</td>
</tr>
</tbody>
</table>

If we look at individual categories, all four models performed best on the assertive category, each classifying ~98-100% of assertive segments correctly. This is an unsurprising result, as the majority of segments in both the training and test data are assertive.

The order of which of the other categories were best predicted by the model varies quite a bit from subset to subset. Second place goes to the expressive category in three out of four cases, and third place goes to directive twice, but interrogative varies from second to last (tied for fourth).

In terms of which model performed best overall, the model trained on data subset #1 (majority-annotation, no AQCs) performed the best on the unseen data by the metric of overall classification accuracy, but not by the metric of category-specific accuracies. The model trained on data subset #3 (majority-annotation, AQCs included) performed the best by that metric, with the highest category-specific accuracies for three categories: assertive (tied), directive, and expressive.
4.2.1.1.5 Summary of Repeat Variables

The following list provides the variables that appeared in two or more regressions or two or more top random forest MDA rankings. The variables are presented according to the number of regressions and top random forest MDA rankings they appear in, starting with the most frequent. Ties are listed under the same number. Association plots for each variable (except the bigrams) will be given in the next section, as we address any variable-specific hypotheses concerning the variable.

1) **Punctuation** (4/4 regressions, 4/4 MDA rankings)

2) **Wh-word fraction** (4/4 regressions, 2/4 MDA rankings)
   - “i love” expressive bigram (4/4 regressions, 2/4 MDA rankings)

3) “i hate” expressive bigram (3/4 regressions, 2/4 MDA rankings)

4) **Verb fraction** (2/4 regressions, 2/4 MDA rankings)

5) **Indirect act choice** (4/4 regressions, 0/4 MDA rankings)
   - Serious yes/no (4/4 regressions, 0/4 MDA rankings)
   - “Ask” verb fraction (4/4 regressions, 0/4 MDA rankings)
   - “Tell” verb fraction (4/4 regressions, 0/4 MDA rankings)
   - **Proper noun fraction** (4/4 regressions, 0/4 MDA rankings)
   - **Numeral fraction** (4/4 regressions, 0/4 MDA rankings)
   - **Nominal verb contraction fraction** (4/4 regressions, 0/4 MDA rankings)
   - **Determiner fraction** (4/4 regressions, 0/4 MDA rankings)
   - **Segment length** (4/4 regressions, 0/4 MDA rankings)
   - “gotta stop” directive bigram (4/4 regressions, 0/4 MDA rankings)
“watch it” commissive bigram (4/4 regressions, 0/4 MDA rankings)

“i’m so” expressive bigram (4/4 regressions, 0/4 MDA rankings)

“i miss” expressive bigram (4/4 regressions, 0/4 MDA rankings)

6) Modal fraction (3/4 regressions, 0/4 MDA rankings)

Adjective fraction (3/4 regressions, 0/4 MDA rankings)

Vulgar word fraction (3/4 regressions, 0/4 MDA rankings)

Preposition fraction (3/4 regressions, 0/4 MDA rankings)

“be safe” directive bigram (3/4 regressions, 0/4 MDA rankings)

“i will” commissive bigram (3/4 regressions, 0/4 MDA rankings)

“i’m gonna” commissive bigram (3/4 regressions, 0/4 MDA rankings)

7) “Give” verb fraction (2/4 regressions, 0/4 MDA rankings)

Common noun fraction (2/4 regressions, 0/4 MDA rankings)

Adverb fraction (2/4 regressions, 0/4 MDA rankings)

Original/reply status (2/4 regressions, 0/4 MDA rankings)

Coordinate conjunction fraction (2/4 regressions, 0/4 MDA rankings)

Emoji in tweet fraction (2/4 regressions, 0/4 MDA rankings)

“i’m going” commissive bigram (2/4 regressions, 0/4 MDA rankings)

“am i” interrogative bigram (2/4 regressions, 0/4 MDA rankings)

“is it” interrogative bigram (2/4 regressions, 0/4 MDA rankings)

“thank you” expressive bigram (2/4 regressions, 0/4 MDA rankings)

“$5.95 chef” exercitive bigram (2/4 regressions, 0/4 MDA rankings)

Interjection fraction (0/4 regressions, 2/4 MDA rankings)
**Tweet length** (0/4 regressions, 2/4 MDA rankings)

In addition to these individual variables, one interaction term survived model selection for all four regressions. This was the interaction of *Tweet length* and *Segment position*. A significant interaction between these variables is not at all unexpected (hence including the interaction term). The four segment position categories were: entire (1/1 segment), beginning, middle, and end. This means that a segment with a tweet length of 1 would only ever have a segment position of “entire,” and vice versa (i.e. no other tweet lengths would yield segment position classifications of “entire”). Also, a tweet with a length of 2 would only ever have “beginning” and “end” positions. These eventualities can be seen in figure 4.4, the association plot for **Tweet length** and **Segment position**.

![Association plot for Tweet length and Segment position, data subset #1.](image)

*Figure 4.4. Association for Tweet length and Segment position, data subset #1.*
4.2.1.2 Discussion

At the most basic level, several overt variables demonstrated significant effects on the models developed in the previous sections. The hypothesis that “Certain overt elements available to Twitter users function as direct illocutionary force indicating devices” is therefore broadly supported. However, this not does mean that there are strong IFIDs available for each direct act category. Some categories appear to have more, and/or more reliable, IFIDs than do others, as evidenced by the confusion matrix data in tables 4.20 and 4.23. Nor is it necessarily the case that the associations found for these variables are really IFIDs in the original sense intended by Searle & Vanderveken. Such issues will be discussed in section 4.2.1.2.2.

Additionally, the two inferred variables included in the analysis (indirect act choice and serious yes/no) also demonstrated significant contributions to the models, particularly the regression models. This suggests that, while overt IFIDs exist, inferred information may also contribute to direct act classification.

4.2.1.2.1 Variable-specific Hypotheses and Results

Focusing in on the variables listed in section 4.2.1.1.5, the specific hypotheses associated with those variables mentioned in section 4.2.1.1.1 were supported in some cases, and not supported in others. Association plots for each of those variables are given below. For each variable, I will address whether any variable-specific hypotheses were supported or refuted, as well as whether the associations observed for that variable would be expected, with regard to reasonable deduction and known conventions. As will become evident, many of the variables discussed here are likely not IFIDs in the sense that their presence directly signals the force of the segment, but they may connect with other variables or concepts that are more likely to have a
role in speech act classification. In the next section, I will step back and assess what these results suggest, all together, about IFIDs and speech act interpretation/classification.

The training portion of data subset #1 (majority-annotation version, AQC’s excluded) was used for these plots, because the model trained on it had the highest overall classification accuracy for both the regressions and random forest analyses, and the best category-specific accuracies for the regressions (though the model trained on subset #3 had the best category-specific accuracies for the random forests). The significance of these facts will be discussed at the end of section 4.2.1.2.2.

In this section, I will use shorthand to summarize the associations observed for each variable listed in section 4.2.1.1.5. A positive association will be noted with an upwards arrow ↑, while a negative association will be noted with a downwards arrow ↓.

In association plots, neither height of the association bar nor width is the more important feature. The height indicates the strength of the association, while the width indicates how many data points the association is based on. Therefore, one approximation of the significance of an association is the area of the bar. For each of the plots below, the bars with the highest approximate areas will be the ones addressed, in addition to any overall trends.

In several cases, the results have been batched for easier viewing. That is, instead of presenting the results of, say, 30 different “Noun Fraction” values ranging between 0 and 1, those values have been grouped into 6 batches of values. Association plots are not given for bigram variables, as their primary associations are inherent to their being a bigram for a given category

---

104 In all cases where the values range between 0 and 1, the batches are: 0; 0 < x ≤ 0.2; 0.2 < x ≤ 0.4; 0.4 < x ≤ 0.6; 0.6 < x ≤ 0.8; and 0.8 < x ≤ 1. For segment length, the batches increase by 4 up to 32 (i.e. x ≤ 4, 4 < x ≤ 8, … 28 < x ≤ 32).
(i.e. strong positive association with that category only). The significance of the surviving bigrams will be discussed, as a whole, near the end of this section.

**Punctuation:** All of the variable-specific hypotheses for punctuation are supported by figure 4.5, and some additional associations can be seen as well: *question mark* ↑ *interrogative,*  
*exclamation point* ↑ *expressive,*  
*period* ↑ *assertive,*  
*exclamation point* ↑ *directive,*  
*comma* ↑ *assertive*.

![Punctuation](image)

**Figure 4.5. Association for Direct act category and Punctuation, data subset #1.**

The *question mark* ↑ *interrogative* and *exclamation point* ↑ *expressive* associations were hypothesized and expected based on standard writing conventions, which appear to hold (at least some of the time) on Twitter.

Though the *period* ↑ *assertive,* *exclamation point* ↑ *directive,* and *comma* ↑ *assertive* associations were not originally hypothesized, they are not unexpected given the standard writing
conventions that the variable-specific hypotheses were based on. Periods can come at the end of any direct act segment except interrogative, exclamation points can be used to represent stress or volume lent to any type of statement (i.e. any utterance could theoretically be exclaimed), and commas can represent a pause between two clauses (of, theoretically, any speech act category), or be necessary in the case of multiple coordinated elements.

However, though not entirely unexpected, these last three results are still interesting with regard to what else they can tell us about speech act interpretation and general communication on Twitter. First of all, the period actually has a marked status in computer-mediated communication.\textsuperscript{105} I would presume, therefore, that it is showing up here because of tweets with multiple sentences that require punctuation to avoid confusion. If we take a look at figure 4.6, the association plot for segment position (a variable that did not prove to be significant by itself), we can see that indeed, although assertive segments positively associate with the appearance of a period, they are negatively associated with a position at the end of a tweet. This also likely explains the association between assertive and the presence of a comma.

\textsuperscript{105} Gunraj et al. (2016) found that text messages that ended with a period were considered less sincere compared to the same message without the period. This difference was not observed for handwritten (i.e. non-CMC) messages (p.1067). Crair (2013) likewise observes that utterance-final periods in text and instant messages have gained an unhappy or even aggressive interpretation.
The follow-up question is then why assertive segments tend not to come at the end of tweets, relative to the other categories. This may simply be a convention of communication on Twitter, that tweets tend to end with something other than an assertive act.

Second, the fact that a segment could end with an exclamation point and be classified as something other than expressive suggests that the expressive quality signalled by the punctuation does not necessarily override the force of an un-exclaimed version of the utterance.\textsuperscript{106} Additionally, it is interesting that, in the present corpus, this appears to have happened the most frequently with the directive category (hence the association). It is not immediately evident why such should be the case.

\textsuperscript{106} This is in contrast to, say, the question mark, which only shows positive associations with interrogative segments.
**Wh-word fraction:** The variable-specific hypotheses for wh-word fraction are supported by figure 4.7: non-zero wh-word fractions ↑ interrogative; low, non-zero wh-word fractions ↑ expressive; absence of wh-words ↑ assertive; absence of wh-words ↑ commissive; absence of wh-words ↑ directive.

![Figure 4.7. Association for Direct act category and Wh-word fraction (batched), data subset #1.](image)

These results are quite expected, seeing as wh-words are necessary for asking wh-questions, and common in certain exclamative constructions (e.g. “What a night!”).

Wh-words also appear in relative clauses. Restrictive relatives were designated as backgrounded (and therefore unavailable to perform a direct act) during the segmentation, and non-restrictive relatives were designated as possibly backgrounded (to be judged for backgroundedness in the annotation). The associations seen here might suggest that
non-restrictive relatives were either usually judged to perform interrogative or expressive acts (the latter seeming more likely), or judged as backgrounded and unable to perform a direct act.

**Verb fraction**: The variable-specific hypothesis that lower verb fractions would be seen for expressives is supported by figure 4.8, but not to the exclusion of other expressive associations. That is, *absence of verbs* ↑ expressive, but also *verb fraction between 0.4 and 0.8* ↑ expressive. The variable-specific hypothesis for assertives is supported: *verb fraction of 0 up to 0.2* ↑ assertive. Other associations observed were: *absence of verbs* ↓ commissive; *absence of verbs* ↓ directive; *low, non-zero verb fraction* ↑ interrogative.

![Figure 4.8. Association for Direct act category and Verb fraction (batched), data subset #1.](image)

Though not necessarily expected, the split in the expressive associations is a reasonable result. Exclamatives are often fragments without any verbs (e.g. “What a night!”), as are
interjections (e.g. “Holy shit!”), but not all expressive statements are exclamatives or
interjections.

For assertive, as stated in section 4.2.1.1.1, verbless fragments are often assumed to be
copular with a contextually-salient or expletive subject (e.g. “Late today” = “I am late today,” or
“Raining” = “It is raining”). This would explain the association with a verb fraction of 0, while
non-fragment assertives explain the non-zero verb fraction association.

Finally, I would draw attention to the fact that while directive acts are not the only ones to
have only non-zero positive associations, they are the only category to exhibit a positive
association with a verb fraction of 1. This is another reasonable result, as a verb is the only
necessary element for an imperative (e.g. “Come”).

**Indirect act choice:** The variable-specific hypothesis, that expressive indirect acts would
be relatively even across the different direct act categories, is refuted by figure 4.9 (though the
expectation that expressive would be the most common category of indirect act is supported).

*Expressive indirect act ↑ assertive direct act and expressive indirect act ↑ interrogative direct act,*
but *expressive indirect act ↓ commissive direct act, expressive indirect act ↓ directive direct act,*
and *expressive indirect act ↓ expressive direct act.* Other associations noted were: *no indirect act
↑ commissive direct act, no indirect act ↑ expressive direct act, and directive indirect act ↑
directive direct act.*
Some of the results for this variable are expected, which clear potential explanations, while others are not. As stated above, I expected expressive to be the most common category of indirect act and that was indeed the case. In fact, the associations of each direct act category with the expressive indirect category are the reverse of their associations with the “none” indirect category in all cases except directive direct (and exercitive direct). So, for most categories, it was either the case that the segment performed an expressive indirect act, or no indirect act at all.

We might be tempted to explain the patterns in figure 4.9 by looking to the next variable, serious yes/no. As stated in section 1.1.3, Clark (1979) proposed that direct acts were usually not meant seriously in the presence of an indirect act. If we found that the direct act categories less likely to perform an indirect act were the categories more likely to be meant seriously, this might
explain why. However, figure 4.10 does not support such a conclusion. Whether Clark’s findings are reproduced by the present study will be further discussed in section 4.2.2.

Another potential explanation might be found in Potts (2007a). Potts details how expressive words may convey an expressive meaning that is both simultaneous and parallel to the descriptive meaning of a given utterance. Whether such an expressive meaning should be analyzed as direct, indirect, or neither from a theoretical point of view is an issue of its own, but in the absence of specific guidance, annotators may have resorted to annotating this expressive meaning as an indirect act, contributing to the high count of expressive indirect acts, particularly for certain direct act categories. This issue will be revisited at the end of this section, as it ties in with several other variables.

**Serious yes/no:** The variable-specific hypothesis for serious yes/no, that the different categories would be equally likely to be serious and the variable would drop out in model selection, is refuted on both counts. The variable did not drop out, and clear differences are seen in the associations in figure 4.10: serious $\uparrow$ assertive; serious $\downarrow$ commissive; serious $\downarrow$ directive; serious $\downarrow$ interrogative; serious $\downarrow$ expressive.
Figure 4.10. Association for Direct act category and Serious yes/no, data subset #1.

This is an interesting result. For the commissive, directive, and interrogative categories, the fact that they were less often interpreted as serious may partially be a result of their defining qualities. Under Alston’s (2000) definition of illocutionary speech act, the serious and felicitous performance of a given act requires that the speaker take responsibility for certain conditions being satisfied. He details those conditions for each category (see section 1.1.2.3). For directive (which includes interrogative, in his taxonomy) and commissive, the speaker must take responsibility for multiple conditions, at least one of which involves the hearer (e.g. laying a responsibility on them, or ascribing to them a desire that the speaker perform a certain act).

But when a speaker is tweeting from a public account on a social media platform, there is not just one hearer. It is reasonable that taking responsibility for conditions that involve an indeterminate number of hearers may be judged as unlikely (i.e. that the act was not meant
Even if the speaker uses an @-tag for their intended hearer, the utterance is still being performed for a larger audience. To judge if the act was meant seriously, the audience would have to know certain things about the speaker, the hearer, and their relationship. In the absence of this knowledge, the audience cannot guarantee that the speaker is holding themself responsible for the necessary conditions. It seems likely that, in such a situation, the audience would default to assuming the act is not serious.

Meanwhile, for assertive and expressive under Alston’s definitions, the speaker only has to take responsibility for one condition which doesn’t even involve the hearer: that a certain proposition is true (assertive), or that the speaker has a certain psychological state (expressive). The fact that these two categories differ in terms of seriousness, specifically that expressives were more likely to be judged as non-serious, must have a different explanation. It is possible that this is simply a norm of communication on Twitter, that assertive acts are usually considered serious, while expressive acts are not.

“Ask” verb fraction: The variable-specific hypothesis for all of the “speech act verbs” from Wierzbicka (1987) was that each verb would positively correlate with whichever speech act category was the super-category for the act denoted by the verb. In the case of “ask,” the hypothesis therefore was that non-zero ask fraction ↑ interrogative. Although this association is observed in figure 4.11 (and therefore the hypothesis is supported), that was not the only positive association observed. There was also non-zero ask fraction ↑ directive.

107 If the speaker really wanted to only speak to the one person, they could just direct message them through the platform.
Figure 4.11. Association for Direct act category and “Ask” fraction, data subset #1.\textsuperscript{108}

The fact that the variable-specific hypothesis was supported is interesting, and would seem to coincide with Austin’s proposed “performative formula” from his original 1957 account (see section 1.1.1). That is, that any speech act can be performed explicitly using the exact verb that denotes the action, for example, “I ask whether you have had lunch,” instead of “Have you had lunch?” Austin does not state that a given “speech act verb” is \textit{only} used to explicitly perform its denoted action, but if the formula is reliable, we might expect the outcome

\textsuperscript{108} It should be noted that the different variable fractions in this figure, and those in figures 4.12 and 4.22, are not batched, as there were so few instances of these verbs in the corpus that only these fractions appeared. A reader points out that presenting the few fractions that exist in this way is misleading, as it suggests that these fractions were the only possible outcomes. While there does seem to be a correlating relationship, it may have been more reflective to designate a single threshold value instead of including all fractions that happen to appear in each plot.
hypothesized by the variable-specific hypothesis here (that each verb would positively correlate with whichever speech act category was the super-category for the act denoted by the verb). Though the performative formula does not hold in many cases (again, see section 1.1.1), it worked here, which is interesting.

Meanwhile, the association with directive was not necessarily expected but is perfectly reasonable, as “ask” is an easy and fulfillable command to give online. This result will be revisited, along with the other speech act verbs, at the end of this section.

“Tell” verb fraction: The variable-specific hypothesis for all of the “speech act verbs,” that the verb will positively correlate with whichever category best reflects the super-category for the act denoted by the verb, is refuted in this case, as it would predict that “tell” positively correlates with the assertive category. The only associations found for this variable in figure 4.12 are: low, non-zero “tell” fraction ↑ commissive; low, non-zero “tell” fraction ↑ directive.
Figure 4.12. Association for Direct act category and “Tell” fraction, data subset #1.

Again, the association with directive was not necessarily expected but is perfectly reasonable. “Tell” is another command that is easy to give and easy to fulfill online. This result will be revisited, along with the other speech act verbs, at the end of this section. The association with commissive is unexpected, and has no immediate explanation.

**Proper noun fraction**: There was no variable-specific hypothesis for proper noun fraction, but strong trends are found in figure 4.13 nonetheless: *presence of proper nouns* ↑
assertive; absence of proper nouns ↑ commissive; absence of proper nouns ↑ expressive; absence of proper nouns ↑ directive;\textsuperscript{109} low non-zero proper noun fraction ↑ interrogative.

Figure 4.13. Association for Direct act category and Proper noun fraction (batched), data subset #1.

These results do not have an immediately obvious explanation, and may be simply noise. However, they do appear to pattern with some other part-of-speech results (i.e. common noun fraction, determiner fraction, etc.) as part of a multivariable pattern. This pattern will be revisited at the end of this section.

**Numeral fraction**: There was no variable-specific hypothesis for numeral fraction, but strong trends are found in figure 4.14 nonetheless: presence of numerals ↑ assertive; absence of numerals ↑ expressive; absence of numerals ↑ directive; low, non-zero numeral fraction ↑ commissive; non-zero numeral fraction of 0.4 ↑ exercitive.

\textsuperscript{109} Small association for proper noun fraction of 0.6 ↑ directive.
First of all, the spike for the exercitive category in figure 4.14 is expected and uninteresting, given that only two segments in the majority-annotation version of the corpus were designated as exercitive and one of them happens to have a numeral in it.

As with proper noun fraction, there is no immediate explanation for the associations we see here and they may be noise, but they may also be part of a multivariable pattern. This will be revisited at the end of this section.

**Nominal verb contraction fraction:** There was no variable-specific hypothesis for nominal verb contraction fraction, but strong trends are found in figure 4.15 nonetheless:

- absence of nominal verb contractions ↑ assertive;
- absence of nominal verb contractions ↑ directive;
- absence of nominal verb contractions ↑ interrogative;
- low, non-zero nominal verb contraction fractions ↑ commissive;
- non-zero nominal verb contraction fractions ↑ expressive.
There is a very reasonable explanation for these particular associations: nominal verb contractions appear in top bigrams that survived model selection and/or have a high MDA ranking for both commissive (“i’m going”, “i’m gonna”) and expressive (“i’m so”).

**Determiner fraction**: There was no variable-specific hypothesis for determiner fraction, but strong trends are found in figure 4.16 nonetheless: *presence of determiners* ↑ *assertive; absence of determiners* ↑ *commissive; absence of determiners* ↑ *interrogative; absence of determiners* ↑ *expressive.*
Figure 4.16. Association for Direct act category and Determiner fraction (batched),
data subset #1.

This is another part-of-speech variable for which there is no immediate explanation for
the associations we see, but which may be part of a multivariable pattern. This will be revisited
at the end of this section.

**Segment length**: The variable-specific hypothesis that short segments would correlate
with expressive is supported by figure 4.17, as *length 1-4 ↑ expressive*. The hypothesis that short
segments would positively correlate with assertive is not supported, at least not for the shortest
segments (1-4 tokens). However, *length 5-16 ↑ assertive*. Other associations noted: *length 1-4 ↑
directive; length 5-8 ↑ commissive; length 5-12 (and 17-20) ↑ interrogative.*
Figure 4.17. Association for Direct act category and Segment length (batched), data subset #1.

The expressive result is reasonable, and likely explained by interjections and adjectival fragments judged to be copular with the speaker or another contextually-salient subject. The assertive result is interesting. As noted above, a verb fraction of 0 was positively correlated with assertive, suggesting that many verbless fragments are indeed assertive. However, they must be longer fragments.

**Modal fraction:** The variable-specific hypothesis that modal fraction would positively correlate with commissive is supported by figure 4.18, as presence of modals ↑ commissive. The hypothesis that it would positively correlate with directive was not supported, as no strong
associations (either positive or negative) are seen for the directive category. Finally, the hypothesis that the presence of modals would positively correlate with the assertive category is supported, \( \text{low, non-zero modal fraction } \uparrow \text{ assertive} \). Other associations noted: \( \text{absence of modals } \uparrow \text{ expressive} \); \( \text{low, non-zero modal fraction } \uparrow \text{ interrogative} \).

![Figure 4.18. Association for Direct act category and Modal fraction (batched), data subset #1.](image)

The commissive result is expected, seeing as “will” was included in the group of modals checked for, and the commissive bigram “i will” was found to be a significant variable.

The expressive results are reasonable, seeing as an expressive was defined as an utterance wherein the speaker expresses a psychological state. If a user were to express a potential future psychological state (e.g. “I will be happy when X happens,”), that would likely be classified as either a commissive or assertive segment, depending on the exact form. Likewise for other hypothetical utterances of psychological state, such as “I should be mad,” which would be an
assertive. In the definition of expressive, therefore, it seems that there is an implicit requirement that the segment purport to express a current, non-hypothetical psychological state.\textsuperscript{110}

The results for assertive and interrogative are what would be expected (general presence of modals), but the results for directive were unexpected, given the hypothesis. The hypothesis that directive acts would positively correlate with the presence of modals was based in the fact that second person subjects combined with deontic modals may be interpreted as orders (e.g. parent to child, \textit{You must clean your room before dinner}).

