

**Sociolinguistic Variation and Enregisterment in an Online Community of Practice:
A Case Study of MetaFilter.com**

Kimberly Ann Witten

PhD

University of York

Language and Linguistic Science

September 2014

Abstract

With the emergence of communities that are primarily based in computer-mediated communication (CMC) environments, we see the prevalence of internet-derived neologisms, i.e., *netologisms*. Often these netologisms are acronyms (e.g., ‘LOL’), blends (e.g., ‘weblog’), or other forms of abbreviation. These new forms may present challenges for English phonotactics, which must be spontaneously resolved by first-time speakers of the netologisms. If the forms contain orthographic characters or sequences that do not directly or consistently correlate to specific English phonemes or phoneme sequences, it is likely that these new forms display phonetic variation.

Netologisms can also be used as linguistic resources in taking stances or asserting aspects of identity, especially where phonetic variation is possible. These stances may represent the identity of the group, or they may become associated with particular identities within the group. The process by which sounds, features and word forms become associated with particular identities is known as *enregisterment* (Agha, 2003, 2005; Squires, 2010). Enregisterment has traditionally been studied in sociolinguistics as a function of individuals interacting in face-to-face (FtF) environments (Johnstone, Andrus and Danielson, 2006; Beal, 2009). However, as more of our daily interactions are mediated by computers and technology, attention must be paid to how enregisterment may take place in primarily text-based social environments.

This research presents the first large-scale mixed-methods study of enregisterment occurring in CMC. The varying pronunciations of two netologisms — the community’s nickname (‘MeFi’, from MetaFilter.com) and the collective nickname for its participants (‘MeFites’) — are naturally-occurring sociolinguistic variables that showcase the ongoing negotiation of community conventions and the development of group identity. An exploration of this kind adds an important piece to our broader understanding of linguistic interaction in CMC, while also exhibiting one of the many new directions of sociolinguistic research today.

Table of Contents

Abstract	II
Table of Contents	III
List of Figures	VII
List of Illustrations	IX
List of Examples	X
Acknowledgements	XI
Author's declaration	XIII
Chapter 1: Introduction	1
1.1 Background	2
1.2 Research Questions	5
1.2.1 Hypotheses and Expected Findings	5
1.3 Relevance and Broader Implications	7
1.4 Thesis Structure	7
Chapter 2: Literature Review	9
2.1 Introduction to Literature Review	9
2.2 Models of Community	9
2.2.1 Community of Practice – Definition	9
2.2.2 Community of Practice versus The Speech Community	16
2.2.3 Social Networks and Weak Ties	17
2.3 Capital	19
2.3.1 Economic Capital	19
2.3.2 Cultural Capital	19
2.3.3 Social Capital	20
2.3.4 Virtual Capital	20
2.3.5 Summary of Capital	20
2.4 Computer-Mediated Communication (CMC)	21
2.4.1 Classification of CMC Research Areas	21
2.4.2 Communicative Benefits of CMC	22
2.4.3 MetaFilter as a Social Network Site, or ...?	23
2.4.4 Describing the Features of an Online Community	25
2.5 Registers and Enregisterment	25
2.5.1 Registers – Definition	25
2.5.2 Enregisterment – Definition	27
2.5.3 Indexicality	28
2.5.4 Sociolinguistic Approaches to Variation and Enregisterment	32
2.5.5 Enregisterment of a Language Variety	32
2.5.6 Enregisterment of Forms Within a Variety	37
2.5.7 Summary of Enregisterment	39
2.6 Onomastics	40
2.6.1 Overview of Naming Categories	40
2.6.2 Place Names	41
2.6.3 Personal Names	44
2.6.4 Product and Brand Names	45
2.7 Summary of Literature Review	47

Chapter 3: MetaFilter and the M-Set Variables	48
3.1 Introduction to MetaFilter and the M-Set Variables	48
3.2 The MetaFilter Community – Culture and Context	48
3.2.1 The MetaFilter Subsites	48
3.2.2 A CMDA Outline of MetaFilter	55
3.2.3 The MetaFilter Userbase	61
3.2.4 MetaFilter Comment and Post Frequency	62
3.3 Previous Research on MetaFilter	64
3.3.1 Previous Study of MetaFilter as a CoP	64
3.3.2 Previous Study of MetaFilter and Capital	65
3.3.3 Previous Study of MetaFilter Ethos and Identity	68
3.3.4 Summary of Previous Studies of MetaFilter	70
3.4 The MetaFilter Register	70
3.4.1 Elements of the MetaFilter Register	71
3.5 A Linguistic Overview of the M-Set Variables	73
3.5.1 The Structure & Composition of ‘MetaFilter’	75
3.5.2 Orthographic Features	76
3.5.3 Grapheme-Phoneme Correspondence (GPC) Rules	76
3.5.4 Phonetic Realizations of the M-Set	79
3.5.5 Stress Assignment and Vowel Length	85
3.5.6 Syllabification	86
3.5.7 Morphological Processes	87
3.6 Summary of MetaFilter and the M-Set	88
Chapter 4: Methodology	89
4.1 Introduction to Methodology	89
4.2 Research Design	89
4.2.1 The Research Approach	89
4.2.2 The Researcher’s Perspective	90
4.2.3 The Researcher’s Influence on Enregistration	91
4.2.4 Overview of the Types of Data Collected	92
4.2.5 Online Surveys	92
4.2.6 Participation and Usage Meta-Data	94
4.2.7 Corpora – Word Frequency Tables	95
4.2.8 Qualitative Data from Community Discourse – Posts and Comments	95
4.3 Sampling Design	96
4.4 Measures	97
4.4.1 Measuring the Distribution of the M-Set	98
4.4.2 Demographic Measures	98
4.4.3 Measures of Social Engagement	98
4.4.4 Measures of Metalinguistic Awareness	98
4.4.5 M-Set Stance Measures	99
4.4.6 Other Measures	100
4.4.7 Summary of Measures	100
4.5 Data Collection Procedures	100
4.5.1 The MetaFilter Surveys	101
4.5.2 InfoDump Collection Procedures	106
4.5.3 Word Frequency Table Collection Procedures	106
4.5.4 MetaTalk Thread Collection Procedures	106
4.6 Data Analysis Procedures	107
4.6.1 Data Correction and Adjustment	107