However, it is also true that imperative utterances cannot contain modal auxiliaries (e.g. \textit{Come}). It would appear that perhaps the imperative type of directive was more common than the deontic modal type.

\textbf{Adjective fraction:} The variable-specific hypothesis that adjective fraction would positively correlate with the expressive category is supported by figure 4.19, \textit{adjective fraction between 0.4 and 1} \uparrow \textit{expressive}. Other associations noted: \textit{absence of adjectives} \uparrow \textit{commissive}; \textit{absence of adjectives} \uparrow \textit{directive}; \textit{low, non-zero adjective fraction} \uparrow \textit{assertive}; \textit{low adjective fraction} \uparrow \textit{interrogative}.

\textsuperscript{110} This does not, of course, preclude the existence of a non-serious expression of psychological state.
Figure 4.19. Association for Direct act category and Adjective fraction (batched), data subset #1.

The expressive result is reasonable, given that predicate adjectives are a common method of expressing a psychological state (e.g. *I’m happy, I’m bored, I’m afraid*). Additionally, in a more general capacity, description can be a type of expression (one’s feelings/opinion about something can come out through one’s description).

Especially interesting is the strong association between expressive and an adjective fraction of 1. This suggests that fragments consisting of only adjectives are usually judged to be copular with the speaker or a contextually-salient subject (as was hypothesized for segment length and verb fraction).

The results for the other categories will be revisited at the end of this section, as they are part of multivariable patterns.

**Vulgar word fraction:** The variable-specific hypothesis is supported by figure 4.20, as
presence of vulgar words $\uparrow$ expressive. Other associations noted: low, non-zero vulgar word fraction $\uparrow$ directive; low, non-zero vulgar word fraction $\uparrow$ interrogative.

![Figure 4.20](image)

**Figure 4.20. Association for Direct act category and Vulgar word fraction (batched), data subset #1.**

The expressive result is expected, as vulgar words routinely occur as interjections or in exclamatives, and possess an inherent expressive content. This meaning is so inherent that segments consisting of only vulgar words (fraction = 1) are classified as expressive, as seen in figure 4.20.

The interrogative and directive results will be revisited at the end of this section, as part of a multivariable pattern in connection with the indirect act results.

**Preposition fraction:** There was no variable-specific hypothesis for preposition fraction, but some associations are seen in figure 4.21 nonetheless: absence of prepositions $\uparrow$ expressive; non-zero preposition fractions $\uparrow$ assertive; non-zero preposition fractions $\uparrow$ commissive; non-zero
preposition fraction ↑ interrogative; zero or middling non-zero preposition fractions ↑ directive.

Figure 4.21. Association for Direct act category and Preposition fraction (batched), data subset #1.

The results for this variable are likely just noise. The strong association between a preposition fraction of zero and the expressive category could be linked to the fact that short segment length exhibits positive associations with the expressive category. That is, longer modifiers like prepositional phrases may simply be uncommon in expressive segments.

“Give” verb fraction: The variable-specific hypothesis for all of the “speech act verbs,” that the verb will positively correlate with whichever category best reflects the super-category for the act denoted by the verb, does not make a very clear prediction for “give.” The act of “giving” does not seem like a speech or speech-related act at all, unless you view it as a declaration of transfer of possession (e.g. “I give you this,”). For this reason, under the hypothesis, I would have predicted exercitive to correlate with “give.” Such was not supported,
however. The strongest association seen for this verb in figure 4.22 is presence of “give” ↑ directive.

Figure 4.22. Association for Direct act category and “Give” fraction, data subset #1.

As with “ask” and “tell”, the association of “give” with directive was not necessarily expected but is reasonable. As long as one was referring to giving some sort of information, “give” is yet another command that is easy to give and easy to fulfill online. This result will be revisited, along with the other speech act verbs, at the end of this section.
Common noun fraction: There was no variable-specific hypothesis for common noun fraction, but some associations are seen in figure 4.23 nonetheless: absence of common nouns ↑ expressive; absence of common nouns ↑ commissive; low and middling non-zero common noun fraction ↑ assertive; low common noun fraction ↑ interrogative; middling common noun fraction ↑ directive; non-zero common noun fraction ↑ exercitive.

Figure 4.23. Association for Direct act category and Common noun fraction (batched), data subset #1.

The spike for exercitive, as with the spike for exercitive in figure 4.14, is due to the fact that only two segments in the corpus were classified as exercitive, and they happen to contain common nouns.
This is another part-of-speech variable for which there is no immediate, reasonable explanation for the associations we see, and may contain quite a bit of noise. But it also may be part of a multivariable pattern. This will be revisited at the end of this section.

**Adverb fraction:** The variable-specific hypothesis that adverb fraction would positively correlate with the expressive category is not well supported by figure 4.24. Although *adverb fraction between 0.4 and 0.6* $\uparrow$ *expressive*, it was also the case that *absence of adverbs* $\uparrow$ *expressive*. A similar split occurred for directive, *adverb fraction between 0.2 and 0.4* $\uparrow$ *directive* and *absence of adverbs* $\uparrow$ *directive*. Other associations noted: *low, non-zero adverb fraction* $\uparrow$ *assertive*; *non-zero adverb fraction* $\uparrow$ *commissive*; *non-zero adverb fraction* $\uparrow$ *interrogative*.

![Figure 4.24. Association for Direct act category and Adverb fraction (batched), data subset #1.](image-url)
This is another part-of-speech variable for which there is no immediate, reasonable explanation for the associations we see, and which is likely mostly noise. But it may also be part of a multivariable pattern. This will be revisited at the end of this section.

**Original/reply status:** The variable-specific hypothesis for original/reply status is supported by figure 4.25, *reply post ↑ directive*. Other associations noted: *reply post ↑ commissive; reply post ↑ expressive; original post ↑ assertive; original post ↑ interrogative* (and also *original post ↑ exercitive*).

![Figure 4.25. Association for Direct act category and Original/reply status, data subset #1.](image)

Together, these results suggest that reply posts are less likely to be general comments on a post (assertive) or follow-up questions (interrogative), and more likely to be: advice about what action the original poster should take with regard to an issue they have raised (directive), a
commitment by the replier with regard to an issue the original poster has raised (commissive), or
an expression of the replier’s emotions about an issue the original poster has raised (expressive).

**Coordinate conjunction fraction**: There was no variable-specific hypothesis for
coordinate conjunction fraction, but some associations are seen in figure 4.26 nonetheless. The
strongest were: low, non-zero coordinate conjunction fraction ↑ commissive; non-zero coordinate
conjunction fraction ↑ interrogative; non-zero coordinate conjunction fraction ↑ expressive;
coordinate conjunction fraction of 0 or between 0.4 and 0.6 ↑ assertive; coordinate conjunction
fraction of 0 or between 0.2 and 0.4 ↑ directive.

![Figure 4.26. Association for Direct act category and Coordinate conjunction fraction (batched), data subset #1.](image)

To interpret these results, it is useful to recall that the segmentation guideline for
cojointed elements dictated that the coordinate conjunction accompany the segment that follows
it (see section 2.2.2.1). Thus, coordinate conjunctions were segmented with the second (or final)
segment of the coordination. Looking at figure 4.26, it seems to be that commissive, interrogative, and expressive segments are the final segment of a coordination much more often than the first (or middle) segment. Meanwhile, the assertive and directive categories exhibit associations that suggest they are usually the first (or middle) segment of a conjunct (or not in one) and only sometimes the final segment.

This pattern is interesting. It may be that the speaker asserts something or gives a directive, and then follows that with a commitment, question, or expression about that assertion/direction. However, it is also quite possible that these results are simply noise.

**Emoji in tweet fraction:** The variable-specific hypothesis for emoji in tweet fraction is supported by figure 4.27, as generally presence of emoji in tweet ↑ expressive. Other associations noted: absence of emoji in tweet ↑ interrogative; non-zero emoji in tweet fraction ↑ commissive; non-zero emoji in tweet fraction ↑ directive; emoji in tweet fraction of 0 or higher than 1 ↑ assertive.
Figure 4.27. Association for Direct act category and Emoji in tweet fraction (batched), data subset #1.

The expressive results are expected, as emoji are generally linked to emotional expression (Riordan, 2017). There is no immediate explanation for the other noted associations, and they are likely simply noise.

**Interjection fraction**: The variable-specific hypothesis for interjection fraction is supported by figure 4.28, as *presence of interjections* ↑ *expressive*. Other associations noted: *absence of interjections* ↑ *assertive; absence of interjections* ↑ *commissive; interjection fraction between 0.4 and 0.6* ↑ *directive; low, non-zero interjection fraction* ↑ *interrogative.*
Figure 4.28. Association for Direct act category and Interjection fraction (batched), data subset #1.

The expressive results are exactly what would be expected, as interjections are essentially expressive by definition. Meanwhile, the interrogative and directive results are not necessarily expected, but may connect to other results in a multivariable pattern which will be discussed at the end of this section.

**Tweet length:** The variable-specific hypothesis here, that assertive acts would appear in longer tweets, is supported by figure 4.29, though not to the exclusion of other assertive associations or other associations with longer tweets. The various associations observed were: *length 1, 4-6, or 9+ ↑ assertive; length 2-3 or 8 ↑ commissive; length 4-9 ↑ directive; length 3-4 ↑ exercitive; length 1-3 or 7 ↑ interrogative; length 2-3 or 7 ↑ expressive.*
These results are quite dynamic, and likely mostly noise. The variable-specific hypothesis was supported, however. One possible explanation for these results, based on that hypothesis, is that assertive segments can be informative and explanatory. They could be used to provide more information about any type of segment. In short, assertive segments may appear in longer tweets because they are what is making the tweet longer. They do not just appear in longer tweets, they comprise them. That cannot be the only explanation for the results we see, however, as directive segments also consistently appear in longer tweets.
Multivariable Pattern #1: Speech Act Verbs

Finally, I will review some patterns that we see if we look at certain variables together. The results for the three “speech act verbs” (ask, tell, and give) were interesting. The variable-specific hypothesis for those verbs was that they would positively correlate with whichever category best reflects the super-category for the act denoted by the verb. As figures 4.11, 4.12, and 4.22 showed, however, they all instead positively correlated with the directive category (though in one case, for ask, the verb did also associate with its “super-category,” interrogative).

There is a reasonable explanation for these results. Essentially, speech act verbs may be less likely to be used to perform the denoted act, and more likely to be used to mention the denoted act. The directive associations are a result of this, as the speaker is naming the act that they are ordering the hearer to perform.

Multivariable Pattern #2: Bigrams

The second multivariable pattern involves the bigrams that appear in the list of repeat variables in 4.2.1.1.5. To review, the bigrams included as explanatory variables in this analysis were those with the highest tf-idf by category. Only some of those bigrams survived model selection and appear in the list of repeat variables. Most of these surviving/repeat bigrams are so naturally associated with the category indicated that they would most likely be classified as such without any further text or context (e.g. “i will” alone would be categorized as commissive, “thank you” alone would be categorized as expressive). By comparing these surviving/repeat bigrams according to category, we can see commonalities that may function as potential IFIDs.

111 The exceptions, in my interpretation, are “watch it” (for commissive), “gotta stop” (for directive), and “$5.95 chef” (for exercitive).
Five expressive bigrams appear on the repeat variable list: “i love,” “i hate,” “i’m so,” “i miss,” and “thank you.” With the exception of “i’m so,” these bigrams all contain an explicit verb of emotive action. And with the exception of “thank you,” they all contain the first person pronoun “I.” Neither of these commonalities is especially surprising, given Alston’s condition for successful performance of an expressive act: that the speaker take responsibility for possessing a certain psychological state. But nonetheless, it is interesting that expressive statements wherein the speaker literally professes to have a certain psychological or emotional state are common enough to generate so many high tf-idf bigrams of the template “[speaker] and/or [emotive verb].” Furthermore, this template proved to be useful enough in classification that they survived model selection.

Four commissive bigrams appear on the repeat variable list: “watch it,” “i will,” “i’m gonna,” and “i’m going.” With the exception of “watch it,” these bigrams have predictable commonalities similar to the expressive bigrams. Commissive acts were defined to the annotators as acts wherein the speaker commits themself to a future action. The three similar commissive bigrams all contain the first person pronoun “I,” and a future auxiliary verb or immediate future verb. Again, it is interesting that commissive statements wherein the speaker literally refers to their own future actions were common enough to generate these results.

Two interrogative bigrams appear on the repeat variable list: “am i” and “is it.” Both of these bigrams exhibit a linguistic phenomenon that is not at all unexpected in an interrogative: subject-verb inversion. This is definitely another potential IFID for interrogative direct acts.

Two directive bigrams appear on the repeat variable list: “gotta stop” and “be safe.” In this case, there is not a strong commonality between the two bigrams. While “be safe” would
most likely be classified as a directive without any additional context, “gotta stop” would not necessarily.

The last bigram that appeared on the repeat variable list was for the exercitive category: “$5.95 chef.” Seeing as there were only two segments in the corpus that were classified as exercitive, it is not surprising that an unusual bigram taken from one of them proved to be a useful tool for classification.

Finally, the assertive category was the only category for which no high tf-idf bigrams appeared on the repeat variable list. If we return to the definitions of the categories, this is a reasonable result. Most of the bigrams above consist of what would most likely be the first two words of a segment (the exceptions, again, are “watch it” as a commissive, “gotta stop” as a directive, and “$5.95 chef”). Moreover, they would be the first two words (or words of the first two lexical categories) in literal examples of these segments. Because the semantic subject of commissives and expressives is the speaker by definition, the first words of literal examples of these categories would be “I” followed by a future verb or auxiliary (commissive) or an emotive verb (expressive). Because the semantic subject of directives is the addressee, the first word of a literal example of such a segment should be either “you” or a verb in the imperative. For interrogative, the subject of the segment is not restricted to the speaker or addressee. However, the first two words of a polar interrogative would be a subject-verb inversion (and the first two words of a wh-interrogative would be a wh-word + verb).

Meanwhile, for assertive segments, the subject of the segment is not restricted to the speaker or addressee, and the first two words of a literal assertive would not exhibit subject-verb
inversion. That is, there are fewer restrictions on what may constitute the first words of a literal example of an assertive segment.

In short, it would appear that, for categories other than assertive, there exist certain initial bigram templates that may function as IFIDs, as least for literal examples of the acts.\textsuperscript{112}

**Multivariable Pattern #3: Determiner Phrase**

The third multivariable pattern involves which of the basic elements of the determiner phrase pattern together across which categories. Elements considered were numerals, determiners, adjectives, and common nouns, which all independently survived model selection. For the assertive category, all of these elements have positive non-zero associations. This trend is echoed by figure 4.13, showing positive non-zero association with proper noun fraction, as well. Together, these results suggest that complete, modified DPs are a common element in assertive segments.

For the commissive category, determiners, adjectives, and common nouns have negative non-zero associations. Again, this trend is echoed by figure 4.13, showing negative non-zero associations with proper noun fraction. Together, these results suggest that modified (or unmodified) DPs are not a common element in commissive segments.

For the expressive category, determiners, numerals, and common nouns have negative non-zero associations, a trend with is again echoed by the proper noun fraction, together suggesting that DPs are not a common element in expressive segments.

\textsuperscript{112} I did not discuss exercitive because there was only the one bigram (taken from one of only two exercitive segments in the corpus), but I believe the first two words of a literal example of an exercitive would also be fairly predictable. In the most literal example, they would be “I hereby.”
However, notably, adjectives were not included in this pattern for expressive. Adjectives show positive non-zero associations with expressive, despite the negative associations with nouns, as mentioned above.

Meanwhile, for directive and interrogative, the various elements of the DP did not pattern together as exactly, but generally exhibited associations with zero or only low levels of each element.

A potential explanation for this multivariable pattern may be related to the variable-specific hypothesis that was made for tweet length. For that variable, I hypothesized that assertive segments would appear in longer tweets because they are informative and explanatory, and may be responsible for any background information/context necessary to interpret the tweet. This hypothesis was supported (see figure 4.29). If assertive segments are responsible for setting the scene and providing context, that may also explain why full determiner phrases appear in them. They provide the background information that other segments (of all speech act categories) can then simply refer to with pronouns.

**Multivariable Pattern #4: Expressive Language**

The final multivariable pattern looks at expressive language in segments that were not ultimately classified as expressive, and is connected to the theory and findings of Potts (2007a). In short, if expressive language is inserted into an otherwise, say, assertive utterance, then is the segment expressive or assertive? Potts says both, but in separate dimensions. Potts terms this quality the “independence” of expressive meaning from descriptive meaning (p.167).

Annotators in the present study were not able to declare two direct acts for a segment. If they were faced with such a conundrum, they were forced to relegate one of the acts to the
indirect level. It may have been the case that expressive was therefore usually the act relegated to the indirect level. If this is the case, then it would suggest that expressive language does not always override what the force of the utterance would be without the expressive language.

If annotators were consistently faced with expressive language in otherwise assertive and interrogative utterances, one possible result could be the associations seen in figure 4.9, (expressive indirect act ↑ assertive direct act, expressive indirect act ↑ interrogative direct act). If that were the case, we should expect to find positive non-zero associations with expressive elements like vulgar words, emotion words, adjectives, adverbs, incorporated emoji, and interjections in assertive and interrogative segments. Four of those six elements survived model selection (see figures 4.17, 4.18, 4.24, and 4.28), while the association plots for the remaining two are given below. The associations for emotion word fraction are given in figure 4.30, and for incorporated emoji yes/no in figure 4.31.
Figure 4.30. Association for Direct act category and Emotion word fraction (batched), data subset #1.
Three of the six elements show positive non-zero associations with the assertive category: adjective fraction, adverb fraction, and emotion word fraction. For the interrogative category, vulgar word fraction, adverb fraction, and interjection fraction show positive non-zero associations, while adjective fraction shows positive associations for both zero and low, non-zero values. Thus, it may be that the expressive elements in the segments contributed to the annotator’s designation of an expressive indirect act.

If we want to tie the presence of expressive language to the presence of an expressive indirect speech act in otherwise non-expressive segments, we should also see that a lack of expressive language correlates with a lack of expressive indirect acts. In the present study, that
would mean that we should see a lack of expressive language in both commissive and directive acts.

While this holds true for commissive (i.e. of the six expressive elements considered, only adverb fraction shows any positive non-zero associations with commissive), it does not hold for directive. The directive category exhibits positive non-zero associations for vulgar word fraction, incorporated emoji, and interjection fraction (and positive associations with both zero and non-zero values for emotion word fraction and adverb fraction). We would have expected, therefore, that directive would also show associations with expressive indirect acts. Such is not the case.

However, directive direct acts do show a positive association with directive indirect acts, curiously. It may be, therefore, that expressive language is a signal of the mere presence of indirect acts, but not necessarily the force. This possibility will be explored more in section 4.2.2, which will address features which may signal the presence of indirect acts.

In short, for the moment, we may at least hesitantly conclude that the presence of expressive language in a tweet does not necessarily override the direct force of the utterance without that language. Whether this expressive language correlates with the presence of an indirect act, especially an expressive one, will be addressed in section 4.2.2.

4.2.1.2.2 Summary and Interpretation

Table A.1 in appendix 2 summarizes the various possible attributes for each speech act category as discussed in section 4.2.1.2.1, except for those associations which were clearly due to correlation with a high-frequency bigram (though the bigrams themselves were included).\textsuperscript{113}

\textsuperscript{113} For example, nominal-verb contraction fraction was omitted, but “i’m so,” “i’m gonna,” and “i’m going” were included.
Additionally, the exercitive category was left out entirely seeing as there were so few exercitive segments that any associations found for that category could hardly be anything other than noise.

In this section, I will step back and assess the actual significance of these results, and what they may together suggest about IFIDs and speech act interpretation/classification.

Most of the features and associations discussed in section 4.2.1.2.1 are not “exclusive” IFIDs, in the sense that presence or absence of the feature indicates that the segment should be classified as one specific type of act. However, viewed together, the various collections of features may be seen to constitute templates of the typical act of each type. Such a template can itself function as an IFID. A hearer vacillating between two interpretations might finally make a decision because the segment seems more “typical” of (i.e. exhibits more of the attributes of), say, an assertive than a commissive.

Some of the features discussed in section 4.2.1.2.1 are clearly stronger or more significant indicators than the rest. Punctuation was the only variable that survived model selection for all four regressions and appeared in the top-ranked variables (by MDA) for all four random forest analyses, no doubt due to the strong, positive association between the presence of a question mark and a classification of interrogative. Wh-word fraction was close behind punctuation, appearing in all four regressions and two of four top MDA rankings. It seems no surprise, therefore, that the interrogative category received the second-best classification accuracy for 5 of the 8 models (first place went to assertive for all 8).

Meanwhile, the second-best classification accuracy for the remaining 3 of the 8 models went to expressive. Unsurprisingly, both of the next two most frequently appearing variables
were expressive bigrams: “i love” (4/4 regressions, 2/4 MDA rankings), and “i hate” (3/4 regressions, 2/4 MDA rankings).

The fact that some features (e.g. the question mark) are more consistent than others in terms of which type of act they associate with does not, however, mean that those features are not still only typical of a certain act. That is, they are not foolproof category indicators. Even in the case of the question mark, not every segment that ended with one was annotated as interrogative. For example, one such segment was, “I’m nervous????”, which was annotated as expressive.

For many of the variables in this study, of course, they were not expected to be categorical, only ever occurring within one type of act. For example, the various part-of-speech variables would not have been expected to be categorically associated with only one type of act. They would be expected to occur in all different types of acts, just in differing amounts. But it is important to note that even the variables for which a categorical result would not have been unreasonable to expect or find (e.g. question mark) did not yield categorical results.

It should also be noted that the features assessed in this analysis are generally only surface features (indirect act choice and serious yes/no excepted). More complex features, such as syntactic structures, were not considered explicitly (although one, the determiner phrase, was discovered as a potential feature after the fact).

In short, this analysis was designed to identify some typical overt attributes of each category given a variety of features, and that is ultimately what it did, generating a potential glimpse at a typical template for each category. These “typical templates” may themselves be what we should be referring to as IFIDs, as opposed to individual features.
Even though these typical templates are not an unexpected result of this type of analysis, they may still be illuminating. First of all, the fact that any such templates were found would appear to refute the most extreme parts of the contextualist branch of speech act theories. If the results had been entirely inconsistent, with no discernible associations between the variables tested and the speech act classification given, that would have supported such a contextualist theory, wherein illocutionary force is entirely inferred and not explicitly encoded. The fact that this was not the case, that several overt elements did exhibit strong associations with certain speech act categories, lends support to the opposing literalist branch of speech act theories.

Additionally, even though the methodology was designed to uncover the typical attributes of the various types of acts, the exact forms that these templates might take were generally unknown. Certain results discussed in section 4.2.1.2.1 were unexpected or even entirely counter to what was hypothesized, and certain variables dropped out that I would not have expected to drop out (e.g. incorporated emoji yes/no).

Finally, with these results in mind, I will briefly return to the topic of which data subset yielded the highest classification accuracy. As stated in section 4.2.1.1.2, the fact that the majority-annotation version of the corpus yielded better results than the full version may suggest that the differing annotations represent more erroneous noise than non-erroneous ambiguity. From the “typical template” view of the results, however, it may be that segments that received different annotations are simply atypical of one or both of those annotations. To borrow an image from exemplar theory, they may be at the edges (or perhaps the intersection) of the exemplar clouds for both speech act categories. While the segment would theoretically exist at this intersection regardless of whether it is labelled with only one of the categories or both of
them, it is reasonable that being labelled with only one of them is less confusing to the model. This typical template assessment does not rule out erroneous annotation, but it makes it more difficult to detect.

Meanwhile, as stated in section 4.2.1.1.2, the fact that the majority-annotation version of the corpus without the attention and quality control questions performed better than the version with them would seem to suggest that the “correct” answers to the AQĈs were poorly chosen, and added inconsistency to the input. However, it is important to note the difference in baseline accuracy between the two models. Assertive was by far the most common category, in all four subsets. The effect of this fact on the models is clear, based on the category-specific classification accuracies in tables 4.20 and 4.23. Assertive was judged by the models to be the default category, so one measure of baseline accuracy is what the result would be if the model simply guessed assertive every time. Data subset #1 (majority-annotation, no AQĈs) consisted of 69.0% assertive segments. Data subset #3 (majority-annotation, AQĈs included) consisted of 66.4% assertive segments. Hence, the baseline accuracy for the model trained on data subset #1 was higher than for the model trained on data subset #3.

In fact, if we judge which model performed better in terms of increase from baseline, the model trained on data subset #3 performed better on the regressions (10.8% improvement on baseline versus 8.3% improvement on baseline for the model trained on data subset #1), while the model trained in data subset #1 performed better on the random forests, though not by much (8.9% improvement on baseline versus 8.5% improvement on baseline for the model trained on data subset #3). Given this fact, and the fact that the AQĈs ultimately did very little gatekeeping during the annotation (see section 3.1.3.1.1), I do not believe that the relatively lower
classification accuracies for the models trained on data subset #3 were a result of poorly chosen AQC answers.

4.2.2 Analysis and Discussion of Hypothesis 2

**Hypothesis 2:** Certain overt elements available to Twitter users signal the presence of an indirect act. Candidate elements include direct act category, lexical choice, capitalization presence and/or choice, punctuation presence and/or choice, emoji presence and/or choice, tag presence and/or choice, unit position within the tweet, and status as an original post vs. reply post.

This hypothesis will be assessed through logistic regression and random forest techniques, with indirect act presence as the response variable. The list of explanatory variables is the same as that used for the assessment of hypothesis 1, with the exception that direct act category is now an explanatory variable and indirect act choice is not. Any variables that are shown by the model to be significant for classification will then be addressed in the discussion section, which will review the variable-specific hypotheses and attempt to propose some potential explanations for the associations observed. It should be noted that the response variable, indirect act presence, yielded poor agreement (see section 4.1), and the results observed in this section for specific features should be viewed with a skeptical lens. Meanwhile, the fact that indirect act presence yielded poor agreement itself supports a conclusion that there are not, in fact, overt elements that signal of an indirect act (at least not systematically, regardless of context).
4.2.2.1 Analysis

The response variable for this analysis was indirect act presence. With the exception of direct act category, the explanatory variables used to assess hypothesis 2 were the same as those for hypothesis 1. This was done because the variables used when assessing hypothesis 1 were mostly inspired by or taken from studies that were attempting to build an automatic classifier for speech act category. But none of those studies had the goal of separating direct and indirect acts. It is quite possible that some of the speech acts annotated in their manually-annotated training data would have been considered indirect by the annotators of the present study. This, in turn, might mean that some of the variables that proved to be useful for classification in the other studies were actually useful because of their correlation with certain indirect acts, as opposed to certain direct acts.

Additionally, the analysis was done using the same set of segments as for hypothesis 1 (only segments that performed a non-automatic direct act). This was done because all three of the variable-specific hypotheses for this analysis involve the direct act in some way, as will become apparent in section 4.2.2.1.1.

As with hypothesis 1, four different subsets of the data were tested, with the same training/test splits. These four subsets are a result of crossing two meta-variables: 1) whether the majority-annotation version (one row per segment) or the full version (three rows per segment, one for each annotator) of the corpus was used, and 2) whether attention and quality control segments were included or excluded.

Testing both the majority-annotation and full versions was done mostly to maintain the parallel with the analysis of hypothesis 1. For hypothesis 1, I proposed that if the
majority-annotation version of the corpus yielded better results than did the full version, this may suggest that, on the whole, the differing annotations represented more erroneous noise than non-erroneous ambiguity. However, a segment can perform multiple indirect acts at once, and it would be a difficult task to judge whether differing indirect annotations were a result of multiple indirect acts or simply “mistakes.” Hence, for the analysis of hypothesis 2, if the models trained on the majority-annotation version of the corpus perform better on unseen data than do those trained on the full version, this does not necessarily mean that the differing annotations represent more erroneous noise than non-erroneous ambiguity.

The reasoning behind testing the corpus with and without the attention and quality control segments was mostly to maintain the parallel with the analysis of hypothesis 1, but it may also function as a control of sorts. The inclusion of the attention and quality control segments should yield better results than their exclusion, as they simply provide more data, and the AQCs were never a matter of indirect act classification (only direct act). If the inclusion of the AQCs seems to yield poorer or less consistent results, that would be an unexpected outcome requiring explanation.

Finally, the same model selection methods were used for the analysis of hypothesis 2 as for the analysis of hypothesis 1. A description of those methods can be found in section 4.2.1.1.2.

4.2.2.1 Explanatory Variables

As noted in the introduction to this section, the variables used for this analysis were the same as those used for hypothesis 1 (with the addition of direct act category and the elimination
of indirect act choice). Descriptions of the variables can therefore be found in section 4.2.1.1.1 and are not repeated here, except that a note must be made about the bigram features used.

The bigram features here tested for the presence or absence of a certain bigram, as did the bigram features in hypothesis 1. Tf-idf was calculated for both categories within indirect act presence (“Yes” and “No”), using the training portions of each data subset. However, only the “No” bigrams were used, as dozens were tied for top rank in the “Yes” category. The top 5 “No” bigrams for each of the four subsets are given in table 4.24. For subset #3 (majority-annotation version, AQC’s included), several bigrams were tied for 5th place, so only the top 4 were used.

### Table 4.24. Bigrams for indirect act analysis.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Bigram rank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>#1</td>
<td>i have</td>
</tr>
<tr>
<td>#2</td>
<td>i love</td>
</tr>
<tr>
<td>#3</td>
<td>to be</td>
</tr>
<tr>
<td>#4</td>
<td>i love</td>
</tr>
</tbody>
</table>

In addition to the general hypothesis that certain overt elements will correlate with the presence or absence of an indirect act, there are a few variable-specific hypotheses to be made, born out of previous literature as well as the results from the analysis of hypothesis 1. There are three such hypotheses, all of which I will address in section 4.2.2.2.1.

The first variable-specific hypothesis is based on Clark’s (1979) study of indirect acts (see section 1.1.3). Specifically, I expect that the seriousness of the direct act is a significant

---

114 More specifically, “Yes, there is an indirect act,” and “No, there is not an indirect act.”
feature for classification. Also, I expect that the presence of an indirect act will positively associate with the designation of the segment’s direct act as non-serious. Clark’s actual findings were that a non-serious direct act is more likely to co-occur with an indirect act than not. That is, in the event of a non-serious direct act, an indirect act should be present. In the event of a serious direct act, an indirect act may still be present, but in such a case the acts should be designated as equally important to the speaker’s message. Hence, in the event that the presence of an indirect act does not positively associate with the designation of the segment’s direct act as non-serious, I will assess whether direct and indirect acts were usually judged to be equally important to the speaker’s message.\textsuperscript{115}

The second variable-specific hypothesis is based on the analysis of hypothesis 1. I expect to find that direct act category is a significant feature for classification, seeing as indirect act choice was significant for the classification of direct acts (see section 4.2.1.2.1).

The final variable-specific hypothesis is that certain expressive elements will positively correlate with the presence of an indirect act. As discussed under multivariable pattern #4 in section 4.2.1.2.1, some results from the analysis of hypothesis 1 suggest that the presence of expressive elements like vulgar words, emotion words, adjectives, adverbs, incorporated emoji, and interjections in segments that perform non-expressive direct acts may signal the presence of an indirect act (usually expressive).

\textsuperscript{115} This is possible because of the additional annotation question asked when a direct act was judged as serious in the presence of an indirect act: “Which act do you think is more important to the speaker’s overall message?”

299
4.2.2.1.2 Regression Analyses

The multinomial logistic regression function `multinom` from the R package `nnet` was used for these analyses. For all four data subsets, after model selection, no variables were marked significant by the final model. But several variables appeared in the final model of more than one regression, as summarized in table 4.25.

Table 4.25. Variables that appeared in the final model for more than one regression for indirect act analysis.

<table>
<thead>
<tr>
<th>Number of final models</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 of 4</td>
<td>Proper noun fraction</td>
</tr>
<tr>
<td>3 of 4</td>
<td>Punctuation, segment has incorporated emoji</td>
</tr>
<tr>
<td>2 of 4</td>
<td>Original/reply status, direct act serious yes/no, direct act category, segment has tone-type emoji adjacent</td>
</tr>
</tbody>
</table>

The classification accuracy of each model on the test portion of its data subset is provided in table 4.26. The baseline accuracy, what the accuracy would be if the model guessed the most common category every time (i.e. “No”), is also given, along with the difference between the two (“improvement on baseline”). The category-specific accuracies are given in table 4.27.
Table 4.26. Classification accuracy of final regression models for test portions of data, for indirect act analysis.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Classification accuracy</th>
<th>Baseline accuracy</th>
<th>Improvement on baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>MajAnno, No AQCs (#1)</td>
<td>90.2%</td>
<td>90.2%</td>
<td>0%</td>
</tr>
<tr>
<td>Full, No AQCs (#2)</td>
<td>83.4%</td>
<td>83.4%</td>
<td>0%</td>
</tr>
<tr>
<td>MajAnno, AQCs (#3)</td>
<td>90.5%</td>
<td>90.5%</td>
<td>0%</td>
</tr>
<tr>
<td>Full, AQCs (#4)</td>
<td>83.8%</td>
<td>83.7%</td>
<td>0.1%</td>
</tr>
</tbody>
</table>

Table 4.27. Category-specific classification accuracy of final regression models for test portions of data, for indirect act analysis.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>#2</td>
<td>99.3%</td>
<td>3.7%</td>
</tr>
<tr>
<td>#3</td>
<td>100%</td>
<td>0%</td>
</tr>
<tr>
<td>#4</td>
<td>99.6%</td>
<td>2.9%</td>
</tr>
</tbody>
</table>

The only model that improved on the baseline accuracy was the model trained on data subset #4 (full version of the corpus, AQCs included).

4.2.2.1.3 Random Forest Analyses

The function randomForest from the R package randomForest was used for these analyses. For all four data subsets, after model selection, no variables were marked significant by the final model. But several variables appeared with high mean decrease in accuracy values.
in multiple final models. Table 4.28 provides the variables that appeared in the top 5 variables (by MDA rank) for more than one final model.

Table 4.28. Variables that appeared in the top 5 for more than one random forest final model, for indirect act analysis.

<table>
<thead>
<tr>
<th>Number of final models</th>
<th>Variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 of 4</td>
<td>Direct act category</td>
</tr>
<tr>
<td>3 of 4</td>
<td>Segment position</td>
</tr>
<tr>
<td>2 of 4</td>
<td>Common noun fraction, segment length, original/reply status, @-tag in tweet, adjective fraction</td>
</tr>
</tbody>
</table>

The classification accuracy of each model on the test portion of its data subset is provided in table 4.29. The baseline accuracy, what the accuracy would be if the model guessed the most common category every time (i.e. “No”), is also given, along with the difference between the two (“improvement on baseline”). The category-specific accuracies are given in table 4.30.

Table 4.29. Classification accuracy of final random forest models for test portions of data, for indirect act analysis.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Classification accuracy</th>
<th>Baseline accuracy</th>
<th>Improvement on baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>MajAnno, No AQC s (#1)</td>
<td>90.8%</td>
<td>90.2%</td>
<td>0.6%</td>
</tr>
<tr>
<td>Full, No AQC s (#2)</td>
<td>83.4%</td>
<td>83.4%</td>
<td>0%</td>
</tr>
<tr>
<td>MajAnno, AQC s (#3)</td>
<td>90.9%</td>
<td>90.5%</td>
<td>0.4%</td>
</tr>
<tr>
<td>Full, AQC s (#4)</td>
<td>83.7%</td>
<td>83.7%</td>
<td>0%</td>
</tr>
</tbody>
</table>
Table 4.30. Category-specific classification accuracy of final random forest models for test portions of data, for indirect act analysis.

<table>
<thead>
<tr>
<th>Data subset</th>
<th>Category-specific accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No</td>
</tr>
<tr>
<td>#1</td>
<td>100%</td>
</tr>
<tr>
<td>#2</td>
<td>100%</td>
</tr>
<tr>
<td>#3</td>
<td>100%</td>
</tr>
<tr>
<td>#4</td>
<td>100%</td>
</tr>
</tbody>
</table>

The models trained on data subsets #1 and #3 both improved on the baseline accuracy.

4.2.2.1.4 Summary of Repeat Variables

The following list provides the variables that appeared in two or more regressions or two or more top random forest MDA rankings. The variables are presented according to the number of regressions and top random forest MDA rankings they appear in, starting with the most frequent. Ties are listed under the same number. Association plots for each variable will be given in the next section.

1) **Direct act category** (2/4 regressions, 4/4 MDA rankings)

2) **Proper noun fraction** (4/4 regressions, 1/4 MDA rankings)

3) **Original/reply status** (2/4 regressions, 2/4 MDA rankings)

4) **Punctuation** (3/4 regressions, 0/4 MDA rankings)

    - **Segment has incorporated emoji** (3/4 regressions, 0/4 MDA rankings)
    - **Segment position** (0/4 regressions, 3/4 MDA rankings)

5) **Direct act serious yes/no** (2/4 regressions, 0/4 MDA rankings)
Segment has tone-type emoji adjacent (2/4 regressions, 0/4 MDA rankings)

Common noun fraction (0/4 regressions, 2/4 MDA rankings)

Segment length (0/4 regressions, 2/4 MDA rankings)

@-tag in tweet (0/4 regressions, 2/4 MDA rankings)

Adjective fraction (0/4 regressions, 2/4 MDA rankings)

In addition to these individual variables, two interaction terms survived model selection for two of four regressions. These were the interactions of Direct act serious yes/no and Direct act category, and Hashtag in tweet and Non-incorporated hashtag adjacent, and they are shown in figures 4.32 and 4.33.

Figure 4.32. Association for Direct act category and Direct act serious yes/no, data subset #1.
4.2.2.2 Discussion

On a basic level, several overt elements survived the model selection phase of multiple analyses and demonstrate strong associations with the presence or absence of an indirect act. Therefore, the hypothesis that “Certain overt elements available to Twitter users signal the presence of an indirect act” would appear to be supported. Section 4.2.2.2.1 will address individual results and the variable-specific hypotheses, and section 4.2.2.2.2 will provide a more general summary and explanation of the results. It is worth noting, once again, that the indirect act present question yielded poor inter-annotator agreement, which itself suggests that there are
not, in fact, overt elements that signal the present of an indirect act (at least not in a systematic way, regardless of context).

4.2.2.2.1 Variable-specific Hypotheses and Results

In this section, I will first provide association plots for all of the repeat variables and summarize their results. At the end of this section, I will discuss whether the three variable-specific hypotheses given in section 4.2.2.1.1 were supported.

The training portion of data subset #1 (majority-annotation version, AQCs excluded) was used to generate these plots, because the random forest model trained on it produced the highest improvement on baseline out of all 8 models. As mentioned in section 4.2.2.1, the fact that the majority-annotation version of the corpus yielded better results does not necessarily mean that the differing annotations were erroneous noise, as a segment may perform multiple indirect acts. Meanwhile, as also mentioned in section 4.2.2.1, the inclusion of the attention and quality control segments should yield better results than their exclusion, as they simply provide more data, and the AQCs were never a matter of indirect act classification (only direct act). Hence, it is somewhat surprising that the model that performed the best was trained on a version of the corpus that excluded the AQCs. However, the only other models that improved on the baseline even marginally (regression #4 and random forest #3) were both trained on versions of the corpus that included the AQCs.

In this section, I will use shorthand to summarize the associations observed for each variable listed in section 4.2.2.1.4. A positive association will be noted with an upwards arrow ↑, while a negative association will be noted with a downwards arrow ↓.
In association plots, neither height of the association bar nor width is the more important feature. The height indicates the strength of the association, while the width indicates how many data points the association is based on. Therefore, one approximation of the significance of an association is the area of the bar. For each of the plots below, the bars with the highest approximate areas will be the ones addressed, in addition to any overall trends.

In several cases, the results have been batched for easier viewing. That is, instead of presenting the results of, say, 30 different “Proper Noun Fraction” values ranging between 0 and 1, those values have been grouped into 6 batches of values.116

**Direct act category:** The strongest associations in figure 4.34 are: presence of an indirect act ↑ assertive direct act; presence of an indirect act ↑ interrogative direct act; absence of an indirect act ↑ commissive direct act; absence of an indirect act ↑ expressive direct act.

---

116 In all cases where the values range between 0 and 1, the batches are: 0; 0 < x ≤ 0.2; 0.2 < x ≤ 0.4; 0.4 < x ≤ 0.6; 0.6 < x ≤ 0.8; and 0.8 < x ≤ 1. For segment length, the batches increase by 4 up to 32 (i.e. x ≤ 4, 4 < x ≤ 8, … 28 < x ≤ 32).
Figure 4.34. Association for Indirect act yes/no and Direct act category, data subset #1.

Proper noun fraction: Figure 4.35 shows that: presence of an indirect act ↑ absence of proper nouns, and vice versa (absence of an indirect act ↑ presence of proper nouns).
Figure 4.35. Association for Indirect act yes/no and Proper noun fraction (batched), data subset #1.

**Original/reply status**: Figure 4.36 shows that: *presence of an indirect act ↑ original post*, and vice versa (*absence of an indirect act ↑ reply post*).
Figure 4.36. Association for Indirect act yes/no and Original/reply status, data subset #1.

**Punctuation:** The strongest associations in figure 4.37 are: *presence of an indirect act* ↑ question mark; *presence of an indirect act* ↑ exclamation point; *presence of an indirect act* ↑ period; *absence of an indirect act* ↑ comma; *absence of an indirect act* ↑ other.
Figure 4.37. Association for Indirect act yes/no and Punctuation, data subset #1.

**Segment has incorporated emoji:** Figure 4.38 shows a strong association (based on relatively few data points): *presence of an indirect act → segment has incorporated emoji.*
Figure 4.38. Association for Indirect act yes/no and Segment has incorporated emoji, data subset #1.

**Segment position:** The strongest associations in figure 4.39 are: *presence of an indirect act* ↑ segment is entire tweet; *absence of an indirect act* ↑ segment is in middle of tweet.
Figure 4.39. Association for Indirect act yes/no and Segment position, data subset #1.

**Direct act serious yes/no:** Figure 4.40 shows that: presence of an indirect act ↑ direct act is non-serious, and vice versa (absence of an indirect act ↑ direct act is serious).
Figure 4.40. Association for Indirect act yes/no and Direct act serious yes/no, data subset #1.

Segment has tone-type emoji adjacent: Figure 4.41 shows that: presence of an indirect act ↑ segment has tone-type emoji adjacent, and vice versa (absence of an indirect act ↑ segment does not have tone-type emoji adjacent).
Figure 4.41. Association for Indirect act yes/no and Segment has tone-type emoji adjacent, data subset #1.

**Common noun fraction**: The strongest associations in figure 4.42 are: *absence of an indirect act* $\uparrow$ *absence of common nouns*; *presence of an indirect act* $\uparrow$ *common noun fraction between 0 and 0.2*. 
Figure 4.42. Association for Indirect act yes/no and Common noun fraction (batched), data subset #1.

Segment length: The strongest associations in figure 4.43 are: absence of an indirect act ↑ segment length 1-4; presence of an indirect act ↑ segment length 5-12.
Figure 4.43. Association for Indirect act yes/no and Segment length (batched), data subset #1.

[@-tag in tweet] Figure 4.44 shows that: presence of an indirect act ↑ absence of @-tag in tweet, and vice versa (absence of indirect act ↑ presence of @-tag in tweet).
Figure 4.44. Association for Indirect act yes/no and @-tag in tweet, data subset #1.

**Adjective fraction**: The strongest associations in figure 4.45 are: absence of indirect act

↑ absence of adjectives; presence of indirect act ↑ adjective fraction between 0 and 0.4.
Variable-specific Hypothesis #1: Clark (1979)

The first variable-specific hypothesis was supported on both counts: the seriousness of the direct act appeared as a significant feature in 2/8 analyses, and the presence of an indirect act was positively correlated with the designation of the segment’s direct act as non-serious. Clark’s main findings are therefore reproduced, and his proposal that seriousness of the direct act may contribute to the inferential process involved in deducing the presence of an indirect act is supported.
Variable-specific Hypothesis #2: Direct Act Category

The second variable-specific hypothesis was also supported: direct act category proved to be a significant feature for classification. In fact, it was the most significant feature for classification, surviving selection in 2/4 regressions and occurring in all 4 top MDA rankings for the random forests. The specific associations noted in figure 4.34 were generally expected, based on figure 4.9 in the analysis of hypothesis 1. The one exception is that directive direct acts, which figure 4.9 shows to positively correlate with the presence of directive indirect acts, show a negative association with the general presence of an indirect act in figure 4.34.

It is unsurprising that the strongest associations for the presence of an indirect act are with assertive and interrogative direct acts, seeing as those are the ones that figure 4.9 showed to positively correlate with expressive indirect acts, and expressive was the most common type of indirect act in the corpus. Additionally, the results for variable-specific hypothesis #2 appear to tie in to the results for variable-specific hypothesis #3, as will become apparent.

Variable-specific Hypothesis #3: Expressive Language

The final variable-specific hypothesis, born out of multivariable pattern #4 in section 4.2.1.2.1, was that certain expressive elements will positively correlate with the presence of an indirect act. Of the 6 expressive elements mentioned (vulgar words, emotion words, adjectives, adverbs, incorporated emoji, and interjections), only adjectives and incorporated emoji appeared in the repeat variable list in section 4.2.2.1.4. Both of these elements do give the hypothesized result, though. Figure 4.45 shows a positive correlation between the presence of an indirect act and a low, non-zero adjective fraction, while figure 4.38 shows a strong positive correlation between the presence of an indirect act and incorporated emoji in the segment.
Association plots for the remaining 4 expressive elements (that did not make the repeat variable list) are provided below, as figures 4.46 through 4.49. In all four plots, low, non-zero values of the element positively correlate with the presence of an indirect act. In 3 of 4 plots (emotion words being the exception), absence of the element positively correlates with the absence of an indirect act.

Figure 4.46. Association for Indirect act yes/no and Adverb fraction (batched), data subset #1.
Figure 4.47. Association for Indirect act yes/no and Emotion word fraction (batched), data subset #1.
Figure 4.48. Association for Indirect act yes/no and Vulgar word fraction (batched), data subset #1.
The plots for the 5 expressive elements that are fractional (i.e. all except incorporated emoji) show an interesting pattern. Specifically, that low, non-zero values of the element positively correlate with the presence of an indirect act, but higher values of the element positively correlate with the absence of an indirect act.

I suspected that the segments with high fractions of these expressive elements were being annotated as performing expressive direct acts -- which, the results have shown (both figures 4.6 and 4.9), are unlikely to perform indirect acts, especially expressive indirect acts (the majority of indirect acts). In other words, my hypothesis was that high values of these expressive elements would positively correlate with an expressive direct act, while lower values of these expressive elements would positively correlate with an expressive indirect act. Based on this suspicion, I
took only the segments that performed either an expressive direct or expressive indirect act\textsuperscript{117} and looked at the associations of the 5 fractional expressive elements against the metric of directness. The results are given below, in figures 4.50 through 4.54. The one non-fractional feature (incorporated emoji presence) was also assessed, for completeness (figure 4.55).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure450.png}
\caption{Association for Expressive direct/indirect and Adjective fraction (batched), data subset #1.}
\end{figure}

\textsuperscript{117} This was a total of 330 segments. The 4 segments that were annotated as performing both an expressive direct act and an expressive indirect act were left out.
Figure 4.51. Association for Expressive direct/indirect and Vulgar word fraction (batched), data subset #1.
Figure 4.52. Association for Expressive direct/indirect and Emotion word fraction (batched), data subset #1.
Figure 4.53. Association for Expressive direct/indirect and Adverb fraction (batched), data subset #1.
Figure 4.54. Association for Expressive direct/indirect and Interjection fraction (batched), data subset #1.
Figure 4.55. Association for Expressive direct/indirect and Segment has incorporated emoji, data subset #1.

These figures generally support the hypothesis that high values of these expressive elements would positively correlate with an expressive direct act, while lower values of these expressive elements would positively correlate with an expressive indirect act. I propose that this pattern may be a result of what Potts (2007a) termed the “expressive dimension.” The expressive dimension is where the expressive meaning of an utterance resides, parallel to the descriptive meaning of that utterance. When expressive language is inserted into an otherwise, say, assertive utterance, Potts would say that the utterance performs both an assertive act and an expressive act, but in different dimensions.

As discussed under multivariable pattern #4 in section 4.2.1.2.1, annotators in the present study were not able to declare two direct acts for a segment. When faced with expressive
language in an otherwise non-expressive utterance, it would appear that the annotators’ solution was to annotate that expressive meaning as an indirect act. This would explain why low fractions of these expressive elements correlate with the presence of an indirect act. Inserting these elements into otherwise non-expressive utterances would result in low, non-zero fractions of these elements.