4.6.2	Quantitative and Qualitative Analyses	109
4.7	Summary of Methodology	109
Chapter 5: Data Results		110
5.1	Introduction to Data Results	110
5.2	The Distribution of the M-Set	111
5.2.1	The Pronunciation of ‘MeFi’ — Survey Data	111
5.2.2	The Pronunciation of ‘MeFite’ — Survey Data	112
5.2.3	The Pronunciation of the M-Set — Panel Data	112
5.2.4	Changes in M-Set Distributions Over Time	113
5.3	Metalinguistic Awareness Factors	118
5.3.1	Exclusivity of Use of Preferred M-Set Variants	118
5.3.2	Amount of Thought Given to the Pronunciation of the M-Set	123
5.3.3	Other Measures of Metalinguistic Awareness	124
5.3.4	Analogies in Survey Rationales	127
5.4	Demographic Factors	131
5.4.1	Language Background and Experience	132
5.4.2	Geography	137
5.4.3	Age	145
5.4.4	Gender	150
5.5	Summary of Data Results	151
Chapter 6: Enregistration		153
6.1	Introduction to Enregistration	153
6.2	The Enregistration Process in CMC	153
6.3	Message Chains	155
6.3.1	Message Chain Components	156
6.3.2	MetaFilter Message Chains and the Frequency of ‘MeFi’	157
6.4	Social Engagement Factors and the Enregistration of the M-Set	161
6.4.1	Year of Joining MetaFilter	164
6.4.2	Frequency of Visitation to MetaFilter Subsites	165
6.4.3	MetaFilter Site Participation and Enregistration	172
6.4.4	Frequency of Podcast Listening	175
6.4.5	Podcast Listening and Enregistration	179
6.4.6	Frequency of Meetup Attendance	181
6.4.7	Meetup Attendance and Enregistration	184
6.4.8	The Effect of Social Engagement on ‘MeFi’ Pronunciation	185
6.4.9	Summary of Social Engagement Factors	186
6.5	M-Set Stances	187
6.5.1	Stances From Survey Rationales	188
6.5.2	Public Stances From MetaTalk Posts about the M-Set	191
6.5.3	Indexicalities of ‘MeFi’ and Popular Pronunciation Variants	198
6.6	Summary of Enregistration	202
Chapter 7: Discussion & Conclusion		205
7.1	Introduction to Discussion & Conclusion	205
7.2	Overview of Research	205
7.3	Summary of Data Results	206
7.3.1	Summary of Findings Related to Data Populations	206
7.3.2	Summary of Findings Related to Change Over Time	206
7.3.3	Summary of Findings Related to the Pronunciation of the M-Set	207
7.3.4	Summary of Findings Related to the Process of Enregistration	208

7.4	General Interpretation of Findings	209
7.5	Research Hypotheses and Inconsistent Outcomes	211
7.6	Research Limitations	215
7.6.1	Internal Validity	216
7.6.2	External Validity	218
7.6.3	Statistical Analysis	218
7.7	Implications of Findings	219
7.7.1	Implications for the Sense of ‘Community’	219
7.7.2	Implications for the Community of Practice (CoP) Model	220
7.7.3	Implications for Enregisterment	221
7.7.4	Implications for Indexicality	223
7.7.5	Implications for Onomastics	224
7.8	Future Research Directions	225
7.8.1	Future Research on Netologisms	225
7.8.2	Future Research on Enregisterment	226
7.8.3	Future Research on Sociolinguistic Behavior in CMC Environments	226
7.8.4	Summary of Future Research Directions	227
7.8.5	The Future of the M-Set	228
7.9	Concluding Remarks	228
Appendices		229
Appendix A: MetaTalk Posts About the Pronunciation of the M-Set		229
Appendix B: Survey MetaTalk Posts		230
Appendix C: Survey Information & Consent Pages		232
Appendix D: The MetaFilter Surveys		234
Appendix E: The MetaFilter Register		239
Appendix F: Orthographically Similar Forms to ‘MeFi’		240
Appendix G: Summary of Findings Tables		241
Appendix H: Regression Models		244
References		246