4.2.2.2 Summary

In this section, I will first briefly review the results in a more general capacity, assessing whether they would be expected given known conventions and reasonable deduction. I will then revisit the issue of expressive language and indirect acts. Table 4.31 summarizes the results shown in figures 4.34 through 4.45.

Table 4.31. Summary of hypothesis 2 analysis results.

<table>
<thead>
<tr>
<th>Absence of an indirect act</th>
<th>Presence of an indirect act</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Commissive or expressive direct act</td>
<td>• Assertive or interrogative direct act</td>
</tr>
<tr>
<td>• Proper nouns</td>
<td>• No proper nouns</td>
</tr>
<tr>
<td>• Reply post</td>
<td>• Original post</td>
</tr>
<tr>
<td>• Segment ends in comma, semi-colon, colon, or other</td>
<td>• Segment ends in question mark, exclamation point, or period</td>
</tr>
<tr>
<td>• No incorporated emoji</td>
<td>• Incorporated emoji</td>
</tr>
<tr>
<td>• Segment is in middle of tweet</td>
<td>• Segment is entire tweet</td>
</tr>
<tr>
<td>• Direct act is serious</td>
<td>• Direct act is non-serious</td>
</tr>
<tr>
<td>• No tone-type emoji adjacent</td>
<td>• Tone-type emoji adjacent</td>
</tr>
<tr>
<td>• Common noun fraction of 0</td>
<td>• Common noun fraction between 0 and 0.2</td>
</tr>
<tr>
<td>• Segment length is 1-4 tokens</td>
<td>• Segment length is 5-12 tokens</td>
</tr>
<tr>
<td>• @-tag in tweet</td>
<td>• No @-tag in tweet</td>
</tr>
<tr>
<td>• Adjective fraction of 0</td>
<td>• Adjective fraction is between 0 and 0.4</td>
</tr>
</tbody>
</table>

The direct act category, seriousness, and adjective fraction results have already been discussed. But many of the other results tie in or pattern together.
The punctuation findings tie in with the segment position finding (absence of an indirect act was positively associated with both a middle segment position as well as typically “middle” punctuation like comma, semi-colon, and colon).

The proper noun findings pattern with the @-tag findings, which is reasonable seeing as @-tags are essentially hyperlinked proper nouns.

Incorporated emoji as well as adjacent tone-type emoji pattern together, suggesting that emoji presence in general may signal the presence of an indirect act.

The segment length result is another outcome of the fact that segments that perform expressive direct acts are unlikely to perform expressive indirect acts (mentioned under variable-specific hypothesis #3 in section 4.2.2.2.1), since segments of length 1-4 have a strong association with expressive direct act (see figure 4.45), and most indirect acts are expressive.

In general, indirect acts are likely with segments that constitute the entire tweet, do not include proper nouns of any kind, include emoji in some capacity, do not perform an expressive direct act, and are original posts.

I will now revisit the conclusion made under variable-specific hypothesis #3, that when annotators were faced with expressive language in an otherwise non-expressive utterance, that expressive meaning was annotated as an indirect act. While it is debatable whether this is an accurate description of the phenomenon in general, I propose that it is at the least an internally consistent outcome within the confines of the present study.

To review, the segmentation and annotation guidelines proposed that backgrounded elements cannot perform direct illocutionary acts, but may perform indirect acts. Incidentally,
this includes instances of “forced accommodation” of a presupposition. For example, *I saw John when I was in Australia.* The subordinate clause *when...Australia* is backgrounded, and therefore cannot directly assert that the speaker went to Australia. Instead, it presupposes that the hearer already knows that the speaker went to Australia. In the event that the hearer did not already know this, they can easily accommodate the information that *The speaker went to Australia* (or, if they like, question it: *Wait, you went to Australia?*).

A speaker can set up this situation on purpose. If they want to inform a hearer of something without telling them directly, they can perform a different direct act which clearly presupposes the information that they really want to give. For example, if the speaker wanted the hearer to know that they had gone to Australia but did not want to say so directly, the result could be the utterance above, *I saw John when I was in Australia.* Such a situation is what I am referring to as “forced accommodation” of a presupposition.

Thus, by allowing backgrounded elements to be annotated for indirect acts, I expected situations of forced accommodation, situations when the speaker knew that the hearer did not in fact know the seemingly presupposed information, to appear as indirect acts. This is an appropriate analysis, I offer, because the annotation guidelines couched the presence or absence of an indirect act in terms of deniability. While the content of a forcibly accommodated presupposition is generally undeniable, the fact that the speaker knew that it was not in the common ground is deniable. In the example above, the speaker cannot deny that they went to Australia, but they can claim that they did not know this information was new to the hearer. In
doing so, they would be felicitously denying the force of an assertive act (though not denying its content).[^118]

Returning to the use of expressive language in otherwise non-expressive utterances, I propose that such language can act as a presupposition trigger similar to the backrounded clause ...when I was in Australia. For example, consider the exchange in example 4.1:

(4.1) Steve: Did you see John while you were there?
Dan: Yeah, I saw that asshole.
Steve: Wait, you don’t like him?

Dan’s use of the word *asshole* presupposes that John has been established as a disagreeable person in the common ground. Steve’s response to Dan suggests that it is not, as his response to Dan is exactly the response that occurs when a presupposition fails.

In short, use of the expressive word *asshole* forces accommodation of the information that *Dan believes John is a disagreeable person* in the same way that the backrounded subordinate clause ...when I was in Australia forces accommodation of the information that *The speaker went to Australia*. In the corpus, I believe that situations like this led to the annotation of an expressive indirect act.

As for why expressive direct acts were unlikely to be annotated as performing expressive indirect acts, I suspect this is due to the fact that expressive meaning can build on itself. Potts points this out with regard to the repeated use of the same expressive (e.g. *Damn, I left my damn keys in the damn car* seems to convey stronger emotion than just *Damn, I left my keys in the car*, p.182), but I would offer that the use of more expressives in general conveys more emotion (e.g.

[^118]: Therein lies the difference between an indirect act and the forced accommodation of a presupposition, and why the equation of the two in this study is appropriate but not perfect. Both the force and content of an indirect act are deniable, while only the force of an accommodated presupposition is.
Damn, I left my fucking keys in the car is stronger than Damn, I left my keys in the car). In the present study, this additive effect may have made it difficult to separate a direct act and indirect act that were both expressive. Consider the exchange in 4.2.

(4.2) Steve: Did you see John there?
Dan: Yeah. I’m afraid of that asshole.

Dan’s utterance purports to convey a psychological state of fear, making it an expressive direct act under the definitions used in this project. Meanwhile, the use of the expressive word asshole conveys that Dan dislikes John. The emotions of fear and dislike are not the same, and I would argue that the use of asshole here actually forces accommodation of the same information that does its use in example 4.1. Hence, if this utterance had occurred in my project, I might have expected it to be annotated for both expressive direct and expressive indirect acts.

However, the fact that expressives can be additive blurs the boundaries a bit. Steve in exchange 4.2 might interpret Dan’s use of asshole as simply a reflection of the fact that Dan is in a heightened emotional state (due to his purported fear) instead of as a presupposition trigger. Under such an interpretation, the use of asshole is simply strengthening the expressive meaning of the direct expressive utterance, not conveying a separate one. This may have been the default assumption of expressive language in expressive direct acts in my corpus, explaining the non-co-occurrence of expressive direct and expressive indirect annotations.

4.2.3 Analysis and Discussion of Hypothesis 3

Hypothesis 3: We will see a speech act landscape for Twitter that is distinct from that of other social media and certain spoken genres.

In section 4.2.3.1, the hypothesis will be assessed by comparing the direct act distributions and other relevant statistics of the corpus with what previous researchers have
found for Twitter and other media. Statistics will be provided for both the majority-annotation version of the corpus and the full version. The data for comparison will come from the studies covered in section 2.1.2. In section 4.2.3.2, I will discuss the results of the analysis, including potential explanations for them.

4.2.3.1 Analysis

In this section, we will first explore the relevant statistics from the present study. The studies for comparison, and any caveats to those comparisons, will be provided in section 4.2.3.1.2. Finally, section 4.2.3.1.3 will compare all of the studies in terms of speech act category ranking and proportion.

4.2.3.1.1 Present Study

Because none of the other studies pursued direct/indirect distinctions in their data, it is not immediately clear exactly which speech act distributions from the present study should be used for comparison. There are three possibilities: 1) the distributions of non-automatic direct acts only, 2) the distributions of the “primary goal” acts (i.e. the indirect act if present, the direct act if there is no indirect act), or 3) the distributions of the categories among all non-automatic acts on the platform (i.e. the distributions when non-automatic direct and indirect acts are lumped together). All three types of distributions, for both the full version of the corpus and the majority-annotation version, are provided in this section (figures 4.56 through 4.61). As will become evident, the question of which of the three sets of figures to use for comparison is somewhat trivialized by the similarity among the three sets (comparison with any one set will yield approximately similar results as would comparison with another). The biggest differences among the sets are in the relative amounts of assertive and expressive acts.
As for which set should be used from a theoretical perspective, the set that counts the “primary goal” acts seems to be the best choice. One of the original motivations for asking annotators whether a given direct act was meant “seriously” was to assess the question of whether indirect acts and direct acts could both be performed at the same time (see section 1.1.3). If the corpus exhibited a correlation between the presence of an indirect act and the designation of the direct act as “non-serious,” this could suggest that direct acts in the presence of indirect acts are not actually performed, but are instead mere conduits for those indirect acts.

As it turns out, the majority-annotation version of the corpus does in fact show a significant (chi-square test of independence p=2.72e-05) correlation between the presence of an indirect act and a designation of “non-serious” for the direct act (as was discussed and explored in section 4.2.2).

Figures 4.56 and 4.57 provide the distributions of non-automatic direct acts among those segments that perform a direct act.\textsuperscript{119}

\textsuperscript{119} Any segment that does not perform a non-automatic direct act but has its own row in the corpus was filtered out, including: backgrounded segments, quotations, hashtags, unincorporated @-tags, and tone-type emoji.
Figure 4.56. Percentage of direct acts of each category, full version of corpus.
Figure 4.57. Percentage of direct acts of each category, majority-annotation version of corpus.

The relative distributions are approximately the same for both versions of the corpus. Most of the acts are assertive (~66%), with expressive as a distant second (~15%), directive comes in at third (~8%), interrogative fourth (~7%), commissive fifth (~6%), and exercitive last (~0.1%).¹²⁰

Figures 4.58 and 4.59 provide the distributions of the primary goal acts among the segments that perform non-automatic direct acts.

¹²⁰ Exact numbers taken from figure 4.57.
Figure 4.58. Percentage of highest acts of each category, full version of corpus.
Figure 4.59. Percentage of highest acts of each category, majority-annotation version of corpus.

The biggest difference between figures 4.56/4.57 and 4.58/4.59 is in the relative proportions of assertive and expressive acts, with a relative decrease in assertive and increase in expressive in the latter pair. The increase in expressive is especially notable for the full version of the corpus (figure 4.58), as the distribution there is almost twice that of the previous set of figures (~28% vs. ~15%).

Figures 4.60 and 4.61 provide the distributions of the categories among all non-automatic acts on the platform (i.e. the distributions when non-automatic direct and indirect acts are lumped together).
Figure 4.60. Percentage of all non-automatic acts of each category, full version of corpus.
The distributions in figures 4.60 and 4.61 are more similar to those for the primary goal acts than to those for the non-automatic direct acts alone (i.e. more similar to figures 4.58 and 4.59 than to 4.56 and 4.57). This is interesting because it suggests that direct assertive acts with indirect expressive acts, and direct expressive acts with indirect assertive acts may occur at approximately equal rates.

Other than a few adjustments in the relative proportions of assertive and expressive acts, figures 4.56 through 4.61 are quite similar, maintaining the same act rankings across all six figures.
Figures 4.62 and 4.63 provide the proportion of direct acts of each category that was judged to be non-serious, for the full and majority-annotation versions respectively.

Figure 4.62. Percentage of non-serious direct acts of each category, full version of corpus.
While the change in exercitive percentages between the two figures seems significant at first glance, they are based on very few data points. As mentioned in chapter 3, only one segment received a majority-annotation of exercitive, and one other was designated as exercitive during the resolution of the three-way annotation splits. Additionally, as discussed in section 3.2.3, it seems likely that the majority of the exercitive annotations in the full version are erroneous, either in the sense that an annotator accidentally selected the wrong category or that
they simply did not fully understand the category’s description. This means that the exercitive bar in figure 4.62 is based on 13/16 likely erroneous annotations.

In addition to the exercitive category, each of the other categories drops at least a little bit in the majority-annotation version. The relative order of those stays the same, though: 1) directive, 2) commissive, 3) interrogative, 4) expressive, and 5) assertive. The fact that directives were the most likely to be judged non-serious is interesting, though not surprising. Directive acts, as defined by Alston (2000), require the speaker to take responsibility for the fact that they have the authority to give the directive in question. As we will see in section 4.2.3.2 with regard to exercitive acts, “authority” and the medium of Twitter are inherently somewhat incongruous.

When put all together, 20.4% of the non-automatic direct acts in the corpus were judged to be non-serious, while that number drops to 14.6% for the majority-annotation version. In terms of the spread of those acts, in the full version of the corpus, 52.3% of tweets had at least one direct act annotation of non-serious. In the majority-annotation version, 20% of tweets have at least one non-serious direct act.

4.2.3.1.2 Data for Comparison

Twitter

In the course of building a Twitter speech act classifier, Vosoughi & Roy (2016) collected tweets in certain topic types and had three annotators independently annotate them. The majority annotations were then used as the training data for the classifier. Direct comparison between Vosoughi & Roy’s annotation results and those of the present study is not possible, due to some fundamental differences in methodology. Vosoughi & Roy classified whole tweets (as opposed
to speech act units), they made no direct/indirect distinctions, they collected and annotated tweets within specific topics, and their taxonomy is slightly different (no commissive or exercitive, but with “request” and “other” categories). Figures 4.64 through 4.66, recreations of portions of Vosoughi & Roy’s figure 1 (p.712), show the act distributions they found for each of the three topic types they investigated: “event,” “entity,” and “long-standing.”121 Figure 4.67 averages the distributions across all three topic types.122

**Figure 4.64. Event topic type Vosoughi & Roy (2016).**

121 These topic types for Twitter were originally defined by Zhao & Jiang (2011).
122 These averages may not, of course, actually provide a good representation of what Vosoughi & Roy’s taxonomy and methodology would actually conclude about Twitter as a whole, because the nature of Twitter may be that these topic types are not equally distributed (i.e. “entity” topic tweets may not constitute a third of Twitter).
Figure 4.65. Entity topic type Vosoughi & Roy (2016).
Figure 4.66. Long-standing topic type Vosoughi & Roy (2016).
Figure 4.67. Averages across topic types Vosoughi & Roy (2016).

Vosoughi & Roy’s “recommendation” has here been relabelled as “directive,” to more directly compare with the present study. “Question” has also been relabelled as “interrogative.” Their category of “request” has been left unaltered, as requests may be either directly directive or directly interrogative. “Other” has also been left unaltered, as its act composition is unknown.

Zhang et al. (2011) and (2012) also manually annotated a corpus of tweets for training data in the course of building an automatic classifier. They do not report whether they used multiple annotators, but the lack of an agreement score report may suggest they did not.
tweets with those of the present study (no segmentation of the tweets, no direct/indirect distinctions, and a different taxonomy, which lumps commissive and exercitive into “miscellaneous”). Figures 4.68 through 4.70, recreations of portions of figure 2 from their (2012) paper (p.652), show the act distributions they found for one topic within each of the topic types. The particular topics chosen to recreate were chosen because they are also the topics whose data the authors offer as examples in their (2011) paper. The caveat about the distribution of topic types (footnote 122) also applies to the averaging of Zhang et al.’s data.

Figure 4.68. Event topic type Zhang et al. (2011, 2012).
Figure 4.69. Entity topic type Zhang et al. (2011, 2012).
Figure 4.70. Long-standing topic type Zhang et al. (2011, 2012).
Zhang et al.’s “statement” has here been relabelled “assertive,” “question” has been relabelled “interrogative,” “suggestion” has been relabelled “directive,” “comment” has been relabelled “expressive,” and the “miscellaneous” category was left unaltered, as it is composed of both commissive and exercitive acts (per their explanation of categories on p. 86).

**Facebook**

Carr et al. (2012) investigated the different speech acts, and the use of humor, in Facebook status messages. Two raters independently segmented each status into speech act units (“punctuation or propositional units”), coded each unit according to their taxonomy, and coded
the unit for the use of humor. Disagreements were discussed and resolved on an individual basis, and a third rater was called to weigh in only when necessary (p.184).

In addition to the inherent difference of media (Twitter vs. Facebook), Carr et al. do not pursue any direct/indirect distinctions, and the taxonomy is slightly different (no interrogative category, and exercitive split into “effective” and “verdictive”). Figure 4.72, a reproduction of Carr et al.’s figure 1 (p.186), shows the act distributions they found. It should be noted that the bars here do not correspond to overall distributions of acts in their data, but to the “frequency of speech acts within Facebook status messages,” i.e. the “means of speech acts within status messages per person.” So instead of distributions over all of the data, this is the average distribution per post.

![Figure 4.72. Frequency of speech acts in Carr et al. (2012).](image)
As Carr et al. use Searle’s (1979) meaning of “directive,” which lumps directive and interrogative acts together, their “directive” category has here been relabelled “directive/interrogative”. “Effective” and “verdictive” have been lumped together as “exercitive.”

Additionally, Carr et al. found that 21% of status messages contained “some modicum of humor,” (p.185).

**Online Forum**

Qadir & Riloff (2011) investigated speech acts on an online discussion forum.\(^{126}\) Once again, these authors manually annotated a set of message board posts in order to develop training data for an automatic classifier. Two independent annotators categorized each sentence in 50 posts according to speech act, discussed and resolved their disagreements, and then each annotated an additional 50 posts. They could indicate one or more acts for a single sentence, or no acts at all.

In addition to the inherent difference of media (Twitter vs. forum), the authors make no direct/indirect distinctions, use a slightly different taxonomy (no interrogative or exercitive categories), and do not consider basic assertions of fact to perform speech acts (only explicit assertions of belief in the truth of a proposition). Figure 4.73, a recreation of their table 2 (p.754), provides the act distributions they found.

\(^{126}\) Jeong et al. (2009) and Feng et al. (2006) also investigated online discussion forums, but they used taxonomies of dialogue acts that translate poorly into something comparable to the speech act taxonomy of the present study.
Figure 4.73. Speech act distributions in Qadir & Riloff (2011).

Qadir & Riloff’s “representative” has here been relabelled “assertive,” while their “directive” has been relabelled “directive/interrogative,” as they used Searle’s original definition of “directive.” Their “none” category has been left unaltered, but is undoubtedly composed of many units that the present study would have classified as assertive (given Q&R’s view that basic statements of fact are not assertions, and in fact do not perform speech acts).

IM Away Message

Nastri et al. (2006) investigated the distribution of speech acts in the away messages of AOL instant messenger users. The authors use the same speech act taxonomy used by Carr et al. (2012) for Facebook status messages, as well as the same definition of speech act unit.
Two independent raters segmented the messages into speech act units, categorized them according to speech act, and coded for humor. The authors do not indicate how disagreements were resolved.

In addition to the inherent difference of media (Twitter vs. instant message away messages) and the taxonomic differences, Nastri et al. do not pursue any direct/indirect distinctions. The authors do indicate that participants completed a “conversation indirectness scale”, but the results are not reported in their study (p.1033). Figure 4.74, a recreation of their figure 1 (p.1035), gives the average act distributions they found across all participants in the study. That is, participants’ individual act distributions were calculated for all of their messages first, and then those distributions were averaged across all participants. Thus, the numbers here correspond to the average distributions per message, not the distributions over all the data together (similar to the distributions provided by Carr et al.).

127 More accurately, Carr et al. used Nastri et al.’s taxonomy and methodology.
As Nastri et al. use Searle’s (1979) meaning of “directive,” which lumps directive and interrogative acts together, their “directive” category has here been relabelled “directive/interrogative”. “Effective” and “verdictive” have been lumped together as “exercitive.”

Additionally, Nastri et al. found that 16% of away messages contained “some element of humor,” (p.1036).

**Spoken Data**

Jurafsky et al. (1997) investigated speech acts in spoken conversations, using the Switchboard (SWBD) corpus. In the course of developing an automatic discourse act tagger,
they labelled 205,000 utterance units (from 1155 conversations) with one of 42 discourse act
tags. They provide agreement scores for this annotation, but do not report how many annotators
worked on each utterance and how differences were resolved.

Although the 42 dialogue act taxonomy is seemingly quite different from the speech act
taxonomy of the present study, Twitchell & Nunamaker (2004) have categorized the 42
SWBD-DAMSL tags into Searle’s (1979) speech act taxonomy (p.1717). Using their
categorizations, we can more readily compare the speech act distributions that Jurafsky et al.
found in their hand-tagged training data with those of the present study.

In addition to the differences of taxonomy and media (Twitter/CMC vs. spoken), the
authors did not pursue any direct/indirect distinctions. They provide the speech act distributions
for the top 18 most frequent tags, which cover 94% of their data. Figure 4.75 gives those
distributions, from their table 2 (p.1720), using Twitchell & Nunamaker’s translation to Searle’s
taxonomy.
Even though Searle’s definition of “directive” includes both of what the present study has termed “directive” and “interrogative,” only “question” directive acts appeared in Jurafsky et al.’s top 18 tags. Thus, what Searle (and Twitchell & Nunamaker) would have labelled “directive” has here been labelled “interrogative.”

Shriberg et al. (2004) developed the Meeting Recorder Dialog Act (MRDA) corpus, a corpus of 180,000 utterance units taken from recordings of spoken meetings, manually annotated for dialogue acts. They used the Switchboard-DAMSL dialogue act tag set as a base, but changed the meaning of some tags, and added a few tags particular to the setting of “meetings.” The authors also divided the tags into “general” and “specific.” Each utterance unit had to
receive one (and only one) general tag, but could also receive any number of (additional) specific
tags.

In addition to the differences of taxonomy and media (Twitter/CMC vs. spoken
meetings), the authors did not pursue any direct/indirect distinctions. Using the distributions of
the general tags, and the help of Twitchell & Nunamaker’s SWBD-DAMSL/Searle translation,
we can make a basic comparison between Shriberg et al.’s data and that of the present study.
Figure 4.76 gives those distributions, from their table 4 (p.100), using Twitchell & Nunamaker’s
translation to Searle’s taxonomy.

![Figure 4.76. Speech act distributions in Shriberg et al. (2004).](image)

As with Jurafsky et al.’s data, of the possible “directive” tags (according to Twitchell &
Nunamaker), only “question” tags appear in Shriberg et al.’s general tag list. “Directive” has
thus been relabelled as “interrogative.” The “other” category here is composed of “floor-holder” and “floor-grabber” acts. It is worth noting that the “assertive” category here includes the “rhetorical question” tag, per Twitchell & Nunamaker’s designations. Rhetorical questions, in the present study, were generally considered directly interrogative, but with an indirect act. This is just one particularly salient example of how the direct/indirect distinctions of the present study affect its comparison to these other studies.

4.2.3.1.3 Comparison

It is worth reiterating here that the comparisons made in this section are inherently reductive. The different taxonomies used by the various studies have been relabelled to fit the taxonomy of the present study as much as possible, but using the same label does not, of course, mean that the same classification method was used to classify those acts. This is especially obvious given the attempt by the present study to separate direct and indirect acts, a task outside the scope of the other studies.

Given that caveat, table 4.32 consolidates figures 4.56 through 4.61, and 4.64 through 4.76, in terms of the rankings of the various categories.
Table 4.32. Category rankings by paper/subset.

<table>
<thead>
<tr>
<th>Medium</th>
<th>Paper/Subset</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Present study(all subsets)</td>
<td>Assertive</td>
<td>Directive</td>
<td>Interrogative</td>
<td>Commissive</td>
<td>Exercitiv</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V&amp;R event</td>
<td>Assertive</td>
<td>Expressive</td>
<td>Interrogative</td>
<td>Directive</td>
<td>Other</td>
<td>Request</td>
<td></td>
</tr>
<tr>
<td>V&amp;R entity</td>
<td>Expressive</td>
<td>Assertive</td>
<td>Interrogative</td>
<td>Directive, Other, Request (tied)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V&amp;R long-standing</td>
<td>Expressive</td>
<td>Directive</td>
<td>Assertive</td>
<td>Other</td>
<td>Request</td>
<td>Interrogative</td>
<td></td>
</tr>
<tr>
<td>V&amp;R average</td>
<td>Expressive</td>
<td>Assertive</td>
<td>Directive</td>
<td>Interrogative</td>
<td>Other</td>
<td>Request</td>
<td></td>
</tr>
<tr>
<td>Twitter</td>
<td>Zhang et al. event</td>
<td>Assertive</td>
<td>Directive</td>
<td>Expressive</td>
<td>Miscellaneous</td>
<td>Interrogative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhang et al. entity</td>
<td>Expressive</td>
<td>Miscellaneous</td>
<td>Assertive</td>
<td>Interrogative</td>
<td>Directive</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhang et al. long-standing</td>
<td>Expressive</td>
<td>Miscellaneous</td>
<td>Assertive</td>
<td>Directive</td>
<td>Interrogative</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Zhang et al. average</td>
<td>Expressive</td>
<td>Miscellaneous</td>
<td>Assertive</td>
<td>Directive</td>
<td>Interrogative</td>
<td></td>
</tr>
<tr>
<td>Facebook Status</td>
<td>Carr et al.</td>
<td>Expressive</td>
<td>Assertive</td>
<td>Commissive</td>
<td>Directive/Interrogative</td>
<td>Exercitiv</td>
<td>(0)</td>
</tr>
</tbody>
</table>
Table 4.32 (Cont.)

<table>
<thead>
<tr>
<th>Medium</th>
<th>Paper/Subset</th>
<th>1st</th>
<th>2nd</th>
<th>3rd</th>
<th>4th</th>
<th>5th</th>
<th>6th</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discussion Forum</td>
<td>Q&amp;R</td>
<td>None</td>
<td>Directive / Interrogative</td>
<td>Expressive</td>
<td>Assertive</td>
<td>Commissive</td>
<td></td>
</tr>
<tr>
<td>IM Away Message</td>
<td>Nastri et al.</td>
<td>Assertive</td>
<td>Expressive</td>
<td>Commissive</td>
<td>Directive / Interrogative</td>
<td>Exercitiv e (0)</td>
<td></td>
</tr>
<tr>
<td>Spoken</td>
<td>Jurafsky et al.</td>
<td>Expressive</td>
<td>Assertive</td>
<td>Interrogative</td>
<td>Other</td>
<td>Commissive (0)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shriberg et al.</td>
<td>Assertive</td>
<td>Expressive</td>
<td>Other</td>
<td>Interrogative</td>
<td>Commissive (0)</td>
<td></td>
</tr>
</tbody>
</table>

Within the other Twitter studies, only the two “event” topic type subsets share the same top-ranked category with the present study (assertive). However, 5 of the 8 Twitter data subsets share the same top two categories as the present study, if we disregard order (assertive and expressive). If we focus on the two subsets that average values across topic type (V&R average and Zhang et al. average), both show expressive in first place, and assertive in second.

These results certainly seem to confirm that assertive and expressive acts are the most common acts on Twitter, even with variations in the exact classifications of these categories.

The fact that the present study had a top-ranked category of assertive, while both of the averages across topic type from the other studies ranked expressive first, may suggest that my corpus (and perhaps, by extension, Twitter as a whole) is composed of more tweets that would be classified...
under the “event” topic type than the other two topic types. This possibility will be revisited in section 4.2.3.2.

As for the other media, almost all of them also exhibit top two rankings for assertive and expressive (if we disregard order). Qadir & Riloff’s study of discussion forums is the exception, though it is worth reiterating that many (if not all) segments in their “none” category would likely have been categorized as “assertive” in the present study (as Qadir & Riloff did not consider general statements of fact to perform speech acts). The high ranking of directive/interrogative for Q&R’s data is interesting, though perhaps unsurprising given that their data comes from a veterinary medicine forum. It appears that, beyond “statements of fact”, much of the forum revolves around asking questions (as opposed to expressing oneself or asserting beliefs).

Summarily, according to table 4.32, the media/topic types with the assertive category ranked first are: 1) the present study (Twitter without a topic focus), 2) Twitter “event” topic type, 3) IM away messages, 4) spoken dialogue in the context of meetings, and, possibly, 5) discussion forums. All the other media/topic types exhibit a top ranking of expressive: 1) Twitter “entity” topic type, 2) Twitter “long-standing” topic type, 3) Twitter topic type average, 4) Facebook statuses, and 5) spoken telephone conversations. What this difference may mean, especially as regards the difference between Twitter and Facebook, will be revisited in section 4.2.3.2.

The following six figures, 4.77 through 4.82, compare the actual percentages of each category used in the present study with those of each paper/subset that used that category. Carr et al. (2012) and Nastri et al. (2006) both reported their results in terms of proportions of the
average post as opposed to distributions among all of the data. The following figures, which represent those proportions as though they do reflect distributions among all of the data (for the sake of comparison), are therefore, again, inherently reductive.

Figure 4.77 provides the percentage of total acts classified as assertive for each paper/subset that included the category.

![Percentage assertive, by paper/subset and medium.](image)

**Figure 4.77. Percentage assertive, by paper/subset and medium.**

Figure 4.77 shows that IM away messages and spoken conversation (in a meeting context), the two non-Twitter media that are akin to the present study in exhibiting assertive as the most common category, have the highest counts of the whole figure. That is, they exhibit higher percentages of assertive than do any of the subsets of the present study. If Qadir &
Riloff’s “none” category was added to their count here, however, they would have the highest assertive distribution.

Figure 4.78 provides the percentage of total acts classified as commissive for each paper/subset that included the category.

![Percentage Commisive, by Paper/Subset and Medium](image)

**Figure 4.78. Percentage commissive, by paper/subset and medium.**

Figure 4.78 reveals an interesting disparity. Carr et al. and Nastri et al. are the two studies that ranked commissive the highest (both have the category ranked third), yet there is a large difference in what that means in terms of raw percentage. As for why away messages
would have such a large number of commissives, the explanation is likely in the name. “Away” messages are ostensibly meant to provide information about (including reasons for) the user’s absence (e.g. “bars all night”), and perhaps forecast their eventual return (e.g. “be back at 5”).

Figure 4.79 provides the percentage of total acts classified as directive for each paper/subset that included the category.

![Bar chart showing percentage of total acts classified as directive for each paper/subset and medium.](image)

**Figure 4.79. Percentage directive, by paper/subset and medium.**

As figure 4.79 shows, the other Twitter studies were the only ones to designate a directive category akin to that of the present study. The percentage of directive acts varied quite a bit for each of the topic types, in both Vosoughi & Roy’s and Zhang et al.’s studies, but their topic type averages are almost identical (9.7 and 9.5), slightly higher than the counts for the present study.

---

Figure 4.80 provides the percentage of total acts classified as exercitive for each paper/subset that included the category.

As figure 4.80 shows, Carr et al. and Nastri et al. were the only other studies to even consider looking for exercitive acts. They found none. This absence will be revisited in section 4.2.3.2.

Figure 4.81 provides the percentage of total acts classified as expressive for each paper/subset that included the category.
Figure 4.81. Percentage expressive, by paper/subset and medium.

Figure 4.81 is where the largest differences in the subsets of the present study can be seen, promoting comparison of different subsets with different outside studies. Considering direct acts only, the expressive rate of the present study is around ~15%, closest to that of Zhang et al. event, IM away messages and spoken conversations (in a meeting context). Meanwhile, the expressive rate when considering the primary goal acts only is closest to that of Facebook statuses for the full version of the corpus (~28% vs. ~27%), and halfway between Facebook status and IM away message/spoken conversation (meeting) for the majority-annotation version of the corpus.
Figure 4.82 provides the percentage of total acts classified as interrogative for each paper/subset that included the category.

![Figure 4.82. Percentage interrogative, by paper/subset and medium.](image)

The scale of figure 4.82 (only up to ~8%) makes the values here appear to vary quite a bit, but most of the values are within ~3% of each other. Zhang et al. event and Zhang et al. long-standing have the lowest counts here, but they are still within ~5% of the highest counts.

Finally, Carr et al. and Nastri et al. also investigated humor in their data, finding that some level of humor could be found in 21% and 16% of posts/messages in their data, respectively. Those numbers are quite close to the percentage of tweets in the present study found to exhibit some element of non-seriousness, 20% (for the majority-annotation version).
4.2.3.2 Discussion

Given the broad nature of hypothesis 3, and the fact that none of the other studies referenced for comparison used the exact same taxonomy as the present study, it is unsurprising (and somewhat uninformative) that the hypothesis “We will see a speech act landscape for Twitter that is distinct from that of other social media and certain spoken genres” was supported. More informative (and more interesting) are the exact similarities and differences we see between the present study and the other studies in terms of speech act distributions.

First, let’s revisit the differences between the results of the present study and those of the other two studies of Twitter speech acts, Vosoughi & Roy (2016) and Zhang et al. (2011, 2012). As stated above, the fact that the present study had a top-ranked category of assertive, while both of the averages across topic type from the other studies ranked expressive first, could suggest that the present corpus (and perhaps, by extension, Twitter as a whole) is composed of more tweets that would be classified under the “event” topic type than the other two topic types.

However, Zhao & Jiang (2011), who proposed the topic types used by Vosoughi & Roy and Zhang et al., did not find such to be the case for Twitter as a whole. They found an overwhelming domination of the “long-standing” topic type (~75% of tweets), with “entity” a distant second (~15%), and “event” last (~10%) (p.15). This could mean that one (or both) of the following statements is true: 1) the present corpus happens to contain a greater proportion of “event” topic types than did Zhao & Jiang’s data, and/or 2) the classifications of speech act categories used for the present study is actually very different from those used by Vosoughi & Roy and Zhang et al., regardless of category labels.
As for the possibility that the present corpus contains a greater proportion of “event” topic types than did Zhao & Jiang’s data, it is not an unreasonable proposition. The data for the present study was collected 5 years after their study was published, and the tendency to use social media as a news source has only gone up over time (Shearer & Gottfried, 2017).

However, either way, the considerable methodological differences between the present study and those of Vosoughi & Roy and Zhang et al. have rendered this comparison highly speculative to begin with. Variation in the precise definitions of speech act categories seems trivial astride such large differences as categorizing speech act units vs. whole tweets and including an indirectness metric vs. not.

Among the studies of other media, only IM away messages and spoken conversations (in a meeting context) have the same top two categories, in the same order, as the present study (assertive and expressive). The away message result is interesting, but perhaps should have been predicted. Jack Dorsey, the creator of Twitter, appears to have been inspired by the conversational status functions of instant messaging when developing his new platform (Zappavigna, 2017, p.206). The results of the current study suggest that, if his goal was indeed to create a platform where away-message-like utterances take center stage, he may have succeeded.

The similarity of away messages diverges from the present study after the second category ranking, but Shriberg et al.’s data exhibits further similarity in the rank and count of interrogative acts. From this perspective, their data could be construed as the most similar (of the non-Twitter data) to that of the present study in terms of both act ranking and act distribution.
A possible explanation for this similarity is the “context collapse” of Twitter. This context collapse, the fact that a Twitter user’s potential audience consists of a wide variety of people of different levels of connection to the user, may mimic the make-up of meetings in a professional setting. That is, for the average meeting participant, professional meetings might consist of someone with a higher professional rank (e.g. their boss), someone with a lower professional rank (e.g. their assistant), and likely several people with the same approximate professional rank but of differing social ranks/relationships to the user. In short, Twitter users may employ the same strategies for appealing to the masses as do meeting participants, resulting in the same types of speech acts.

Turning to the comparison with the other popular social media platform assessed, Facebook, the difference in dominant speech act category is interesting. Given that Twitter and Facebook are both considered “social media platforms,” and are often referenced together in discussions of social media, we might expect their speech act compositions to be more similar. The domination of assertive acts on Twitter and expressive acts on Facebook suggests a fundamental difference between the two sites. That fundamental difference could either be in how users approach each site (a sort of intra-user variation), or in which people approach which site (a sort of inter-user variation). Most likely, both factors are relevant.

One easily identifiable difference that likely affects the speech act compositions of the two sites is the difference in post prompts, or “taglines.” On Facebook, “What’s on your mind?” appears immediately adjacent to the user input box, while on Twitter it’s “What’s happening?” These taglines “invite participation, at the same time suggesting what the normative purpose and

---

129 This term is taken from Marwick & boyd’s (2011) discussion of the phenomenon, as discussed in section 2.4.
nature of that participation is,” (Burgess, 2015, p.284). It is unsurprising that an invitation to share “what’s on your mind” would result in an expression of your psychological state, while an invitation to tell the audience “what’s happening” would result in an informative assertive act.

Another potentially contributing explanation comes from Panek et al. (2013), one of several studies to explore the relationship between social network site usage and narcissistic personality traits. While other studies have linked either general Twitter or general Facebook usage (or both) to the presence of narcissistic traits, Panek et al. link specific narcissistic traits with each site, at least among college-age users. The qualities of “superiority” (feeling that one is generally superior to others) and “exhibitionist” (a tendency to perform attention-seeking exhibitionist acts) are two of the common diagnostic traits for narcissistic personality disorder. Panek et al. found an association between Twitter use and the “superiority” trait, as well as between Facebook use and the “exhibitionist” trait.

The suggestion here is not that college-age Twitter users have superiority complexes and Facebook users are exhibitionists, but rather that each site may lend itself more easily to the fulfillment of those respective needs, which in turn may correlate to higher counts of either assertive or expressive acts. If Facebook is a place where exhibitionism is possible, this might explain the correlation with a high count of expressive acts, which reveal personal and emotional vulnerabilities. Meanwhile, if Twitter is a place where feelings of superiority can be entertained, this might explain the correlation with a high count of assertive acts, which include statements of fact that may portray the user as knowledgeable.

---

130 E.g. Hughes et al. (2012) and Davenport et al. (2014).
Declaring that Facebook is for exhibitionism and Twitter is for feelings of superiority based only on the results of the present study would be misguided. Such claims would need to be addressed by a separate dedicated study. If one were to explore this possibility and search for evidence that Facebook may be more amenable to exhibitionist behavior while Twitter may be better suited to entertaining feelings of superiority, some places to start could be the differences in user-interfaces and connection mechanics between the two platforms. Facebook profiles literally display more information about the user, with photo albums, group memberships, events, and of course a timeline of posts/statuses, which can be over 60,000 characters long.\textsuperscript{131} There are myriad opportunities and methods to share personal details and aspects of oneself within those parameters, possibly creating an environment where such (over)sharing is implicitly encouraged. Twitter profiles, on the other hand, are much more spartan, consisting mostly of a brief (optional) bio. The focus of Twitter is instead on the timeline of posts, which are also considerably shorter than Facebook posts (280 characters at the time of writing, but 140 characters at the time of data collection). Exhibiting oneself to a satisfactory level may be more difficult within those limitations.

Meanwhile, the follower/followee mechanic of Twitter may be more capable of nourishing feelings of superiority. On Facebook, “friendships” are mutual (if I am your “friend,” you are my “friend”). Twitter “following” mechanics are not inherently mutual (I may “follow” you without you “following” me back). A Twitter user with the need to feel superior to others may need to do nothing besides post tweets of interest to a wide variety of users, and then gather followers without following anyone back.

\textsuperscript{131} The exact number is 63,206 (Protalinski, 2011).
The last issue to be discussed here in section 4.2.3.2 is the (almost) complete lack of exercitive acts in the various media assessed. While not unexpected, this result is interesting in its near absoluteness. In order to actually pass judgment or effect change with one’s words, one must have the unquestioned, sometimes legalized-recognized, authority to do those things. Unquestioned authority is not possible on Twitter due to the lack of guarantee that the person behind the profile is indeed who they say they are,132 as well as to the fact that a court would most likely not consider social media posts legally-binding. In Alston’s definition of exercitive, the conventional effect associated with a given exercitive act does not need to actually happen for an utterance to be classified as an exercitive. That is, speakers are perfectly free to make exercitive utterances without possessing the prerequisite legal (or other) authority (if there is any). But without that unquestioned authority, the speaker would be assumed to be either joking or mistaken.

As figure 4.83 demonstrates, even the current president of the United States, a man with considerable power in an official capacity outside the confines of Twitter, appears to suspect that this power does not extend to his Twitter account (regardless of its verified checkmark):

---

132 There is no guarantee that the person posting a given tweet is in fact the person that the profile “belongs” to both because “hacking” is possible, and because users can hire other people to tweet for them as a form of personal brand management.
As president, Mr. Trump certainly has the authority to at least request that the Department of Justice “look into” something, and the use of “hereby” would suggest that he is attempting to invoke that authority via the tweet itself. But the addition of the phrase “and will do so officially tomorrow” clearly reveals that even he doubts that a Twitter post is an appropriate avenue through which he can legally enact his authority.

Responses to Mr. Trump’s tweet reveal that other Twitter users also found the use of “hereby” (and perhaps exercitive acts in general) incongruous with the medium, regardless of Mr. Trump’s hedge:
As stated above, a speaker is free to make exercitive utterances in the obvious absence of the prerequisite legal (or other) authority, but would be assumed to be either joking or mistaken in such a situation. The tweet in figure 4.84 is mocking Mr. Trump not for tweeting an exercitive act, but for seemingly trying to do so seriously (before doubting himself and hedging his utterance).

In short, a seriously-meant and non-mistaken exercitive is unlikely on Twitter because there are usually prerequisite conditions regarding authority that are inherently unachievable. However, non-serious (or mistaken) exercitives are theoretically perfectly possible. Why even these tend not to occur in the medium is unclear.

4.2.4 Analysis and Discussion of Hypothesis 4

Hypothesis 4: The speech act profiles of unincorporated hashtags vs. the profiles of incorporated hashtags will be distinct, suggesting that "incorporated vs. unincorporated" is pragmatically meaningful.

In section 4.2.4.1, the hypothesis will be assessed by looking for significant differences in the characteristics of hashtags that are incorporated versus those that are not. As noted in section 4.1.4, the inter-annotator agreement scores for the hashtag-specific annotation questions were poor. For this reason, characteristics assessed here will include several annotation variables, as well as several objective contextual variables. For the annotation variables, the assessment will be made using both the majority-annotation version of the corpus and the full version. The objective variables will use only the majority-annotation corpus. In section 4.2.4.2 I will discuss whether the results of the analysis actually suggest that “incorporated vs. unincorporated” is pragmatically meaningful.
4.2.4.1 Analysis

For each of the variables considered, a chi-square test of independence was performed for that variable crossed with the incorporation variable.

4.2.4.1.1 Annotation Variables

The following annotation variables were assessed, for both the majority-annotation version of the corpus and the full version:

**Used:** This variable corresponds to whether the hashtag was judged to be one which a Twitter user would actually utilize as a tag (e.g. one they would search for or click on).

**Affiliation:** This variable corresponds to whether the hashtag was judged to be signaling affiliation with a group or cause, as denoted by the content of the hashtag.

**Contribution:** This variable corresponds to whether the hashtag was judged to be a contribution to a meme or trend, as denoted by the content of the hashtag.

**EmoMeta Yes/No:** This variable corresponds to whether the hashtag was judged to be performing an emotive or meta-commentative indirect act.

**EmoMeta Choice:** This variable corresponds to the category of emotive/meta-commentative act chosen, including “none” as a choice.

**EmoMeta Yes Choice:** This variable corresponds to the category of emotive/meta-commentative act choice, excluding those hashtags for which the response was “none.”

**Direct Act Fractions:** Each of these six variables indicates what fraction of the direct acts performed in the enveloping tweet (i.e. the tweet housing the hashtag) were judged to be of the given category.
**Serious Fraction:** This variable corresponds to what fraction of the direct acts performed in the enveloping tweet were judged to be serious.

For the majority-annotation version of the corpus, the only chi-square test to yield a significant p-value ($\leq 0.05$) was for the EmoMeta Yes/No variable ($p=0.015$). Figure 4.85 provides the association plot for that variable.

![Association plot for Emotive/meta-commentative yes/no and Incorporation, majority-annotation version of corpus.](image)

As can be seen in figure 4.85, incorporated hashtags are negatively associated with the performance of an emotive/meta-commentative indirect act, while unincorporated hashtags are positively associated with such a performance.

For the full version of the corpus, the chi-square tests for four of the variables yielded significant p-values: EmoMeta Yes/No ($p=4.5e-5$), EmoMeta Choice ($p=0.0029$), Contribution
(p=0.0064), and Fraction Interrogative Direct Act (p=0.05). Figures 4.86 through 4.4.5 provide the association plots for these variables crossed with incorporation.

**Figure 4.86.** Association for Emotive/meta-commentative yes/no and Incorporation, full version of corpus.
Figures 4.86 and 4.87 show the same primary association, which is the same as that shown by figure 4.85: incorporated hashtags are negatively associated with the performance of an emo/meta indirect act, while unincorporated hashtags are positively associated with such a performance. As far as what category of emo/meta act is performed by unincorporated hashtags, assertive appears to be the most common, followed by expressive.

Figure 4.87. Association for Emotive/meta-commentative choice and Incorporation, full version of corpus.
Figure 4.88. Association for Contribution and Incorporation, full version of corpus.
Figure 4.88 indicates a negative association between incorporated hashtags and a contribution by the hashtag to a meme or trend. Meanwhile, unincorporated hashtags show a positive association with such a contribution.

Figure 4.89 shows positive associations between unincorporated hashtags and an interrogative fraction of 0 or 0.5. The bottom row provides the associations for tweets which performed no direct acts (interrogative or otherwise), which also shows a positive association with unincorporated hashtags.

4.2.4.1.2 Objective Variables

The following objective variables were assessed, using the majority-annotation version of the corpus:
**Segment Position**: This variable corresponds to the position of the hashtag within the enveloping tweet. For incorporated hashtags, “segment position” refers to the position of the segment into which the hashtag is incorporated.

**Emoji Count in Tweet**: This variable corresponds to the total number of emoji that appear in the enveloping tweet.

**Hashtag Length**: This variable corresponds to the length of the content of the hashtag (i.e. the hash symbol excluded), in characters.

**Hashtag Length Batch**: This variable corresponds to the length batch of the content of the hashtag. Batches were made in 10-character increments (e.g. 0-10, 11-20, 21-30, 31+).

Of these four variables, the only chi-square test to yield a significant p-value was for Segment Position (p=2.2e-16). Figure 4.90 provides the association plot for this variable.
Figure 4.90 indicates a positive association between unincorporated hashtags and a position at the end of the enveloping tweet. Incorporated hashtags, on the other hand, show positive associations with a position at the beginning of the enveloping tweet. Additionally, the second row from the bottom represents the segment position of “entire.” There is a negative association between unincorporated hashtags and a segment position of “entire,” which really just means that users tend not to send tweets consisting only of a single hashtag.

4.2.4.1.3 Summary

Table 4.33 summarizes the findings from 4.2.4.1.1 and 4.2.4.1.2. In all cases, the positive associations are listed in the table, with the understanding that hashtags with the other
incorporation status have a negative association with that variable. Whether the association was found using the majority-annotation version of the corpus or the full version is given in parentheses next to the association.

At this time, it is prudent to reiterate that the hashtag-specific annotation categories produced low inter-annotator agreement scores (i.e. Used, Contribution, Affiliation, and EmoMeta), as did the “serious yes/no” direct act category, so these results should be taken with a grain of salt. Meanwhile the incorporation variable itself was not annotated, but was noted by the author during data segmentation.

Table 4.33. Summary of findings from sections 4.2.4.1.1 and 4.2.4.1.2.

<table>
<thead>
<tr>
<th>Incorporation Status</th>
<th>Association</th>
</tr>
</thead>
</table>
| **Unincorporated**   | Performance of emo/meta indirect act (full, maj-anno)  
Performance of assertive or expressive emo/meta indirect act (full)  
Contribution to a meme or trend (full)  
Interrogative direct act fraction of 0 or 0.5 (full)  
Lack of direct acts in enveloping tweet (full)  
Position at end of enveloping tweet (maj-anno) |
| **Incorporated**     | Interrogative direct act fraction of 0.25, 0.33, or 1 (full)  
Position at beginning of enveloping tweet (maj-anno)  
Position as only segment of a tweet (maj-anno) |

4.2.4.2 Discussion

The associations listed in table 4.33 include several speech act performance characteristics (emo/meta indirect act, contribution indirect act, etc.), so the original hypothesis that “The speech act profiles of unincorporated hashtags vs. the profiles of incorporated hashtags will be distinct, suggesting that "incorporated vs. unincorporated" is pragmatically meaningful” appears to be supported (with the poor agreement caveat noted). The addition of
objective variables has also shown some objective differences between the contextual environments of incorporated and unincorporated hashtags, providing additional support for the hypothesis that incorporated and unincorporated hashtags are used in different ways.