List of Figures

Figure 1. Wenger's (2011) 'Levels of Participation' in a CoP	15
Figure 2. The Message Chain Process	28
Figure 3. Screenshots of the MetaFilter Redesign, Before (Classic) and After (Modern)	49
Figure 4. Screenshot of the Front Page of MetaFilter, "the Blue"	50
Figure 5. Screenshot of the Front Page of MetaTalk, "the Gray/Grey"	51
Figure 6. Screenshot of the Front Page of AskMetaFilter, "the Green"	52
Figure 7. A MetaFilter Post Page and Comments	56
Figure 8. Flat Commenting Versus Threaded Commenting Structure	60
Figure 9. The MetaFilter Userbase – All Registered Participants	61
Figure 10. Average Number of Posts Per Day on MetaFilter Subsites	63
Figure 11. Average Number of Comments Per Day on MetaFilter Subsites	63
Figure 12. Image of Word Frequency Table Format	95
Figure 13. Screenshot of 'MeFi' Pronunciation Survey Question	102
Figure 14. Pronunciation of 'MeFi' by Variant and Survey Year	111
Figure 15. Pronunciation of 'MeFite' by Variant and Survey Year	112
Figure 16. Pronunciation of 'MeFi' by Variant and Survey Year – Panel Data, N=769	113
Figure 17. Pronunciation of 'MeFite' by Variant and Survey Year – Panel Data, N=769	113
Figure 18. Measure of Exclusive Use of 'MeFi' Variant – 2010 and 2012 Panel Data	119
Figure 19. Measure of Exclusive Use of 'MeFite' Variant – Survey Data	120
Figure 20. 'MeFi' Pronunciation by Exclusivity of Use – 2010 Survey Data, N=2,473	121
Figure 21. 'MeFi' Pronunciation by Exclusivity of Use – 2012 Survey Data, N=1,929	121
Figure 22. 'MeFi' Pronunciation by Exclusivity of Use – 2010 Panel Data, N=755	122
Figure 23. 'MeFi' Pronunciation by Exclusivity of Use – 2012 Panel Data, N=761	122
Figure 24. Amount of Thought Given to the Pronunciation of the M-Set	123
Figure 25. English as a Native Language – 2010 and 2012 Survey Data	133
Figure 26. 'MeFi' Pronunciation by Language Experience – 2010 Survey Data, N=2,460	136
Figure 27. 'MeFi' Pronunciation by Language Experience – 2012 Survey Data, N=1,868	136
Figure 28. 'MeFi' Pronunciation by Language Experience – 2010 Panel Data, N=758	136
Figure 29. 2010 World Map of Survey Respondents' Country of Residence	138
Figure 30. 2012 World Map of Survey Respondents' Country of Residence	138
Figure 31. 'MeFi' Pronunciation by Country of Residence – 2010 Survey Data, N=2,453	139
Figure 32. 'MeFi' Pronunciation by Country of Residence – 2012 Survey Data, N=1,906	139
Figure 33. 'MeFi' Pronunciation by Country of Residence – 2010 Panel Data, N=749	140
Figure 34. 'MeFi' Pronunciation by Country of Residence – 2012 Panel Data, N=751	140
Figure 35. Age Distribution of Survey Participants	146
Figure 36. 'MeFi' Pronunciation by Age Group – 2010 Survey Data, N=2,462	147
Figure 37. 'MeFi' Pronunciation by Age Group – 2012 Survey Data, N=1,906	147
Figure 38. 'MeFi' Pronunciation by Age Group – 2010 Panel Data, N=759	148
Figure 39. 'MeFi' Pronunciation by Age Group – 2012 Panel Data, N=763	148
Figure 40. Example of a Text-Based MetaFilter Message Chain	156

Figure 41. Example of a Multi-Modal MetaFilter Message Chain	156
Figure 42. Timeline of Notable Events in the Enregistration of ‘MeFi’	158
Figure 43. Early Stage: Frequency of ‘MeFi’ by Month, 2000–2001	160
Figure 44. S-Shaped Curve Showing Adoption of ‘MeFi’ by Year	160
Figure 45. Social Engagement Factors and the Enregistration Timeline	163
Figure 46. Participants’ Year of Joining MetaFilter – 2010 and 2012 Surveys	164
Figure 47. MetaFilter Subsite Visitation Frequency – 2010 Surveys, N=2,521	166
Figure 48. MetaFilter Subsite Visitation Frequency – 2012 Surveys, N=1,957	166
Figure 49. Frequency of ‘MetaFilter’ by Subsite and Year	169
Figure 50. Frequency of ‘MeFi’ by Subsite and Year	169
Figure 51. Frequency of ‘MeFite’ and ‘MeFites’ by Subsite and Year	169
Figure 52. ‘MeFi’ Pronunciation by AskMetaFilter Visitation – 2012 Survey Data, N=2,477	170
Figure 53. ‘MeFi’ Pronunciation by MetaTalk Visitation – 2010 Survey Data, N=2,473	171
Figure 54. ‘MeFi’ Pronunciation by MetaTalk Visitation – 2012 Survey Data, N=1,921	171
Figure 55. Self-Reported Podcast Listening Frequency – 2010 and 2012 Survey Data	176
Figure 56. Self-Reported Podcast Listening Frequency – 2010 and 2012 Panel Data	176
Figure 57. ‘MeFi’ Pronunciation by Podcast Listening – 2010 Survey Data, N=2,477	178
Figure 58. ‘MeFi’ Pronunciation by Podcast Listening – 2010 Panel Data, N=756	178
Figure 59. Self-Reported Meetup Attendance Frequency – 2010 and 2012 Survey Data	182
Figure 60. Self-Reported Meetup Attendance Frequency – 2010 and 2012 Panel Data	182
Figure 61. ‘MeFi’ Pronunciation and Meetup Attendance – 2010 Survey Data, N=2,469	184
Figure 62. ‘MeFi’ Pronunciation and Meetup Attendance – 2012 Survey Data, N=1,920	184
Figure 63. Social Engagement and ‘MeFi’ Pronunciation – 2010 Survey Data, N=603	186
Figure 64. Indexical Field For ‘MeFi’	199
Figure 65. Indexical Field For 1b - [mifaɪ]	200
Figure 66. Indexical Field For 3a - [mɛfi]	200
Figure 67. Indexical Field For 3b - [mɛfaɪ]	201
Figure 68. Indexical Field For 4b - [maɪfaɪ]	202
Figure 69. 2010 Survey MetaTalk Post	230
Figure 70. 2012 Survey MetaTalk Post	231
Figure 71. 2010 MetaFilter Survey Information and Consent Page	232
Figure 72. 2012 MetaFilter Survey Information and Consent Page	233
Figure 73. The 2010 MetaFilter Survey	234
Figure 74. The 2010 MetaFilter Survey (continued)	235
Figure 75. The 2012 MetaFilter Survey	236
Figure 76. The 2012 MetaFilter Survey (continued)	237
Figure 77. The 2012 MetaFilter Survey (continued)	238

List of Tables

Table 1. CMDA Etic Grid Classification of Medium and Situational Facets of MetaFilter	55
Table 2. Examples of Elements from the MetaFilter Register	72
Table 3. Orthographic Features of ‘MeFi’ and ‘MeFite’	76
Table 4. Phonotactically Viable Phonetic Realizations of ‘MeFi’	80
Table 5. M-Set Variants and Their Vowel Codes	81
Table 6. Data Subsets	110
Table 7. Change in Variant Choice for ‘MeFi’ Over Time	114
Table 8. Change in Variant Choice for ‘MeFite’ Over Time	114
Table 9. Shift in Variant Choice for ‘MeFi’ Over Time – Panel Data	115
Table 10. ‘MeFi’ Variant Switching in Panel Data Participants	116
Table 11. Change in Variant Choice for ‘MeFite’ Over Time – Panel Data	117
Table 12. ‘MeFite’ Variant Switching in Panel Data Participants	117
Table 13. Self-Reported Language Experience from the 2012 Survey	134
Table 14. Gender Distribution of MeFites in the 2010 and 2012 Surveys	150
Table 15. Characteristics of Social Engagement Measures	162
Table 16. Positive and Negative Evaluations of Variants	189
Table 17. Analogy by Orthography	240
Table 18. Summary of Panel Data as Compared to Survey Data by Various Measures	241
Table 19. Summary of Change Over Time by Various Measures	242
Table 20. Summary of Pronunciation Outcomes by Various Measures	243
Table 21. Binary Logistic Regression Model Results – 2010 Model Data	244
Table 22. Binary Logistic Regression Model Results – 2012 Model Data	245

List of Examples

Example 1. Rationales Citing Linguistic Variety as an Influence on Pronunciation Choice . . .	133
Example 2. Innovation of ‘MeFi’	159
Example 3. Early Adoption of ‘MeFi’	159
Example 4. Comment #19 in 2010 Survey MetaTalk Discussion Post	173
Example 5. Comment #81 in 2010 Survey MetaTalk Discussion Post	173
Example 6. Podcast Transcript Excerpt: Episode #2	179
Example 7. Comment #32 in MetaFilter Podcast Post: Episode #2	180
Example 8. “meh”-fee versus mee-fie	193
Example 9. Analogous Words Plus Evaluation	193
Example 10. ‘MeFi’ by Association	193
Example 11. Indirect ‘correctness’	194
Example 12. “Me Phi Me” Hip-Hop Performer and Musical Group: 2006 Post	194
Example 13. “Let’s Call the Whole Thing Off” Song by the Gershwins: 2006 Post	195
Example 14. “Raymond Luxury Yacht” in Monty Python’s Flying Circus: 2006 Post	195
Example 15. “Family Circus” Comic Strip in the Sunday Funnies: 2010 Post	195
Example 16. MetaFilter: Taglines	197
Example 17. Special Snowflakes	197

Acknowledgements

“if knowledge is worth having, it is worth sharing.” — Cameron et al., 1992, p. 24

This is a thesis mostly about a single word. Specifically, ‘MeFi’, and what it means to groups of people. As such, I couldn’t have written this without any and all of them. I am indebted to each and every MeFite of MetaFilter, as well as to all of the linguists, advisors, techies, stats gurus, friends, and family members who helped me along the way.

Even though I was aided by the help of many, I also couldn’t have done this unless I really understood what ‘MeFi’ meant to me. And that was (and still is) forever changing, evolving, expanding. Some days ‘MeFi’ was a moving target; others it was my stable respite; the unwavering focus of my day, but sometimes my escape at the end of it. ‘MeFi’ ran the gamut from being this nebulous global phenomena called “online community” to a random string of disembodied sounds...which often became stranger and more unrelatable as they tumbled over and over in my puzzled head. How to make sense of it all then?

I turned it in on itself. Every challenge I faced informed my understanding ‘MeFi’. For example, my fear of public speaking was partially overcome by the realization that I could use every presentation opportunity to poll my audience — how did they pronounce ‘MeFi’? What did ‘MeFi’ mean to them? In learning from several people from my research community about how to build and use databases, I gained some valuable perspectives on how people parse and structure data, specifically, ‘MeFi’ data. This sort of meta-‘MeFi’ analysis applied to all things, and my confidence and my thesis were both the better for it.

I am also grateful for many others who had a more direct hand in shaping the outcome of this project, and ultimately ensuring my success. My two supervisors, Carmen Llamas and Dominic Watt gave me solid feedback throughout these years, always making great suggestions, challenging my ideas, and preparing me for every possible outcome by forcing me to think through all of the possibilities.