We may now turn to the question of why these particular differences appear between the usages of unincorporated and incorporated hashtags, and whether “incorporated vs. unincorporated” may indeed be pragmatically meaningful.

Returning to table 4.33, unincorporated hashtags are more likely to perform emotive/meta-commentative indirect acts (of, specifically, assertive and expressive categories), a usage termed “evaluative” by Page (2012). At the same time, unincorporated hashtags are more likely to come at the end of the enveloping tweet. It makes sense that a segment performing an evaluative act would come at the end of the tweet -- you need to have said something first in order to be able to evaluate it.133

However, what is not immediately clear is whether unincorporated hashtags tend to come at the end of the tweet because they can perform these evaluative acts, or if they are chosen to perform these acts because they come at the end of the tweet. That is, does the unincorporation itself matter to the evaluative act, or only the segment position?

Given the design of the annotation task, the unincorporation certainly matters for the present study. For every question given to the annotators, an image of the encompassing thread was available to view, while the exact segment being asked about was given alongside the question (and, in the case of duplicate segments, highlighted on the image itself). For the

---

133 Scott (2015) also noted the tendency of meta-commentative hashtags to come “after the main content of the tweet...rather than integrated into the main sentence,” (p.13).
hashtag-specific questions, the exact segment provided alongside the question was only the hashtag, excluding any incorporating text.\textsuperscript{134}

Now, consider the difference between the following two hypothetical posts, the first of which is an example of a meta-commentative hashtag provided by Wikstrom (2014, p.140):

\begin{center}
\begin{itemize}
\item \textbf{(4.3)} Fo Reals, it is blizzarding. \#AndiLikeIt
\item \textbf{(4.4)} Fo Reals, it is blizzarding. And i \#LikeIt
\end{itemize}
\end{center}

During the annotation task, for post \textbf{4.3}, the question of \textit{Does this hashtag perform an emotive or meta-commentative indirect act?} would have been asked of \#AndiLikeIt, while that question for post \textbf{4.4} would have been asked of just \#LikeIt. Additionally, in the case of post \textbf{4.4}, before any questions about the hashtag specifically, the annotator would have been asked to categorize the direct act of the segment \textit{And i \#LikeIt}. It is possible, if not likely, that an annotator who has just categorized the direct act of the segment \textit{And i \#LikeIt} would not also attribute an evaluative act to the hashtag \#LikeIt by itself.

Consider also hypothetical posts \textbf{4.5} and \textbf{4.6}, wherein the hashtag itself is exactly the same in both cases:

\begin{center}
\begin{itemize}
\item \textbf{(4.5)} Can’t find my keys again. \#ridiculous
\item \textbf{(4.6)} Can’t find my keys again. This is \#ridiculous
\end{itemize}
\end{center}

The hashtag in post \textbf{4.5} seems to mean, by itself, approximately what the second segment of post \textbf{4.6} states explicitly: “this is ridiculous.” Thus, an annotator given the unincorporated hashtag in post \textbf{4.5} would likely say that it performed an evaluative act. But an annotator given the incorporated hashtag in post \textbf{4.6} -- having just been asked to categorize the direct act of \textit{This is

\footnote{134 The incorporating segment itself was annotated in a separate question.}
#ridiculous -- would likely not say that the hashtag performs a (seemingly additional) evaluative act on its own.

In short, the unincorporation may be no more important for the performance of evaluative acts than is the semantic completeness of the proposition. Unincorporated hashtags are interpreted as semantically complete by themselves (with gaps filled in mentally as needed), while incorporated hashtags are interpreted as semantically part of another segment. A single word or phrase can perform an evaluative function by itself (e.g. #ridiculous), but if that word is part of a larger, explicit segment (e.g. this is #ridiculous), the performance of that evaluative act shifts from that word alone to the entire segment. Thus, when considering only the hashtags themselves, unincorporated hashtags may perform more evaluative acts than incorporated hashtags simply because they are interpreted as semantically complete.

The other primary hashtag-specific annotation finding from table 4.33 is that unincorporated hashtags tend to be perceived as signalling that the tweet is a contribution to a meme or trend more often than do incorporated hashtags. At the least, this finding supports the proposal that there is a “template that must be adhered to for successful participation” in a meme/trend (Wikstrom, 2014, p.135). But it may also suggest that such a template more often calls for an unincorporated hashtag, as opposed to an incorporated one. That is, the template may more often call for adding the relevant hashtag, unincorporated, to the end (most likely) of the tweet, as opposed to working it into a larger utterance.

The findings regarding interrogative direct act fractions are somewhat odd, in the sense that there is not a very clear and interpretable pattern to the results. Incorporated hashtags are associated with non-zero fractions of interrogative direct acts, except for 0.5. Even if we write
that exception off as noise, the fact that incorporated hashtags are more likely when there are higher proportions of interrogative acts in the tweet does not immediately lend itself to explanation. One possibility is that when questions and a hashtag appear in the same tweet, the questions tend to be about the hashtag (or the group/cause/trend that the hashtag references), meaning that it would probably be incorporated.

The finding that unincorporated hashtags are more likely in the case that the enveloping tweet does not perform any direct acts is somewhat unexpected, given that unincorporated hashtags are unlikely to be tweeted out by themselves (as will be reiterated next). It must therefore be the case that unincorporated hashtags are more likely when a tweet consists of only backgrounded segments or quotations, units that do not perform non-automatic direct acts. The question then becomes why quotations or backgrounded elements would be more likely to be accompanied by an unincorporated hashtag than to include an incorporated one.

For quotations, it may be because the content is fixed, not flexible. You cannot change a word in a direct quote so that it fits the hashtag you want to tweet, so you would likely just add the hashtag unincorporated at the end. As for backgrounded segments, it may simply be that it is difficult to background something that is bright blue and underlined when the rest of the text is black and not underlined. That is, hashtags may seem inherently at-issue due to their automatic orthographic emphasis, and this at-issueness may make them seem out of place in a backgrounded segment.

Finally, as mentioned under figure 4.90, the fact that incorporated hashtags are more likely to appear in the “entire” position just means that users are unlikely to post a tweet consisting of only a single (unincorporated) hashtag. Given the various functions of hashtags
proposed in the literature, this result is perhaps not surprising. All of those functions in some way invoke or refer to the other (i.e. non-hashtag) content of the tweet, with the exception of the affiliation function (and, thereby, the indexing function of affiliating hashtags). The basic tagging/indexing function of hashtags categorizes the tweet’s content according to topic, the contribution function marks the tweet’s content as a contribution to a meme/trend, and the evaluative function evaluates the tweet’s content -- all of which would be meaningless (or impossible) in the absence of any other text in the tweet.

However, as noted above, the affiliation function is an exception. A hashtag meant to signal affiliation with a group or cause does not need to refer to any other text to accomplish that function, and so could theoretically appear alone in a tweet. For example, a tweet consisting of only #BlackLivesMatter can signal the poster’s affiliation with or support for the Black Lives Matter movement without the need for any other (non-hashtag) text.

It is possible that such tweets, though theoretically possible, simply did not occur in the corpus. It is also possible that signalling affiliation with a community without contributing to the discourse around that community is generally discouraged. Especially when the community itself is or has been actually constructed online, there may be an implicit rule that one must help to construct that community if one is going to signal affiliation with it. In Zappavigna’s (2011) analysis of affiliating hashtags, she emphasizes that they “render searchable the coupling that occurs in the tweet” (p.801). This coupling is the connection between the community represented by the tag and the values that the community is built around (as conveyed through the non-hashtag content of the tweet). This would explain why such a hashtag is unlikely to appear without any non-hashtag content.
4.2.5 Analysis and Discussion of Hypothesis 5

**Hypothesis 5:** There is a distinct communicative style to Twitter, and we will be able to see evidence for it in the corpus. If we look at how the categories compare against how long the original poster has been on Twitter, we'll see a convergence on a certain style (e.g. more indirectness over time, more expressives, more emoji, more emotive hashtags, or perhaps just more of the structures that are not used in spoken conversation).

This hypothesis will be assessed by looking for trends in the behavior of Twitter users over time. For this purpose, “time” is defined as the number of days between the join date of the user and the date of the post in question. Because the date that a user joined Twitter is not necessarily the date that the person became an active user of the site, trends in behavior will also be assessed over the number of total tweets the poster has ever sent. The assumption here is that the number of days since joining and the number of total tweets ever sent are proxies for a user’s fluency in the communicative norms of the site. User information (such as join date and total tweets sent) was only gathered for the “original post,” the first tweet in a conversation. Therefore, only segments from those tweets (2681 out of 3631 segments) were considered in the search for trends over time/total tweet count.

In addition to looking for trends in the behavior of the original posters, indications of style and communicative norms will also be sought in the behavior of those who responded to the original post in some way. That is, when corrected for the number of followers the original poster has, which characteristics of an original post correlate with an increased likelihood to receive a like, retweet, or reply post from another user? Additionally, if a post tags another user, how often does that user respond in some way (like/retweet/reply post)? Can we identify
contextual differences between @-tags that are responded to by the relevant user and ones that are not? These questions and others will be addressed in the analysis and discussion to follow.

4.2.5.1 Analysis

As mentioned in section 4.1.4, the categories whose behavior will be assessed over time/total tweets include the majority of the annotation categories (e.g. direct force, indirect act presence/force, etc.) as well as a variety of objective categories (e.g. number of emoji, number of hashtags, etc.). Annotation categories with poor agreement will be noted in the summary of findings.

4.2.5.1.1 Trends in the Behavior of Original Posters

Annotation Categories Over Time/Total Tweets

The following annotation categories were assessed over time/total tweets sent: direct act category, indirect act presence, indirect act category, “tagging” hashtag usage, affiliating hashtag usage, contributing hashtag usage, and emotive/meta-commentative hashtag usage.

Let’s begin with direct act category, specifically with whether and how the distribution of the various forces changes over time/total tweets. First, all segments in the original posts that performed a non-automatic direct act were ordered by how many days the poster had been a Twitter user before posting the tweet that appeared in the corpus. The ordered segments were then divided into batches of 100 segments each, and the percentage of each batch that was a given category was calculated and plotted over time. This same process was then repeated, but with ordering the segments by how many tweets the poster had ever posted on the site. The resultant plots are given below, as figures 4.91 and 4.92.
Figure 4.91. Distribution of direct act category over time, batches by number of observations.
Figure 4.92. Distribution of direct act category over total tweets sent, batches by number of observations.

While none of the trends on either plot are particularly robust, both plots do show a slight decrease in the percentage of directive acts over time/total tweets, as well as a slight increase in the percentage of expressive acts over time/total tweets.

For indirect act presence, all segments with a potential indirect act were coded as either receiving an indirect act annotation (1) or not (0). The results were then plotted once over time and once over total tweets. The resultant plots showed no trends, suggesting that neither the presence nor absence of an indirect act typifies communication on Twitter.

In addition to no trend in the overall presence of indirect acts over time/total tweets, neither does there appear to be a trend in the force of the indirect acts that do occur. As was done
for the direct act category assessment, the segments with a potential indirect act were ordered by time, divided into batches of 100 observations, and the percentage of each indirect act response were calculated for each batch and plotted over time. Then the process was repeated, ordering the segments by total tweets ever sent by the original poster. The resultant plots show no strong trends.

Let’s now turn to the hashtag annotation categories. For these plots, the data was restricted to only hashtag segments that occurred in original posts, a total of 172 segments. Those segments were then plotted over time and total tweets sent for each category. The resultant plots are given, by category, below.

Before we proceed, however, a brief discussion of outliers is necessary. Due to significant outliers in terms of total tweets sent, that plot for each category below excludes segments from users whose total tweet count exceeds 75,000 (which means excluding 12 of 172 segments, or ~7% of the hashtags). Those 12 data points extend from ~75,000 total tweets sent to ~400,000 total tweets sent. This means that if we look at a plot that includes all 172 hashtag segments, 93% of the data points have x values (total tweet counts) that are less than 20% of the total domain. That is, the 12 outlying data points range over 80% of the domain. This can make a trendline of the whole plot misleading. The first 20% of the trendline is based on 93% of the data, while the remaining 80% is based on 7% of the data. This is why segments with a total tweet count over 75,000 were excluded, so that we may view a trendline that is more reliable. Consider the difference in trendlines between figures 4.94 and 4.95. The former shows all 172 hashtag segments, while the latter zooms in on the 93% of the data with a total tweet count less than 75,000.
Returning to the data, the first hashtag category is whether or not a given hashtag would be “used” (i.e. actually searched for or clicked on). While the trendlines of these plots are dynamic, there seems to be no overall trend, over time or total tweets sent, with regard to whether or not the original poster used hashtags that would actually be clicked on or searched for.

Next is the question of whether or not a given hashtag is signalling affiliation with a group. Figure 4.94, which includes all data points over total tweets sent, is only included here to form a point of comparison with figure 4.95 (the graph with total tweet counts over 75,000 excluded), for the benefit of the outlier discussion above. The analogous plot for the remaining hashtag categories will not be included.

Figure 4.93. Likelihood that hashtag is affiliating over time on Twitter.
Figure 4.94. Likelihood that hashtag is affiliating over total tweets sent.
Figure 4.95. Likelihood that hashtag is affiliating over total tweets sent, x>75,000 excluded.

Figure 4.93 does appear to show an overall decrease in affiliating hashtags over time, while the plot over total tweets sent (figure 4.95) is dynamic without an overall trend.

Next is the question of whether or not a given hashtag is contributing to a meme or trend.
Figure 4.96. Likelihood that hashtag is contributing over time on Twitter
Once again, the trendlines are very dynamic with large margins of uncertainty. There are no robust trends over time or over total tweets sent with regard to whether the hashtag is contributing to a meme or trend. Interestingly, however, layering figures 4.93/4.96 and 4.95/4.97 over each other does demonstrate one seemingly reliable trend: that hashtags tend not to be both affiliating and contributing at the same time.
Figure 4.98. Likelihood that hashtag is contributing/affiliating over time on Twitter.
Figure 4.99. Likelihood that hashtag is contributing/affiliating over total tweets sent, $x > 75,000$ excluded.

In numbers, of the 172 hashtags that occur in the original posts, 84.9% (146) of them are either affiliating or contributing, but not both. 10.5% are both, and 4.7% are neither.

Finally, the question of whether or not a given hashtag is performing an emotive/meta-commentative act, and if so, what type that act is. The first two figures here, 4.100 and 4.101, address the first half of that question (is there an emotive/meta-commentative act, yes or no), and the next two figures (4.102 and 4.103) address the question of type.
Figure 4.100. Likelihood that hashtag is emotive/meta-commentative over time on Twitter.
Figure 4.101. Likelihood that hashtag is emotive/meta-commentative over total tweets sent, x>75,000 excluded.
Figure 4.102. Hashtag emotive/meta-commentative distribution over time on Twitter.
Figure 4.103. Hashtag emotive/meta-commentative distribution over total tweets sent.

The likelihood that a given hashtag will be of an emotive/meta-commentative nature does seem to go up over time, according to figure 4.100. Interestingly, this is not true for the data viewed over total number of tweets sent. As for which type(s) of emotive/meta-commentative acts go up as the rate of “None” goes down, figure 4.102 shows a slight increase in the assertive category over time.

Objective Categories Over Time/Total Tweets

The following objective categories were assessed over time/total tweets sent: overall emoji count, hashtag presence/count, @-tag presence/count, segment length in tokens, tweet length in segments, tweet-final punctuation, use of standard capitalization, use of vulgar words, and use of personal pronouns.
Let’s start with general hashtag presence/count. As a user becomes more familiar with Twitter (through time spent or tweets sent), do they use more or less of the most notorious elements of Twitter communication? As figures 4.104 and 4.105 demonstrate, the answer appears to be generally no.

![Graph showing number of hashtags over time on Twitter](image)

**Figure 4.104. Number of hashtags over time on Twitter.**
As mentioned in the previous section, and evident in figure 4.105, a few tweets are significant outliers by the metric of “total tweets sent.” When looking at only the hashtags, 93% of the data covered only ~20% of the domain, so I zoomed in on that 20% of the domain to get a look at the trendline as based on 93% of the data (i.e. excluded segments with total tweet counts over 75,000). Now, looking at all 1017 of the original tweets, the skew of the data is even greater, with 98% of the data located in that first ~20% of the domain (i.e. total tweets sent is less than 150,000), and the remaining 2% spread out over the remainder (up to ~900,000). Hence, for the “over total tweets sent” plots here in 4.2.5.1.1.b, I have zoomed in on the 20% of the domain that contains the vast majority of the data, as it shows a more reliable trendline.
Additionally, some outliers were excluded for the y-dimension in these categories, as well. For example, as evident in figure 4.105, only 2 (out of 1017) tweets contained more than 3 hashtags. Other y-axis outliers will be noted as relevant for the following categories.

For hashtag count, the plot excluding the 21 data points with a total tweet count over 150,000 and the 2 data points with a hashtag count over 3 is shown in figure 4.106. The trendline remains mostly flat, although there is a slight initial drop, accompanied by a fairly narrow margin of uncertainty.

Figure 4.106. Number of hashtags over total tweets sent, 23 outliers removed.

The average number of hashtags in a given tweet stays mostly consistent over both time and total tweets sent. Is this also true of @-tags? Figures 4.107 and 4.108 appear to confirm that it is, but with the same curious initial drop (and accompanying narrow margin of uncertainty) when the data is viewed over total tweets sent. This drop can be seen in figure 4.108, which
excludes the 21 data points with a total tweet count over 150,000 and 2 tweets with an @-tag count over 2.

Figure 4.107. Number of @-tags over time on Twitter.
Next is the count of emoji in tweets over time and total tweets sent. No strong increase or decrease in emoji count over time or total tweets sent was observed.

Utterance length is another objective category that may reflect style accommodation over time. Utterance length was assessed in two ways: tweet length in segments, and segment length in tokens. For tweet length, the number of segments used was the total number of segments for that tweet minus any incorporated segments (e.g. incorporated hashtags or @-tags). As figure 4.109 shows, tweet length does not appear to vary greatly over time. There is a small initial drop, even if we remove the one outlier with y>10 (see figure 4.110), but the margin of uncertainty covers it.
Figure 4.109. Tweet length over time on Twitter.
Figure 4.110. Tweet length over time on Twitter, y>10 excluded.

When we look at tweet length over total tweets sent, however, there is a definite initial drop in the trendline (figure 4.111). For this plot, the one data point with more than 10 emoji was excluded, as were the 21 data points with total tweet count over 150,000.
While there may be a change in tweet length over total tweets sent, figures 4.112 and 4.113 show that the average number of tokens in a segment remains fairly consistent over both time and total tweets sent. Figure 4.113 excludes segments with $x>150,000$, as outliers.
Figure 4.112. Segment length over time on Twitter.
Figure 4.113. Segment length over total tweets sent, x>150,000 excluded.

Figure 4.112 does show a rise starting at ~2,750 days on Twitter, but this is also the exact point that the data begin to thin and the margin of error increases. It would appear that, in general, the average segment is about six tokens long, and remains so over time and total tweets sent.

Two other traditional elements of written style, punctuation and capitalization, were assessed as well. For the former, the presence of tweet-final punctuation was plotted over time and total tweets sent (figures 4.114 and 4.115).\textsuperscript{135} Once again, data points with total tweet counts over 150,000 were excluded from the plot over total tweets sent.

\textsuperscript{135} Only tweet-final punctuation was assessed because assessing segment-final punctuation would have required a judgment for each segment of whether some sort of segment-final punctuation would be expected/appropriate according to traditional rules of written punctuation.
Figure 4.114. Presence of tweet-final punctuation over time.

Figure 4.115. Presence of tweet-final punctuation over total tweets sent, x>150,000 excluded.
Both figures show an initial drop before an increase. This is particularly pronounced in figure 4.115.

Likewise, the likelihood that a given segment will exhibit standard capitalization practices stays fairly consistent over both time and total tweets sent, as figures 4.116 and 4.117 show. For these purposes, standard capitalization meant three things: 1) proper nouns were capitalized, 2) tweet-initial words and any word following a period, exclamation point, or question mark were capitalized, and 3) no words were in all capital letters.

Figure 4.116. Standard capitalization over time on Twitter.
Figure 4.117. Standard capitalization over total tweet sent, x>150,000 excluded.

There is a slight increase in standard capitalization over time, but that is followed by a decrease with a wide margin of uncertainty as the data begin to thin out. Meanwhile, the data over total tweets sent shows many ups and downs, but no overall trend.\textsuperscript{136}

Finally, let’s look at some lexical choices over time and total tweets sent. Specifically, the use of vulgar words and the various personal pronouns. Figures 4.118 and 4.119 show the average number of vulgar words in a tweet over time and total tweets sent.\textsuperscript{137} Figure 4.119 excludes the 3 data points with vulgar word counts over 2, and those with total tweet counts over 150,000.

\textsuperscript{136} Again, the final drop occurs as the data thins out dramatically and the margin of error greatly increases.

\textsuperscript{137} The vulgar word list was discussed more in section 4.2.1, but is generally composed of various permutations of the most common four-letter words.
Figure 4.118. Number of vulgar words over time on Twitter.

Figure 4.119. Vulgar words count over total tweets sent, y>2 and x>150,000 excluded.
The average number of vulgar words per tweet stays remarkably consistent over time, and shows only small changes over total tweets sent. Figure 4.119 shows a slight increase in vulgar word count until around 50,000 tweets sent, at which point the trendline decreases and the margin of uncertainty begins to widen.

Finally, the use of personal pronouns was assessed over time and total tweets sent. First person, second person, and third person pronoun counts were each assessed individually, followed by an assessment of total personal pronoun count (all three together). For each set of plots that follows, the “over total tweets sent” plot excludes total tweet counts over 150,000. No strong trends were observed for either first or second person pronouns.

Figures 4.120 and 4.121 show the third person pronoun count over time and total tweets sent.

Figure 4.120. Third person pronoun count over time on Twitter.
Figure 4.121. Third person pronoun count over total tweets sent, x>150,000 excluded.

Figure 4.120 shows a general decrease in third person pronoun count over time, at least up until about 1200 days on Twitter.

Finally, figures 4.122 and 4.123 show the total personal pronoun count over time and total tweets sent.
Figure 4.122. Total count of personal pronouns over time on Twitter.

Figure 4.123. Total count of personal pronouns over total tweets sent, x>150,000 excluded.
Figure 4.122 shows a consistent decrease in total personal pronoun count over the first 1000 days on Twitter, after which the count stays fairly constant. Figure 4.123 also appears to show an overall decrease in pronoun count over total tweets sent.

Summary

The results mentioned above are summarized here, organized into annotation categories and objective categories.

Annotation category results

1) Slight decrease in percentage of directive acts over time/tweets sent
2) Slight increase in percentage of expressive acts over time/tweets sent
3) No trends in presence or type of indirect act
4) No trend in likelihood that a hashtag would actually be used
5) Affiliating usage of hashtags trends slightly down over time
6) Affiliating and contributing usages of hashtags tend not to overlap
7) Likelihood that a hashtag performs an emotive/meta-commentative act trends upward over time

It should be reiterated here that the annotation categories behind #3-#7 all had low inter-annotator agreement scores. This low agreement may even be partially responsible for the lack of significant trends.

Objective category results

1) Brief initial drops in hashtag and @-tag counts over total total tweets sent
2) No robust trend in emoji count
3) Initial drop in tweet length over total tweets sent
4) Initial drop in tweet-final punctuation followed by a steady increase over total tweets sent
5) Slight increase in standard capitalization over time, followed by a decrease with a wide margin of uncertainty
6) Very slight increase in vulgar word count over total tweets sent, followed by a decrease with a wide margin of uncertainty
7) No overall trends in first or second person pronoun counts over time/tweets sent
8) Initial decrease in third person pronoun count over time
9) General decrease in total personal pronoun count over first 1000 days on Twitter, and over total tweets sent

What these trends may tell us about the communicative style of Twitter will be discussed in section 4.2.5.2.

4.2.5.1.2 Trends in the Behavior of Respondents to Original Posts

Indications of style and communicative norms were also sought in the behavior of those who responded to the original posts in some way. The main question asked in this section was therefore: what do users tend to respond to? This question was addressed from two fronts. In part A, regression analyses were used in an attempt to find which characteristics of the original posts were significant predictors of responses. In part B, for tweets that included at least one @-tag, another regression analysis was performed in search of which characteristics of the tweets appear to be significant predictors of whether the (or an) @-tagged user responded to the tweet.

Of course, in order for a user to respond to a tweet, they have to first see the tweet. The issue of tweet visibility is non-trivial to this analysis. An original poster with 50,000 followers is probably more likely to receive responses than is a poster with 2 followers because the tweets of
the former are showing up on 49,998 more Twitter feeds. Hence, there is a good chance that the number of followers that a given original poster has is a confounding variable in the analysis of responses.

To assess whether this was indeed the case, I plotted the instances of each type of response over the number of followers. A positive correlation would suggest that follower number is indeed a confounding variable.

At first glance, with all data included, positive correlations seem obvious, as the next three figures show. Figure 4.124 plots the number of unique repliers to a given tweet over the follower number of the tweet’s poster. The trendline consistently rises until ~80,000 followers (before apparently decreasing due to three of the most significant outliers). Figure 4.125 plots the number of likes that the tweet received over the number of followers of the tweet’s poster. Finally, figure 4.126 plots the number of retweets the tweet received over the number of followers of the tweet’s poster.
Figure 4.124. Number of unique repliers over number of followers.
Figure 4.125. Number of likes over number of followers.
When we exclude outliers and zoom in on the most dense portions of these figures, however, we see a bit of a different story. Figure 4.127 is the same as figure 4.124, but with $x>2500$ (i.e. tweets whose poster has over 2500 followers) excluded. This exclusion leaves 90% of the data, for which any confounding effect of follower number does not appear significant. The trendline only just begins to rise at the upper end of this range. Figure 4.128 is the same as figure 4.125, but with $x>2500$ excluded. Again, for the remaining 90% of the data, the number of likes is fairly consistent, regardless of follower number. Finally, figure 4.129 is the same as figure 4.126, but with $x>2500$ excluded. The same trend appears here as in 4.127 and 4.128.
Figure 4.127. Number of unique repliers over number of followers, x>2,500 (10% of data) excluded.
Figure 4.128. Number of likes over number of followers, x>2,500 (10% of data) excluded.
Figure 4.129. Number of retweets over number of followers, $x>2,500$ (10% of data) excluded.

Though the average numbers of unique repliers, likes, and retweets appear mostly consistent for the data with $x>2500$ excluded, the question remains about confounding effects on the lower end. Figure 4.130 is the same as figure 4.124, but zoomed in on the bottom 10% of the data (up to 67 followers). Likewise, figure 4.131 is the same as figure 4.125, but zoomed in on the bottom 10% of the data. Finally, figure 4.132 is the same as figure 4.126, but zoomed in on the bottom 10% of the data.
Figure 4.130. Number of unique repliers over number of followers, x>67 (90% of data) excluded.
Figure 4.131. Number of likes over number of followers, x>67 (90% of data) excluded.
Interestingly, a low number of followers (or even no followers) does not seem to consistently affect the number of unique repliers to a given tweet. However, it does affect the numbers of likes and retweets that a given tweet receives. As the follower number nears ~67, the trendlines of figures 4.131 and 4.132 approach the values of those in 4.128 and 4.129.

In short, for all types of response, follower number does appear to be a confounding variable, but only for the top ~10% of the data (for follower numbers over 2500). For likes and retweets, follower number is also a confounding variable for the bottom ~10% of the data (for follower numbers up to 67). To remove (or at least minimize) these confounding effects, only the data within those bounds (follower numbers between 67 and 2500) was used in the analyses of response predictors.
Significant Predictors of Response to Original Tweet

Four (linear) regression analyses were performed for four different response variables: total number of unique repliers, total number of likes, total number of retweets, and total number of responses (unique repliers + likes + retweets). In all cases, the explanatory variables were the same, and are listed below.

**Direct act fractions by category (6):** Each of these six variables consists of the number of segments in the tweet that perform a given type of direct act over the total number of (non-automatic) direct acts performed in that tweet (e.g. a tweet that performs four direct acts of which two are assertive would have a Fraction Assertive Direct Act value of 0.5).

**Fraction of possible indirect acts that occur:** This variable consists of the total number of segments in a given tweet that perform an indirect act over the total number of segments that could perform an indirect act in that tweet.

**Hashtag count:** This variable consists of the total number of hashtags in the tweet.

**@-tag count:** This variable consists of the total number of @-tags in the tweet.

**Emoji count:** This variable consists of the total number of emoji in the tweet.

**Vulgar word count:** This variable consists of the total number of vulgar words that appear in the tweet.

**Tweet-final punctuation:** This variable indicates whether the final (or only) segment of the tweet ends with punctuation.
**Fraction standard capitalization:** This variable consists of the number of segments in the tweet that exhibit standard capitalization over the total number of relevant segments in the tweet.\(^\text{138}\)

**Personal pronouns (4):** Each of these four variables consists of the count of personal pronouns of one category in the tweet: first person pronouns, second person pronouns, third person pronouns, and all three together.

**Interaction terms:** Interaction terms were included for each combination of direct act fraction variables, and for the interaction of the total personal pronoun count variable with the other three personal pronoun count variables.

Additionally, all four regressions were subjected to the same method of model selection. Specifically, “drop one” methodology was used. The regression was first performed with all explanatory variables. Then, a variable was dropped from the model, and the two models (the version with the variable and the one without) were compared using chi-squared ANOVA. A significant variance (p ≤ 0.05) meant that the variable was kept, while an insignificant variance meant the variable was dropped.

**Regression #1: Response variable = total number of unique repliers**

After model selection, two variables remained for this regression: the fraction of possible indirect acts that occurred, and the interaction term for the fraction of interrogative direct acts and the fraction of expressive direct acts. Within that model, both of those terms proved to be

---

\(^{138}\) "Relevant" here means that the segment does not consist solely of a hashtag, an @-tag, or an emoji.
significant, as did the interrogative fraction term alone.\textsuperscript{139} The association plot for the fraction of possible indirect acts with the number of unique repliers is given below, as figure 4.133.

\textsuperscript{139} An interaction term in the model means that both component terms are kept in the model as well.
Figure 4.133. Association for Fraction of possible indirect acts that occur and Number of unique repliers.
Generally, it appears that more indirectness means fewer repliers. If we look at the top row, we see that an indirect fraction of 0 was associated with non-zero numbers of repliers. Meanwhile, looking at the bottom row, we observe the other side of that trend (i.e. indirect fraction of 1 is associated with 0 repliers).

Meanwhile, a significant interaction term may point to a crossover interaction between the main terms. Figures 4.134 and 4.135 plot the associations of those main terms (fraction of direct acts that are interrogative and fraction of direct acts that are expressive, respectively) with the number of unique repliers.
Figure 4.134. Association for Fraction of direct acts that are interrogative and Number of unique repliers.
Figure 4.135. Association for Fraction of direct acts that are expressive and Number of unique repliers.
As expected, the associations shown in these two figures are frequently (though not categorically) opposed. This trend is best observed by looking at the columns for 0 repliers and 1 replier. For example, a replier count of zero is negatively associated with an interrogative direct act fraction of 1, but positively associated with an expressive direct act fraction of 1.

As for the interrogative fraction plot alone, we can see a logical trend in the left-most column of figure 4.134. For an interrogative fraction of 0-0.25, there is a positive association with a replier count of 0. But for higher interrogative fractions, there are negative associations with a replier count of 0. This trend is then inverted in the next column (replier count = 1). Thus: if more of the tweet is asking a question, there is more of a chance that someone will reply.

**Regression #2: Response variable = total number of likes**

After model selection, no variables remained for this regression. That is, none of them produced a significant change in the model when dropped. The variable that came the closest was the number of vulgar words (p = 0.19). The association plot for this variable with number of likes is given below, as figure 4.136.
Figure 4.136. Association for Total number of vulgar words and Total number of likes.

The top row, for which the total number of vulgar words is 0, shows a positive association with all non-zero numbers of likes except 3 and 9. Meanwhile, there is a negative association between a vulgar word count of 0 and a like count of 0. The next row down, for which the total number of vulgar words is 1, shows exactly opposite associations (and the bottom three rows are based on very little data, as indicated by the increasingly thin bars). But the main takeaway is that more vulgar words in a tweet may reduce the tweet’s chances of getting likes.

Regression #3: Response variable = total number of retweets

After model selection, two variables remained for this regression: the total hashtag count, and the total @-tag count. Within that model, only the hashtag count variable proved to be significant (p = 0.037, @-tag count p-value = 0.067). The association plots for this variable with the total number of retweets is given below, as figures 4.137.
Figure 4.137. Association for Total number of hashtags and Total number of retweets.

For 0 to 1 hashtags, there is almost no association with 0 retweets either way. That is, a tweet with 0 or 1 hashtags is neither likely nor unlikely to get 0 retweets. For non-zero retweet counts (and a hashtag count of 0 or 1), there is a general trend that a higher retweet count is negatively associated with 0 hashtags and positively associated with 1 hashtag. For the higher hashtag counts (2, 3, and 5), the associations are inconsistent (likely because they are based on much less data).

Regression #4: Response variable = total number of responses (repliers+likes+retweets)

After model selection, as with regression #2, no variables remained. However, as with regression #2, the vulgar word count variable came the closest to causing a significant change in the model (p = 0.17). Figure 4.138 shows the association between vulgar word count and the total number of responses.
Figure 4.138. Association for Total number of vulgar words and Total number of responses

Figure 4.138 roughly suggests that more vulgar words in a tweet may reduce the tweet’s chances of receiving any responses at all. This figure is very similar to figure 4.136 (vulgar word count vs. total number of likes), which is unsurprising seeing as there are more than twice as many likes in the data (1011) as there are unique repliers and retweets combined (360).

Summary

The significant associations from these four regressions are few in number. The ones that do exist may be summarized as follows: 1) more indirectness in tweets is associated with fewer repliers, 2) a higher fraction of interrogative acts is associated with a higher chance of a replier, and 3) a tweet with 0 hashtags is less likely to get a high retweet count than a tweet with 1 hashtag. Meanwhile, none of the variables tested showed a significant association with either the
number of likes or the total number of responses. Vulgar word count came the closest, showing an association between fewer vulgar words and a higher chance that the tweet will be liked.

Despite the significant association between indirectness and replier number in the data, it should be reiterated here that the indirectness variable is built on shaky ground. That is, the inter-annotator agreement for whether or not a given segment was performing an indirect act was low.

**Significant Predictors of Responses to @-tags**

Another regression analysis was performed on only the tweets that included at least one @-tag, in search of which characteristics of the tweets appear to be significant predictors of whether the/an @-tagged user responded to the tweet. In pure numbers, 37.5% of the tweets that include at least one @-tag received at least one reply from an @-tagged user. The question is whether we can identify differences between the characteristics of @-tags that are responded to by the/an @-tagged user and those that are not.

This regression used the same set of explanatory variables as the regression analyses in part A. The response variable was binary, indicating whether or not an @-tagged user had responded to the tweet. After model selection, two variables remained: the interaction term for directive fraction and expressive fraction, and the interaction term for third person pronoun count and total personal pronoun count. Within that model, only the third person pronoun count and the interaction term it was a part of proved to be significant.

As stated above, significance of an interaction term may point to a crossover effect. Figures 4.139 and 4.140 show the associations for the components of the interaction term, third person pronoun count and total personal pronoun count, respectively.
Figure 4.139. Association for Total number of third person pronouns and @-tagged user response yes/no.

Figure 4.140. Association for Total number of personal pronouns and @-tagged user response yes/no.
A crossover effect is not immediately apparent between these two plots. For a variable value of 0 (the top row), the two plots do show opposite associations. But for non-zero variable values, the associations are the same.

As for the third person pronoun count alone, 0 or 1 third person pronouns is associated with a higher chance of no response, while 2 third person pronouns is associated with a higher chance of response.

In short, only the third person pronoun count variable proved to be a significant predictor of whether an @-tagged user would respond. Specifically, if the tweet used more than 2 third person pronouns, the @-tagged user was more likely to respond.

4.2.5.2 Discussion

There are many more possible elements of style than those I have assessed here in section 4.2.5. The slight trends and associations I have found do not encompass all of the communicative style of Twitter. But their presence supports the conclusion that such a style exists (and that they represent a part of it). Hence, the hypothesis that “There is a distinct communicative style to Twitter, and we will be able to see evidence for it in the corpus” has been supported. As their familiarity with Twitter grows, a poster is more likely to post in a certain way, to accommodate to the style of Twitter. Meanwhile, there are also certain qualities of a post that make a user more likely to respond to it.

Now that we have seen evidence that there exists a communicative style to Twitter, and that it is reflected in the corpus, we may discuss how my findings may reflect the community itself. Stylistic choices can index higher goals and values of a community. What higher goals and values do the stylistic choices I’ve observed for Twitter index?
As discussed in section 2.4, a higher goal and value of authenticity has been observed by several previous studies of Twitter. Such a goal can help explain some of my findings. Users posting more expressive acts, which can signal honesty and vulnerability, over time is one such finding from section 4.2.5.1.1. Another is the initial drop in tweet length over total tweets sent, as a shorter tweet may be interpreted as a more off-the-cuff thought. Likewise for the drop in total personal pronoun count over time, as subject-dropping definitely suggests a certain offhandedness. Meanwhile, from section 4.2.5.1.2, the fact that a higher fraction of indirectness is associated with a lower chance of reply may be explained as a rejection of inauthenticity by potential repliers. That is, potential repliers may not want to encourage indirect acts because they are inherently covert (which may be interpreted as inauthentic). Also, the fact that a higher fraction of interrogative acts is associated with a higher chance of reply makes sense through the lens of authenticity, as the audience would interpret the question as genuine.

Another potential goal is an appropriate level of formality, though where exactly Twitter falls on the spectrum of formality appears to be as unclear to Twitter users as it has been to previous researchers of the topic (see section 2.4). In my data, in section 4.2.5.1.1, the usage of standard markers of formality like punctuation and standard capitalization practices exhibited dynamic trends. The use of tweet-final punctuation initially went down over total tweets sent, before steadily increasing. The use of standard capitalization showed the opposite trend -- an initial increase followed by a decrease (with a wide margin of uncertainty). These trends may reflect uncertainty on the part of users as to how formal they should be. This result is unsurprising, given the continually evolving nature of ideologies for new media.
Trying to find an explicit goal or value to explain every communicative norm on Twitter is likely a fool’s errand. Some of my findings are probably best explained just by accommodation itself. That is, users were accommodating to what everybody else was doing, simply because everybody else was doing it. For example, the initial drops in hashtag and @-tag counts over time, in section 4.2.5.1.1, seem best explained by new Twitter users coming to realize that the elements of Twitter communication that they had probably heard most about (before joining Twitter) are not actually used as commonly as the level of discussion around them would suggest. There is also likely a “novelty wearing off” effect there. Meanwhile, the emotive/meta-commentative hashtag, while also an oft-discussed element of Twitter communication, is more esoteric. The fact that their usage goes up over time may therefore be due to a learning curve in how best to construct and deploy them to fit in.

4.3 Introduction to Summary Discussion and Conclusions

In this section, I will briefly step back and assess how choices made in the design of this study may have affected the results. In the next section, I will summarize the results and conclusions of the individual hypotheses, and take a step back to assess the overarching question of what we can learn about language by studying the language of Twitter.

There are three main areas in which the design choices may have affected the results of the study. The first is the version of speech act theory that was used to define the speech act categories used in the annotation. Most significantly, if a contextualist version of speech act theory had been used, there would have been no direct/indirect distinctions to annotate. Meanwhile, if a more conventional version of speech act theory had been used for the definitions (e.g. Searle & Vanderveken), the distinction between direct and indirect would have existed but
been more difficult to delineate to the annotators. This is because Alston’s (2000) theory and category definitions lay great importance on the fact that direct illocutionary acts have *purported* conditions and effects. This stipulation paves the way for jokes and serious-but-infelicitous (e.g. seriously giving an order while not realizing you lack the authority) utterances to be annotated for direct acts. Speech act category definitions without the “purported” stipulation may have made the distinction between an indirect act and, say, a non-serious (or serious-but-infelicitous) act less clear.

The second area in which a design choice likely affected the results of the study is in the choice of medium. By using Twitter data, I was able to build a corpus of mostly initial turns, utterances made apropos of nothing, potentially interpreted by strangers. This meant, theoretically, that users would have to put everything necessary for interpretation of their tweet, by a variable audience, into the tweet itself. The pre-existing discourse context is therefore minimized, in comparison with other forms of social media. For example, for both YouTube video comments and online discussion forums, the fact that they have an underlying topic that they are supposedly meant to discuss may mean that information critical to interpretation is not made explicit in the utterance, as it is already known by the majority of people who would read the comments or participate in the forum.

The same goes for Facebook posts, which are initially only viewable by a list of approved and known people (i.e. one’s “friends”). Presumably, one’s friends know a little bit of what is going on in one’s life to begin with, so posts need not include certain critical information. On Twitter, if you have a public account, anyone can follow you. While you may or may not care
whether strangers understand your tweet, the knowledge that they could be reading it undoubtedly affects what and how you tweet.

Finally, the last area in which a design choice likely affected the results of the study is in the annotation question distribution. As discussed in section 4.1.1, annotating indirect acts meant more questions for the annotator. That is, when an annotator answered Yes to the question of whether a given segment was performing an act indirectly, they were then asked to classify that act. Also, if they had said that the direct act was serious, they were asked which was more important. This meant that answering that Yes, there was indeed an indirect act, meant an additional question or two. Annotators may have been tempted, therefore, to say No more often than they would have if the question distribution had been even.

4.4 Summary Discussion and Primary Contributions

In this section, I will first summarize the findings from the five hypotheses addressed in section 4.2. I will then address any answers that these findings may together suggest regarding what we have learned about language through this investigation of the language of Twitter. The section will conclude with a summary of the contributions made by this project, as well as recommendations for future study.

Section 4.2.1 addressed the original hypothesis of this project, that Certain overt elements available to Twitter users function as direct illocutionary force indicating devices (IFIDs, from Searle & Vanderveken (1985). Through regression and random forest analyses of the corpus, with direct act category as the response variable, a variety of features were shown to have a significant effect on the model, and be either positively or negatively associated with a given direct act category. Some of these associations appear to be almost categorical (e.g. question
mark \dagger interrogative), while others are weaker and/or associated with more than one category. Some of these are explained as taken from standard writing conventions (e.g. question mark), some of them seem linguistically grounded,\textsuperscript{140} and some do not seem to have an immediate explanation (and are possibly either new conventions, or simply noise).

Nonetheless, viewed together, the various collections of features may be seen to constitute templates of the typical act of each type. Such a template can itself function as an IFID. A hearer vacillating between two interpretations might finally make a decision because the segment seems more “typical” of (i.e. exhibits more of the attributes of), say, an assertive than a commissive.

Section 4.2.2 addressed the hypothesis that *Certain overt elements available to Twitter users signal the presence of an indirect act*. The question of whether a given segment performed an indirect act exhibited poor agreement in the annotation. While this could be interpreted to mean that the analysis of this hypothesis was compromised to begin with, it could also be interpreted as a result relevant to the hypothesis. More plainly, it is essentially a refutation of the hypothesis, as the poor agreement may in fact suggest the lack of the hypothesized overt elements. What we may learn from the agreement data will be further addressed later in this section.

With the poor agreement noted, the basic result from the analysis of hypothesis 2 was another typical template, this time for the sort of segment that performs an indirect act: such a

\textsuperscript{140} For example, a segment with an adjective fraction of 1 is often classified as expressive as it is usually either copular with the speaker or with a contextually-salient subject, in which case it may reveal the speaker’s feelings about that subject.
segment constitutes the entire tweet, does not include proper nouns of any kind, includes emoji in some capacity, does not perform an expressive direct act, and is part of an original post.

Additionally, Clark’s (1979) findings regarding indirect acts and seriousness were generally reproduced. That is, segments that perform an indirect act were more likely to co-occur with a direct act segment that was annotated as non-serious, suggesting that seriousness of the direct act may be one inferential tool that contributes to the process of deducing whether an indirect act was performed.

Section 4.2.3 addressed the hypothesis that *We will see a speech act landscape for Twitter that is distinct from that of other social media and spoken conversations*. This hypothesis was supported (in terms of frequency of the different direct act types), but the more interesting result is actually which other media the present corpus most resembled. In short, only IM away messages and spoken conversations (in a meeting context) have the same top two categories, in the same order, as the present study (assertive and expressive). The similarity of away messages diverges from the present study after the second category ranking, but the spoken meeting data exhibits further similarity in the rank and count of interrogative acts. On the other end, the other major social media platform assessed, Facebook, does not even have the same top-ranked act as Twitter (assertive on Twitter, expressive on Facebook).

The similarity to IM away messages is actually an intended result of Twitter’s creator, Jack Dorsey (Zappavigna, 2017, p.206). Meanwhile, the similarity of the platform to the spoken meeting data may be a result of “context collapse” (Marwick & boyd, p.117), since Twitter users may employ the same strategies for appealing to the masses as do meeting participants, resulting in the same types of speech acts.
One other result from section 4.2.3 was that the lack of exercitive acts observed in the present corpus was observed across the various other CMC media. The reason for this was proposed to be that the unquestioned authority necessary to perform an exercitive is inherently unattainable in the medium, due to the lack of guarantee that the person behind a post is indeed who they purport to be.

Section 4.2.4 addressed the hypothesis that The speech act profiles of unincorporated hashtags vs. the profiles of incorporated hashtags will be distinct, suggesting that "incorporated vs. unincorporated" is pragmatically meaningful. The annotated questions involving the hashtags all received poor to low agreement, so some objective qualities of the hashtags were also assessed. Unincorporated hashtags were more likely to come at the end of a tweet, while incorporated hashtags were more likely to be incorporated into a segment at the beginning of the tweet, or into a segment that comprises the entire tweet. Unincorporated hashtags were also more likely to be judged as performing an evaluative indirect act (though, as just noted, this question yielded poor agreement). Combined with the positional information, this is a logical result. One must have first said something in order to evaluate it.

Unincorporated hashtags were also more likely to be judged as contributing to a meme or trend (though, again, this question yielded poor agreement). Hashtag memes/trends usually have a template that the user is supposed to follow if they want to participate (Wikstrom, 2014, p.135). It appears that this template more often calls for including the hashtag unincorporated than for incorporating it.

Finally, section 4.2.5 addressed the hypothesis that There is a distinct communicative style to Twitter; and we will be able to see evidence for it in the corpus. A variety of features
were assessed over time on Twitter (i.e. how many days since the poster joined Twitter) as well as over the total number of tweets sent by the poster ever. Several features displayed trends over time, total tweets, or both, suggesting that users were accommodating to a certain style (e.g. more expressive acts and fewer personal pronouns over time). The behavior of respondents was also assessed, and certain features displayed trends there as well. I proposed that several of the findings from this section could be explained by a style that promotes (or at least aspires to) authenticity and is uncertain about its own level of formality, both attributes of Twitter attested in the literature.

Each of the hypotheses addressed in section 4.2 had a specific motivation. Hypothesis 1 was motivated by uncertainty in the literature regarding whether and how illocutionary force is overtly encoded. Hypothesis 2 was motivated by the fact that indirect acts have not received nearly as much attention in the literature as have direct acts, despite the general agreement that they exist. Hypothesis 3 was motivated by literature which suggests that computer-mediated communication should not be viewed as a communicative monolith, and that the specific choice of media matters. Hypothesis 4 was motivated by an observation I made during a pilot for this project. Finally, hypothesis 5 was motivated by the general proposal that Twitter users are a linguistic community, a fact which should be reflected by accommodation to a set of stylistic norms.

All of these specific motivations also had a general, underlying motivation, which was: what can we learn about language by analyzing the language of Twitter? As the specific hypotheses and specific motivations were addressed in section 4.2 and summarized above, we may now summarize any answers we have found to this underlying question.
Most pervasively, I would propose that we have seen evidence for instances of linguistic meaning as composed of a set of norms, whether more formal linguistic meaning (e.g. these norms together typify a speech act of X category), or more sociolinguistic meaning (e.g. these norms together identify a user as a member of X community). In both cases, the actual form of a given norm (or at least, the ones assessed in this study) is represented as the presence, absence, or distribution of a feature. And crucially, in both cases, one such norm alone rarely (if ever) has absolute power. On the speech act side, even a segment ending in a question mark (one of the most consistent features for speech act classification) was not guaranteed an annotation of interrogative. On the sociolinguistic side, outliers can be identified in just about every figure of section 4.2.5. This leads to the stipulation that negotiated meaning-making requires reference to a set of features, as opposed to individual ones.