My linguist friends provided an endless source of technical, emotional, and caffeine-fueled guidance of all sorts — I cannot stress the importance of this trifecta enough! Three people embodied this rock-solid support most exceptionally — Dr. Marta Szreder, Tom Devlin, and Clare Cunningham. Through many mornings, days, and nights we pondered, pedanted, and drank coffees (and sometimes whiskies). The laughter has kept me sane, thank you.

There is an even smaller circle of those who I consider my nearest and dearest and I have been truly blessed to experience these years with their love and encouragement to carry me through. First and foremost, my Mom and Dad — their ceaseless positivity, always only a Skype call away, meant that I had an endless source of energy to get through the challenges,

the frustrations, the winters. Marta — worthy of a second mention, as she has been so much more than a linguist friend — someone who sees the world in similar, hilarious ways as me, and has been there to make me laugh every step of the way. And to Alex — I wish I'd met him earlier on in this journey. Throughout this last year and a half he has consistently supported me with his kindness, creativity, attentiveness, and sensibilities — comforting me when I've been low, celebrating with me when I've found success, and encouraging me to “kill it with words” when I felt I couldn't make peace with writing. I needed all of that.

Lastly, these acknowledgements would be completely lacking without reference to the group that started this all. MetaFilter has changed my life and I don't think I could ever give back what the people of this wonderful place have given to me. The MetaFilter community, and especially MetaFilter moderators Matthew Haughey, Josh “cortex” Millard, pb, and Jessamyn West, who made this research possible and at many points along the way took extra time to help me with data requests, or gave advice and support. Several other MetaFilter participants also went above and beyond in helping me with this project and deserve special mention: Adam Gutteridge, James Young, Kanani Kauka, Karthik Ram, Toby, Richard... there are countless others, who I shall not mention; you know who you are. Thank you again.

Author's Declaration

I hereby declare that I, Kimberly Ann Witten, am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1: Introduction

Since the beginning of the world wide web, people have gathered together in online spaces to discuss topics, share ideas, and form communities. Within these communities, online participants have found new and creative ways to express who they are, and to differentiate themselves from others. Through this interaction, they have also established who they are collectively, and ways in which that group identity is distinct from other online collectives.

In a primarily text-based medium, such as is often the case in computer-mediated communication (CMC) environments, distinctions between people and/or between groups are usually communicated via the written word. In face-to-face (FtF) environments, these distinctions may be communicated in a range of additional ways, including, but not limited to: speech, prosody, gesture, and the visual appearance of the speaker. While the modality used to communicate is one of the key differences between the means with which people establish identities within these media, neither environment is compromised in its ability to convey information about identities. This is because, like members of FtF communities, internet users who participate regularly in online communities can exchange ideas, take stances, and form social bonds. Through this interaction, participants create and share a history. That history includes the events that those who belong to the group have participated in, situated in the always-changing environmental context in which they have experienced those events. While this process is mediated by language, the medium in which language occurs (e.g., text, speech, both text and speech, other communicative means) does not necessarily prevent or otherwise hinder this shared history from unfolding. From this history, a collective register emerges, including linguistic conventions, stereotypes, in-group behaviors, and other features that belong to the community and are reflective of its identity.

More generally, this process of establishing a register and linguistic conventions that are associated with (e.g., are *indexical* of) a group and its members is known as *enregisterment* (Agha, 2003, 2005). This is a relatively new concept in sociolinguistics, but it describes a process that becomes familiar to most people from early ages, when we first become aware of distinctions such as stereotypes, labels, in-groups, and other social divisions. However, for most of us, how these stereotypes, labels, and conventions come to be is often opaque.

This research is the first in-depth case study of enregisterment occurring online, in one of the oldest and most respected online community weblogs, MetaFilter (MetaFilter.com, established in 1999). Owing to the text-based medium, the unique social structure of MetaFilter, and the development of a novel mixed-methods approach employed here, this

investigation allows the entire process of enregisterment to be observed through the site's written record of events and communicative exchanges.

On this global discussion forum (MetaFilter.com), participants have had an ongoing negotiation about their collective nickname, 'MeFi', and the name they use to refer to themselves, 'MeFites'. While orthographically consistent with one another, the forms 'MeFi' and 'MeFite(s)' each have at least eight phonotactically viable pronunciation variants for native English speakers. The terms 'MeFi' and 'MeFite(s)' will be henceforth referred to as the 'M-Set' throughout this research. This allows both variables to be referred to as a collective unit, as well as providing a way to refer to these terms in speech without biasing hearers with a particular pronunciation.

The M-Set is also representative of a new category of word forms. *Netologisms* are words that are derived from or primarily used in CMC spaces (Witten, 2012). These new words may display pronunciation ambiguity, often owing to unique features such as CamelCase (mixed-case lettering, e.g., 'PowerPoint', 'iPhone', 'MetaFilter'), or non-alphabetic characters (e.g., 'L337SP34K', 'G2G', '.gif', '#', '@'), or grapheme sequences which are unusual or disallowed in English orthography (e.g., 'Imgur', 'Flickr', 'Tumblr'). Netologisms may also be acronyms (e.g., '.gif', 'LOL', 'SQL'), or other types of abbreviated forms (e.g., 'WiFi'). Lastly, netologisms may lead to pronunciation ambiguity simply because there are no standardized spoken reflexes of them yet, and grapheme-phoneme correspondence (GPC) rules may lead to more than one outcome (e.g., 'doge', 'Linux', 'vi', 'Ubuntu'). 'MeFi' and 'MeFite(s)' share many of these properties (e.g., CamelCase, being an abbreviation/clipping, having ambiguous GPC mappings), making them ideal examples for studying linguistic variation in CMC. However, netologisms and other ambiguously pronounceable words with similar properties can be found in other media, and in other languages as well. Therefore, while this research is a case study of just two English words used in CMC, it can help shed light on how other, similar words might be learned, pronounced, and enregistered in other communicative spheres and linguistic varieties.

1.1 Background

The first use of 'MeFi' was seen on the website in 2000, and since that time the nickname's pronunciation has become a recognized shibboleth within the community. Participants have been debating the "correct" pronunciation of 'MeFi' (and 'MeFites') for over fourteen years. As opposed to most FtF environments, however, the nature of the text-based medium and the archive of comments and posts on the site have allowed the history of these discussions to be perpetually revisited and reviewed.

While MeFites generally cannot be sure what the actual distribution of ‘MeFi’ pronunciation is, more involved members in the community may have an accurate sense of what some of the prevailing pronunciations and attitudes are, based on their observations of others’ comments and their involvement in off-site MetaFilter activities, such as attending face-to-face meetups or listening to the MetaFilter podcast. However, the majority of MeFites have never attended meetups or heard podcasts, and/or may have arrived at their pronunciations of the M-Set without any linguistic difficulty. They may become aware of the variability of the M-Set through discussions on the site, usually in the form of playful linguistic banter, such as when participants mock each other about pronunciations they recently heard at meetups or know others to use based on previous discussion threads. These discussions usually contain phonetic respellings of variants (e.g., ‘Meffy’, ‘may-fight’, ‘meeee-fie’) and evaluative statements about those variants (e.g., “*sounds cute, like a pet name*”, “*may-fight is argumentative and that’s what we do here*”, “*it’s all about ME and FIE on you for disagreeing!*”).

While the debate has continued on for more involved participants, new members are continually joining the community and therefore will inevitably encounter the M-Set for the first time. These new participants will probably not initially be aware of the debate over pronunciations, or that pronunciation variation exists. It seems that it is usually not until a MetaFilter participant hears or reads about a pronunciation that differs from their own that any conscious reflection about pronunciation occurs.

To MetaFilter participants, the task of selecting a pronunciation is not a particularly hard one, and in most cases it is not done under specific instruction or with close attention paid to the task. Nevertheless, there are a lot of conflicting grammatical rules and social influences to consider in selecting a pronunciation, and these factors may be implicitly resolved by participants with little conscious thought about their doing so.

The importance of the pronunciation of the M-Set is also relatively minimal, as is the risk of negative consequences of making a choice. This, however, does not mean that the matter itself is trivial. On the contrary, this particular linguistic resource is part of a much larger array of forms that individuals can use to socially position themselves, take stances, or construct identities. The investigation of a relatively low-risk example such as the M-Set can be a valuable resource for understanding other, similar matters where the stakes are much higher (e.g., political names or territories, controversial and inflammatory terms, etc.), but where those variables are heavily bound by the contexts that make the very decisions such a high risk in the first place. That is, in higher-risk naming situations, it can be extremely hard to disentangle the political factors and specific history that gives the debate its importance and social significance relative to the other factors that may contribute to various outcomes in the debate.

The matter of pronunciation is also important in the sense that each choice is the outcome of a specific grammatical path traveled, indexical of participants' specific sociolinguistic backgrounds and communicative goals. The pronunciation variant(s) participants choose is at least partly reflective of their identities, as well as partly an outcome of the identities they wish to present of themselves and for the community (e.g., how they think 'MeFi' should be perceived and understood via its identifiable sound pattern and all of the associations that go with that pattern).

As such, there are several internal and structural linguistic factors that may influence the pronunciation of the M-Set. These include, but are not limited to:

- The presence of CamelCase (mixed-case letters) in 'MeFi', which may visually prompt an open-syllable stressed vowel, as in [mi-] or [meɪ-].
- The frequency of words in English that have strong, consistent mappings of <e>→[ɛ] in a stressed syllable, such as 'bet', 'belly', 'deli', 'menu,' 'met', 'wet'.
- The frequency of the word 'me' in English, perhaps prompting a [mi-] pronunciation; also the favorable semantic associations of 'me' for many users in characterizing the site or their involvement with it.
- Consistency with the pronunciation of 'MetaFilter', from which the abbreviation 'MeFi' originates, prompting a [mɛ-] or [meɪ-] pronunciation.
- Analogy with other forms ending in <i>, such as 'WiFi', 'Semper Fi', 'wiki', 'kiwi', etc., prompting either a [faɪ(t)] or [fi(t)] ending.

These are just a few of the internal and structural factors that will be expanded upon throughout this thesis.

Personal associations people have and other social factors may also come into play, such as how much participants interact with each other and in which ways. These serve to further complicate the picture, showing that pronunciation variation operates at multiple levels of linguistic structure, from mental organization of language in the mind to sociolinguistic features such as dialect, usage norms and peer influence.

Regardless of whether or not community participants can easily hear spoken instances of sociophonetic variables, their awareness and understanding of the variation is a way in which they can make sense of their linguistic worlds. Community participants have a desire to contextualize their experiences and their understandings of others through language. This is directly evident on MetaFilter in metalinguistic discussions surrounding pronunciation of the M-Set, usernames, cities, or people who are the subjects of posts, and general discussions about speech, behavior, or backgrounds of people on the site. In sum, MeFites want to know who their fellow community members are, what they sound like, and why.