Such a proposal, based on this study alone, does require a disclaimer. In short, it is worth mentioning the omnipresent caveat that correlation does not equal causation. Several of the features found to have a significant effect on the direct act classification of a segment are more likely to be incidental byproducts of certain inferential steps than to be directly (causatively) linked to that classification. For example, the presence of a feature like the question mark, which has a strong history as a standard writing convention to mark questions, probably does have a causative effect on the direct act classification of a segment. That is, a person sees the question mark and automatically thinks, “This is probably an interrogative segment.” Meanwhile, a feature like “verb fraction” seems unlikely to have such a direct relationship. The fact that “verb fraction” was found to have a significant effect on the direct act classification model is hypothesized to be a byproduct of the inferential step that verbless fragments are assumed to be
copular with the most contextually-salient subject (which usually results in an assertive or expressive). Is it the case that this assumption is so inherent to our inferential process that we skip over the logic of that assumption itself and interpret a lack of verbs as directly signalling an assertive or expressive act? From this study alone, one would not be able to say.

The agreement results also provide some answers to the question of what we can learn about language from analyzing the language of Twitter. First of all, the agreement results for direct act categorization were the best out of all of the questions asked, even with the AQCs excluded. The mildest conclusion to be made from this is that the direct act that a segment performs is less ambiguous and more obvious than is the indirect act (if there is one). This conclusion is precisely what we would expect given the literature on direct and indirect acts, seeing as there is a consensus that indirect acts are generally inferred (while most of the literature on direct acts agrees that it is at least partially overtly encoded).

Meanwhile, the question of whether a segment was performing an indirect act at all yielded poor agreement results. As mentioned above, this supports a conclusion that overt elements which signal the presence of an indirect act are not available to Twitter users, and indeed may not exist (at least not as systematic signals that function consistently regardless of context). That is, perhaps we are seeing a root cause of the oft-maligned pragmatic ambiguity of social media communication.

141 Including the AQCs would have artificially inflated the agreement results seeing as all three answers to each one were required to be the same. However, as mentioned in section 4.1, segments used for AQCs were so chosen because they seemed unambiguously of a certain type, and half of them actually received perfect agreement with no interference (i.e. only three attempted responses were necessary to yield three complete responses). In short, if the present study had not designated AQCs and those segments had instead simply been annotated like the rest, the agreement score would most likely have been higher.
As for why we do not see systematic signals that function to indicate the presence (or force) of an indirect act regardless of context, I would posit a fundamental motivation: the preservation of deniability. In this study, the metric of deniability was the ultimate determiner of whether an act was direct or indirect. If a given feature is so systematic that speakers will always interpret it as signalling the presence/force of an indirect act, this fact significantly reduces the deniability of that act. The lack of consistently-used overt signals preserves that deniability (and hence the direct/indirect boundary) instead. It may be that communication needs deniability, which is why we have indirect speech acts in the first place.

One might argue that, in using deniability as the assessment of directness, this study was inherently precluded from discovering overt signals of indirect presence. However, I would reiterate that the existence of such an element would only reduce the deniability of an indirect act, not eliminate it entirely. If such a feature existed, a speaker could use it to intentionally signal the presence of an indirect act while giving no further clues about the exact nature of that act. Furthermore, annotators were only asked to use deniability to assess whether an act itself was deniable, not whether the presence of any additional acts was deniable.¹⁴²

On the topic of indirect acts, it is worth noting that in the few cases where all three annotators agreed that the segment did indeed perform an indirect act, the agreement score for the category of that act was actually moderate. This may suggest a spectrum as regards the inferrability of indirect acts. That is, that some indirect acts are easier to infer, requiring fewer cognitive jumps be made by the hearer.

¹⁴² In other words, there is quite a difference between a denial of “I didn’t mean [a specific indirect act]!” and a denial of “I didn’t mean anything except [the direct act]!”
The agreement results for all of the hashtag-related questions were informative, as well. As discussed in section 4.1, the agreement score on these questions was poor to low. However, the extent to which this score reflects the difficulty of the task is debatable, seeing as the use of only three annotators for binary questions such as these means the chance of agreement is already fairly high, which greatly affects the Fleiss kappa score. To get a better sense of the true difficulty of these questions, more than three annotators would need to be used.

The indirect act data may provide another interesting answer to the question of what we can learn about language from analyzing the language of Twitter. As touched on in sections 4.2.1 and 4.2.2, most of the indirect acts annotated were expressive. While this could mean, of course, that most of the indirect acts that occurred in the corpus were expressive, one might instead interpret this to mean that expressive direct acts which were occurring in parallel to non-expressive direct acts (i.e. in the expressive dimension, parallel to the descriptive dimension) were annotated as indirect, seeing as there was no option for annotating both concurrently. That is, in my corpus, without the option to declare a parallel expressive direct act, it may be that situations of expressive language in otherwise non-expressive utterances led to the annotation of an expressive indirect act.

As a final conclusion, I would like to address the more general question of what linguistics can learn from an empirical study of speech acts. What is the primary benefit of identifying, say, an empirically-supported taxonomy and/or empirically-supported overt force indicators?

In the case of an empirically-supported taxonomy, I would posit that the primary benefit of identifying such is that it would allow us to investigate the underlying dimension. That is,
with the categories themselves in hand, we could then ask why these categories? As I noted in section 1.1.2.3, my main hesitation with regard to Alston’s account of speech acts is the fact that he does not provide a unifying notion or dimension within which the only options yield the final categories in the taxonomy. Other scholars (including Searle & Vanderveken, and Bach & Harnish) provide more theoretical grounding for the taxonomies they propose. However, instead of beginning with the theory and working up to a taxonomy, we could seek to empirically identify a taxonomy that best fits natural language and then use that taxonomy to identify the unifying dimension. Such a dimension would be a locus of pragmatic information.

I would posit a similar benefit to identifying empirically-supported overt force indicators. With such in hand, we can then investigate what they have in common. Additionally, it would be interesting to assess the potential differences between arbitrary force indicators (e.g. question mark = interrogative) and grounded ones (e.g. verb fraction of 0 = assertive or expressive). Do certain speech act categories favor arbitrary indicators over grounded ones (or vice versa), and why might that be? Do certain linguistic communities favor one type of indicator over the other? In short, the empirical identification of overt force indicators opens up an entire new avenue of research possibilities.

### 4.4.1 Primary Contributions of the Project

At the time of writing, speech act theory has been around for about 60 years, but Twitter has only been around for about 13. Studies of speech acts on Twitter have remained infrequent. The present study has thus contributed to both speech act literature and Twitter linguistic literature by investigating the intersection of the two topics.
Additionally, the present study has taken some novel steps. While previous studies have built manually-annotated corpora of speech acts on Twitter in the course of developing automatic speech act classifiers (Vosoughi & Roy, 2016; Zhang et al., 2011 and 2012), these studies did not seek to separate direct acts from indirect acts, nor did they accommodate the possibility of a tweet containing multiple speech act units. The present study has done both.

More specific contributions made by this study have been both theoretical and tangible. The main theoretical contributions may be found in the proposed answers to the questions asked.

- We have seen support for the proposal that direct illocutionary force is at least partially overtly encoded, specifically as a template of typical features.
- We have seen support for the proposal that expressive language has the potential to perform additional speech acts, separately from the act performed by the utterance without the expressive language.
- We have seen that users of a social media platform form a linguistic community, complete with norms to which they accommodate over time and/or through experience.
- We have seen some ways in which context collapse may ultimately affect communication in a medium.

The project made additional theoretical contributions in the successes and failures of the methodologies utilized.

- We have seen support for crowdsourced annotation as a feasible method of obtaining linguistic data.
• We have also seen, however, that this feasibility varies, seemingly according to
the level of inference required to address a given phenomenon. Consequently,
inferred phenomena may especially benefit from higher numbers of raters.

Finally, the tangible contribution is the corpus itself, as others may use the corpus that I
have built to address their own questions.

4.4.1.1 Recommendations for Future Study

The results of this study recommend several avenues for future study. First of all, a
version of the present study with crucial modifications would provide an informative point of
comparison. For example, the low agreement scores may be improved by increasing the number
of raters (seeing as that would lower the percent chance of perfect agreement, even for the binary
questions). Likewise, as mentioned above, the low agreement scores for the indirect act
questions may be improved by making sure that the annotator is answering the same number of
questions whether they declare an indirect act or not. Additionally, most if not all aspects of the
present study would be improved by collecting and annotating a larger corpus.

The results of the present study also recommend a study that focuses on the question of
whether differing direct annotations are more often due to mistakes or more often due to
non-erroneous ambiguity. If the latter, the exact nature of the ambiguity would also need to be
addressed. Within the framework of the present study, there are essentially three different
situations that would explain non-erroneous ambiguity: 1) the segment in question does not
contain any of the features found to be significant for any of the speech act categories; 2) the
segment in question contains significant features, but only ones which are themselves
ambiguous; and 3) the segment in question contains significant features, but equally as many for each of the categories in question.

Finally, the results of the present study might also recommend a study which utilizes a more direct method of inquiry. That is, instead of having annotators choose the direct act category and move on, directly ask the annotators why they chose a given category. For example, a follow-up question after the classification question would ask something like, “You chose commissive for this act. Was there any particular in-text indication that swayed your decision? If so, what was it?” The answers to such a question, if not primarily in the negative, could then be grouped according to feature type.
Appendix 1

Instructions and Sample Annotated Tweets

Instructions and Examples

The task will be organized in the following way:

- You will be shown an image (or images) of a Twitter conversation.
- You will then answer a few questions about each relevant segment of the conversation separately.

For most segments, you will be asked to classify the segment’s direct speech act, as well as to assess whether the segment is also performing an indirect speech act (and, if so, to classify it).

[Note: The following instructions and examples will be available to you during the task for reference. In addition, definitions of significant terms will be available in hover-text, for quicker reference.]

Figure A.1. Annotation task introduction block.
Speech Act Categories

The speech act of a segment is what the speaker accomplishes in saying the words. In this task, you will be categorizing these acts into one of six mutually exclusive categories based on what exactly the speaker purports to accomplish:

- **Assertive**: The speaker has purported to make an assertion.
  - John went to the store.

- **Directive**: The speaker has purported to give a directive.
  - Go to the store for me, please.

- **Interrogative**: The speaker has purported to ask a question.
  - Who will go to the store?

- **Commissive**: The speaker has purported to commit to an action.
  - I will go to the store.

- **Expressive**: The speaker has purported to express a psychological state.
  - I'm angry that the store is closed.

- **Exegetic**: The speaker has purported to bring about a change in the real world by way of their utterance.
  - I hereby name this store “John.”

Figure A.2. Task speech act category explanation.
**Direct vs. Indirect Speech Acts**

A given segment performs one of these acts *directly*, and may also perform another act *indirectly*.

- The direct speech act of a segment is the one that is *on the surface*, and *undeniable* in normal social circumstances.
- The indirect speech act of a segment, *if there is one*, is *beneath the surface* -- it is a speech act performed *by way of* another speech act (the direct act). Unlike a direct act, an indirect act is *deniable*.

Consider the panels below. In panel 1, the speaker *directly asserts* something ("I couldn’t hear you"). A reasonable conversational partner might infer that, in asserting what they did, the speaker in fact means to *indirectly request* something ("Please repeat yourself"). In panels 2 and 3, we see the speaker attempting to deny the direct and indirect acts, respectively.

*Figure A.3. Task direct and indirect explanation.*
Figure A.4. Task direct and indirect example.
Serious vs. Non-serious Direct Acts

A direct speech act may be performed *seriously* or *non-seriously*. Consider the panels below. In panel 5, the speaker’s direct act is clearly non-serious (the speaker clearly does not intend for the hearer to actually believe the assertion).

**Panel 4**

**Panel 5: A non-serious direct act**

"I couldn’t hear you!"

---

Figure A.5. Task serious and non-serious direct act explanation.
Literal vs. Non-literal Direct Acts

A direct speech act may be performed *literally* or *non-literally*. Consider the panels below. Both perform the same assertive direct act, but they differ in whether that act is performed literally.

**Panel 6: Literal assertive direct act**

**Panel 7: Non-literal assertive direct act**

"I couldn't catch that" is idiomatic for "I couldn't hear you/that." You will not be asked to differentiate between literal and non-literal direct acts, but take care not to confuse *non-literal* with *indirect*.

Figure A.6. Task literal and non-literal direct act explanation.
Backgrounded Segments

Some of the segments you will be annotating are backgrounded. A backgrounded segment is one that provides context or information that is necessary or helpful in understanding the main point of an utterance, but is not itself the main point. These segments, by default, do not perform a direct speech act. But they may perform a speech act indirectly. Consider the panels below.

Panel 8: Speaker A
Wanna go get dinner?

Panel 9: Speaker B
I was gonna get pizza after I leave work.

Speaker B may have meant for the backgrounded segment "after I leave work" to indirectly inform speaker A that B is at work. That indirect act, something like "I am currently at work," would be assertive.
Final Notes

Regarding direct and indirect acts:

- Not every segment performs an indirect act.
- A segment may perform one type of act directly, and another type of act indirectly.
- A segment may perform a direct act and an indirect act that both happen to be of the same type.
- You may feel that a segment is actually performing multiple indirect acts. In such a case, please code for the indirect act that you think is most obvious.
- Some of the segments in the corpus perform a certain direct act by default (or, no direct act by default). For these segments, you will not be asked to classify the segment’s direct act (although, if there is one, you will be told what it is). You will only be asked if the segment performs an indirect act (and if so, what type).

Regarding the segments you will be annotating:

- Many of the segments in the corpus are fragments. To annotate these segments, please mentally fill out the segment into a “full sentence”, and then annotate according to what that “full sentence” would be accomplishing.
- The segments may include hashtags, @-tags, or emojis (e.g. “Excited for #TheBachelorette finale tonight”, “Saw @John yesterday”, or “Wanna go get some 🍺?”). They may even consist of only emojis (e.g. “😢?”). Please treat these segments as though the hashtag, @-tag, or emoji was just another word in the segment (e.g. “Beers?” for “🍺?”).
- Some of the segments you will be annotating consist of only a hashtag. The questions for these segments will be slightly different from those for the rest of the segments, but in the same vein.
Sample Annotated Tweets

The following tweets have been annotated as examples, with the speech act segments divided by color-coded lines. As a reminder, indirect acts are inferred. Therefore, you may find that your intuition differs with regard to the indirect act classifications given in these samples. Likewise regarding the classification of direct acts as serious or non-serious.

Figure A.9. Task sample annotation tweet introduction.
Sample Tweet 1

Direct: Interrogative, Non-serious
Indirect: Expressive

Direct: Interrogative, Non-serious
Indirect: Expressive

Direct: Assertive, Serious
Indirect: Expressive

i want garlic bread????? at this hour???? i always pick a bad time to crave something 😞

Sample Tweet 2

Direct: Assertive, Serious
Indirect: None

Direct: Assertive, Non-Serious
Indirect: None

These political attack ads are getting pretty ridiculous. Next they'll be saying 'if you vote for so and so, the whole province will get chlamydia and it will be all your fault'.
Figure A.11. Tasks sample tweets #3-4.

Details Regarding HITs Excluded from Final Corpus

This appendix details the AQC s and responses for the four HITs that were ultimately excluded from the corpus.

The first problematic HIT, #90, had two AQC s, the second of which was ambiguous. The segment in question was *wait you have a job*, from conversation #307:
Figure A.12. Excerpt from conversation #307.

On first reading, I thought the original poster was saying “I’m out of work” to mean “I lost my job,” and that the reply was asserting that “No, you DO have a job” (hence I designated that the segment must be annotated as assertive). Upon re-reading, it seems more likely that the original poster was saying “I can leave work now,” and the reply is pretending to be shocked, asking for confirmation, as in “You have a job??” And indeed, the respondents expelled by the second AQC thought likewise.

Hence, out of 21 respondents:

- 10 voluntarily dropped out before the first AQC
- 0 were expelled by the first AQC
- 9 were expelled by the second AQC
  - 7 selected interrogative
• 1 selected directive

• 1 selected exercitive

• 2 successfully completed the HIT\textsuperscript{143}

For the second problematic HIT, #96, both AQCs were ultimately ambiguous. The first AQC was a fragment, and so I believe may have been ambiguous for that reason. The segment is 

*True life*, from conversation #322:

![Screen capture of Twitter conversation](image)

**Figure A.13. Conversation #322.**

I read this tweet as saying, “This is true life. I’m poor, but…” Thus, I marked the segment as assertive. I still think that is true, but I acknowledge that in choosing this segment as an AQC I violated one of the guidelines I had set for myself about them (choosing full or almost full clauses).

The second AQC for #96 was an imbedded interrogative, which I declared must be annotated as assertive. While I stand by that declaration (the segment is undeniably assertive,

\begin{small}
\textsuperscript{143} It is only according to Qualtrics that 2 people successfully completed the HIT, as this was the HIT wherein the same person was able to complete the survey twice.
\end{small}
while the associated interrogative act is ultimately deniable), I also acknowledge that it was perhaps the trickiest AQC I chose, seeing as “I was wondering if…” is a very conventionalized way of indirectly asking a question. The segment is *i was wondering if you could help me find out how many chains 2 chainz has*, from conversation #326:

![Figure A.14. Excerpt from conversation #326.](image)

Hence, out of 16 respondents to HIT #96:

- 3 voluntarily dropped out before the first AQC
- 4 were expelled by the first AQC
  - 2 selected expressive
  - 2 selected directive
- 1 voluntarily dropped out between AQCs
- 6 were expelled by the second AQC
  - 6 selected interrogative
- 2 successfully completed the HIT (including 1 locally recruited annotator)

For the third problematic HIT, #118, the first AQC was fine (a simple directive), but the second AQC was definitely a poor choice for two different reasons. The first problem was that
subject deletion in the segment made the segment appear directive. When read in the context of the rest of the tweet, I believe it is clearly not a directive. But I can understand why it might have been confusing.

The second problem with that AQC, however, is that it is actually ambiguous, on further consideration, between assertive and commissive. This segment is *go to Kavos next Friday*, from conversation #378:

![Tweet](image)

**Figure A.15. Conversation #378.**

In my interpretation, the “full clause” version of this segment is “I go to Kavos next Friday,” which I would classify as assertive. The speaker is not so much committing herself to a future action as stating a fact to set the scene for her other segments (i.e. “I go next week, and I still have these issues,”). However, if someone interpreted the “full clause” version of this segment as “I **will** go to Kavos next Friday,” that might be a commissive.

Hence, out of 16 respondents to HIT #118:

- 9 voluntarily dropped out before the first AQC
- 1 was expelled by the first AQC
1 selected commissive

4 were expelled by the second AQC
   - 2 selected directive
   - 2 selected commissive

2 successfully completed the HIT (including one locally recruited annotator)

Finally, for the fourth problematic HIT, #137, both AQCs were designated as expressives.

In both cases, they were what I would consider classic examples of an expressive interjection.

The first is segment *What a perfect start [...]!*, from conversation #430:

![Figure A.16. Conversation #430.](image1)

The second segment is *what a great night*, from conversation #431:

![Figure A.17. Conversation #431.](image2)
However, I believe this particular type of interjection may be an example of where the line between assertive and expressive becomes fuzzy. On my interpretation, “what a great night” is expressive because it is a fragment closer to “Wow!” or “Holy shit!” than to a propositional articulation like “It was a great night!” But, I can understand the opposite view, that “What a great night!” should be considered assertive because it is conveying the proposition that it was a great night, though likely at least with an expressive indirect act of something like “I’m happy that it was a great night.” In fact, 5 of the respondents who were expelled by the first AQC of #137, and the 2 respondents who were expelled by the second AQC, gave exactly these answers (that the segment was assertive with an expressive indirect act).

In short, on further consideration, these segments were ambiguous and therefore poor choices for AQC.

Hence, out of 24 respondents to HIT #137:

- 7 voluntarily dropped out before the first AQC
- 13 were expelled by the first AQC
  - 13 selected assertive
- 2 were expelled by the second AQC
  - 2 selected assertive
- 2 successfully completed the HIT

\[^{144}\text{Though, as stated before, it is only according to Qualtrics that two Turkers successfully responded to this HIT, as the second successful respondent failed to submit their confirmation code to Turk.}\]
## Appendix 2

Table A.1. Summary of results from section 4.2.1.

<table>
<thead>
<tr>
<th>Category</th>
<th>Possible Attributes</th>
</tr>
</thead>
</table>
| **Assertive** | Final character is period or comma  
Absence of wh-word  
Low verb fraction  
Presence of an expressive indirect act  
Direct act is serious  
Presence of proper noun(s)  
Presence of numeral(s)  
Presence of determiners  
Low, non-zero modal fraction  
Low, non-zero adjective fraction  
Presence of common nouns  
Segment is part of an original tweet  
Segment length between 5 and 16 tokens  
Low, non-zero adverb fraction  
Zero or middling coordinate conjunction fraction  
Zero emoji in tweet, or emoji in tweet fraction greater than 1  
Absence of interjections  
Encompassing tweet consists of only 1 segment, or 4-6 segments |
| **Commissive** | Non-zero verb fraction  
Absence of an indirect act  
Direct act is non-serious  
Presence of verb “tell”  
Absence of proper nouns  
Low, non-zero numeral fraction  
Absence of determiners  
Segment length of 5-8 tokens  
Non-zero modal fraction  
Absence of adjectives  
Absence of common nouns  
Segment is part of a reply tweet  
Low, non-zero adverb fraction  
Low, non-zero coordinate conjunction fraction  
Low to middling non-zero emoji in tweet fraction  
Absence of interjections  
Encompassing tweet consists of 2-3 segments  
Presence of certain bigrams: “watch it,” “i will,” “i’m gonna,” “i’m going” |
Table A.1. (Cont.)

<table>
<thead>
<tr>
<th>Category</th>
<th>Possible Attributes</th>
</tr>
</thead>
</table>
| **Directive** | Final character is exclamation point  
Absence of wh-word  
Non-zero verb fraction  
Direct act is non-serious  
Presence of verb “ask”  
Presence of verb “tell”  
Absence of proper nouns  
Absence of numerals  
Segment length of 1-4 tokens  
Absence of adjectives  
Low, non-zero vulgar word fraction  
Zero or low adverb fraction  
Presence of verb “give”  
Segment is part of a reply tweet  
Zero or low coordinate conjunction fraction  
Low to middling non-zero emoji in tweet fraction  
Middling interjection fraction  
Encompassing tweet consists of 4-9 segments  
Presence of certain bigrams: “gotta stop,” “be safe” |
| **Interrogative** | Final character is question mark  
Presence of wh-word  
Low, non-zero verb fraction  
Presence of an expressive indirect act  
Direct act is non-serious  
Presence of verb “ask”  
Low, non-zero proper noun fraction  
Absence of numerals  
Absence of determiners  
Segment length between 5 and 12 tokens  
Low, non-zero modal fraction  
Zero to low adjective fraction  
Low, non-zero vulgar word fraction  
Segment is part of an original tweet  
Low, non-zero coordinate conjunction fraction  
Lack of emoji in tweet  
Low interjection fraction  
Encompassing tweet consists of 1-3 or 7 segments  
Presence of certain bigrams: “am i,” “is it” |
### Table A.1. (Cont.)

<table>
<thead>
<tr>
<th>Category</th>
<th>Possible Attributes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expressive</td>
<td>Final character is exclamation point&lt;br&gt;Low, non-zero wh-word fraction&lt;br&gt;Absence of an indirect act&lt;br&gt;Direct act is non-serious&lt;br&gt;Absence of proper nouns&lt;br&gt;Absence of numerals&lt;br&gt;Absence of determiners&lt;br&gt;Segment length of 1-4 tokens, or 13-16 tokens&lt;br&gt;Absence of modals&lt;br&gt;Presence of adjectives&lt;br&gt;Presence of vulgar words&lt;br&gt;Absence of prepositions&lt;br&gt;Absence of common nouns&lt;br&gt;Segment is part of a reply tweet&lt;br&gt;Zero or middling adverb fraction&lt;br&gt;Low, non-zero coordinate conjunction fraction&lt;br&gt;Presence of emoji in tweet&lt;br&gt;Presence of interjections&lt;br&gt;Encompassing tweet consists of 2-3 or 7 segments&lt;br&gt;Presence of certain bigrams: “i love,” “i hate,” “i’m so,” “i miss,” “thank you”</td>
</tr>
</tbody>
</table>
BIBLIOGRAPHY


Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of personality and social psychology*. 89(6), 925-936.


Yang, L., Sun, T., Zhang, M., & Mei, Q. (2012). We know what@ you# tag: does the dual role affect hashtag adoption? *Proceedings of the 21st international conference on World Wide Web* (pp. 261-270).