The M-Set variables, therefore, are central to issues of identity, accommodation and

social capital. While the latter two topics are beyond what can be covered by this research, they have relevance and will be touched upon where it makes sense to do so. The connections between these sociolinguistic variables and identity are the main focus of this research, and will be explored through the lens of enregisterment — the ways in which the M-Set is indexical of the values of individuals and the group, and how those associations came to be known as such.

1.2 Research Questions

This study aims to provide an account of an online community of practice negotiating in-group conventions in a primarily text-based environment. This initial exploration sets the foundation to then give an account of enregisterment in progress, so as to demonstrate how the process works. Using a combination of quantitative and qualitative measures, the process of enregisterment can be tracked and quantified, thereby providing a model for future research. Lastly, this case study also demonstrates one example of how English speakers pronounce new names that enter the language through the text-based medium of CMC, where often no widely recognized or standardized pronunciations exist. These outcomes can be exemplified through the following three research questions addressed in this thesis:

1. How do speakers pronounce ‘MeFi’ and ‘MeFite(s)’? What are the statistically significant social factors that correlate with pronunciation variants of these names as used by members of a text-based community?
2. How do the M-Set variables acquire social meaning (e.g., associations between the form and the group it represents, stereotypes, and other meanings)? How does this acquisition operate in a text-based medium?
3. How do MetaFilter participants use the M-Set variables to co-construct a group identity? How do participants use the M-Set variables to assert things about themselves, i.e., to construct their individual identities?

These research questions, when combined, give a multi-faceted account of the ongoing enregisterment of word forms within an online community. Owing to the modality and medium in which the participants primarily communicate (i.e., text-based interaction in CMC), the entire enregisterment process for the M-Set can be observed and analyzed. To date, no other sociolinguistic study has been undertaken to account for enregisterment this comprehensively and in such optimal conditions — naturally occurring interaction (i.e., non-experimental) in a clearly delimited community with a written record of its history and communicative exchanges.

1.2.1 Hypotheses and Expected Findings

An investigation such as this has necessitated the adaptation of previously successful

methodologies from other, related investigations. As such, these initial hypotheses have not been as fully circumscribed as one might find in other studies, where both the methodologies and the environments in which they are employed are well-researched and documented. Therefore, all of the hypotheses suggested in this section are more general in nature, but will be expanded upon further in later chapters of the thesis.

To address the first research question, concerning which social factors may bias participants in M-Set pronunciation choices, it is hypothesized that participants' linguistic choices will be partially influenced by their linguistic backgrounds (e.g., their dialect, language experience, geographic origin and country of residence), but also partially influenced by the amount and type of their social engagement on MetaFilter. Participants who are deeply involved in the community, and who spend time participating in areas of the site which pertain to discussions of community-related matters, are likely to be more aware of the goings-on of fellow participants, popular stances in the debate, and influential authority figures' pronunciation choices. This knowledge is predicted to bias these participants' linguistic choices toward those they deem as important, and/or toward what they perceive the emerging standard pronunciation(s) to be. Less-involved participants (e.g., "lurkers" and those who read the site infrequently and rarely make contributions) are predicted to be less aware of community norms, others' pronunciations, or debates about M-Set pronunciation. These participants are therefore expected to choose pronunciations that are largely influenced by their personal linguistic histories (e.g., their dialects and language experience).

For the second research question, regarding how the M-Set variables might acquire social meaning (i.e., indexicalities), this is predicted to be primarily occurring through the messages that are exchanged by participants communicating online. That is, the ideas, stances, and evaluations that online participants share with each other in the text-based medium inform the impressions that they have about the variables, as well as help co-create the written record of those publicly available conceptions of the variables. The types of associations that are made and shared with others are expected to be qualitatively different than those that would be formed by participants who communicated in primarily FtF modalities; the structure of the CMC environment allows ideas to be more explicitly communicated via text, but they are not easily linked to visible characteristics of those who participate.

Finally, the ways in which these variables are employed in the construction of identities is hypothesized to be a function of the salience of popular or otherwise notable stances. It is predicted that the more prominent stances about the M-Set within the community are linked with particular participants in the community, as well as become representative of particular characteristics and values of the community itself.

1.3 Relevance and Broader Implications

Very few large-scale, in-depth studies of enregisterment — occurring in FtF communities or otherwise — have been undertaken thus far. The procedures described in **Chapter 4: Methodology, p. 89**, outline a comprehensive way to collect and assess several types of data for these investigative purposes. This research model allows for an approach to understanding the ways we create stereotypes, make meaningful associations between words and identities, and develop and acquire registers. Therefore, the success of the approach and its outcomes are crucial for advancing theory in sociolinguistics and related disciplines. Additionally, sociolinguistic researchers studying enregisterment processes may gain further insight into CMC environments to explore and various socio-structural factors to consider when conducting research in these areas.

Further still, while there have been many informal studies or questionnaires about the pronunciation of neologisms, there have been no in-depth published studies of global orientations to particular new words or names which are ambiguously pronounceable thus far. As there is much ambiguity in English spelling-to-sound correspondences, it is important to provide a sociolinguistically-based starting point and methodology for exploring these issues. This is done in the hope that others will adopt, adapt and improve the methodologies for their own studies of words, and we can collectively add to the knowledge base of the social underpinnings of our linguistic choices.

1.4 Thesis Structure

Following this introduction, **Chapter 2: Literature Review, p. 9** covers a range of topics from the fields of education, CMC, sociolinguistics, and onomastics. This chapter sets the groundwork for understanding MetaFilter as an online *community of practice* (Lave and Wenger 1991; Wenger, 1998), and introduces and explores concepts such as *enregisterment* (Agha, 2003, 2005) and *indexicality* (Labov, 1972; Silverstein, 2003; Johnstone, Andrus, and Danielson, 2006).

Next, in **Chapter 3: MetaFilter and the M-Set Variables, p. 48**, I will provide historical and cultural background information necessary to understand and contextualize the MetaFilter community. This insight is based on over eight years of daily social engagement with the site, in addition to participatory research specifically focused on addressing the research questions. Previous studies of MetaFilter will also be reviewed in this chapter. The M-Set variables will be introduced, with a detailed phonetic primer and more general review of English pronunciation, as they both relate to the pronunciation of ‘MeFi’ and ‘MeFite(s)’.

Chapter 4: Methodology, p. 89 will outline the approach and procedures used in this

research to collect and analyze data. These methods primarily include the implementation of large-scale surveys, but also borrow some techniques from ethnographic and discourse analysis methods. These data are analyzed using both quantitative and qualitative approaches.

In the first of two data presentation chapters, **Chapter 5: Data Results, p. 110**, demographic results from the surveys are presented. This encompasses possible influences relating to linguistic experience, native language, country of residence, and indicators of participants' levels of metalinguistic awareness.

In the second data presentation chapter, **Chapter 6: Enregisterment, p. 153**, results pertaining to indicators of social engagement (such as frequency of website visitation, and frequency of attending meetups or listening to the podcast) are shown, and the indicator's relationship to M-Set pronunciation is explored. Additionally, this chapter investigates ways in which M-Set variants have acquired social stereotypes through the exploration of *message chains*; the mechanism by which evaluations about features or variables are disseminated in a population (Agha, 2003, 2005). The increased frequency of 'MeFi' use over time is shown through timelines and other data visualizations, and key points in M-Set history are highlighted and discussed.

In **Chapter 7: Discussion & Conclusion, p. 205**, all of the results across both data presentation chapters are summarized. An interpretation of these results is given, with special focus on how these results relate to enregisterment processes. Research limitations are also discussed. Lastly, this chapter provides a final statement on the aims of this study and the findings related to the research questions. The implications of this research for the areas of enregisterment, indexicality, and onomastics are considered, followed by some possibilities for future research.

Chapter 2: Literature Review

2.1 Introduction to Literature Review

The models and theories selected to support this thesis — while originating from different subfields in linguistics (and sometimes outside of linguistics entirely) — all share a commonality. From communities of practice to social capital and beyond, these constructs recognize the relationships between micro- and macro-levels of analysis. The connections between micro-level features (such as a linguistic unit, an instance of interaction, or an individual) and macro-level features (such as the creation of norms, the process of enregisterment, and group identity) will be showcased wherever possible.

2.2 Models of Community

Three sociolinguistic models to describe communities have prevailed in recent years: *communities of practice*, the *speech community*, and *social networks*. In an extensive review of the community of practice model, it will be shown that this is a best fit in accounting for the salient properties of the MetaFilter community with respect to the two sociophonetic variables investigated in this thesis. Following this exploration, some concepts from social network theory, such as the value of weak ties in the spread of information and innovation, are also relevant to the research and will be explained in this section.

2.2.1 Community of Practice — Definition

The community of practice model (CoP) was first proposed by two education researchers, Jean Lave and Étienne Wenger (1991; Wenger, 1998). They formally define a CoP as follows:

“An aggregate of people who come together around mutual engagement in an endeavor. Ways of doing things, ways of talking, beliefs, values, power relations – in short, practices – emerge in the course of this mutual endeavor.”

This definition allows for a range of analytic possibilities, from the examination of smaller, micro-units such as the individual or a specific feature of a practice, to the more global, macro-units such as the practice itself, the group identity or the agreed-upon norms. The CoP model can account for both qualitative and quantitative data, as both types of data result from any shared practice (Bucholtz, 1999, p. 221).

A CoP model has three main criteria. First, involvement in the community includes some source of coherence of its members, described as *mutual engagement*. This consists of relationships and activities organized around what it is that the community does and it is essential to any practice. *Joint enterprise* reflects mutual engagement in that it is something

that is negotiated by community members — they all have a stake in the enterprise, and through their participation and sharing of ideas, they decide what’s important to them. The third criterion — a *shared repertoire* — involves not just objects or requirements of group membership, but also includes the terminology (part of its register), stories, inside jokes, ways of doing things, symbols, concepts and all other things the community can produce or are a part of the practice itself (Lave and Wenger, 1991; Wenger, 1998).

These three tenets of the CoP model have been successfully applied to MetaFilter in two previous studies (Ali-Hasan, 2005; Silva, Goel, and Mousavidin, 2008). Both papers aimed to describe features of the social structure of MetaFilter and explain its success as a long-standing and thriving online community. These studies are covered in greater detail in **3.3.1 Previous Study of MetaFilter as a CoP, p. 64**. This thesis expands on these researchers’ work by applying more aspects of the CoP model, incorporating related theories (e.g., message chains, social capital — both addressed in this chapter), and using the model as a foundational framework for contextualizing and interpreting the linguistic data in this research.

2.2.1.1 *Additional Criteria in Defining a CoP*

While MetaFilter easily meets the three main criteria for a CoP (i.e., mutual engagement, joint enterprise, shared repertoire), there are additional indicators of a viable CoP to consider. Wenger (1998, pp. 125–126) has outlined several of these; they include, but are not limited to:

- sustained mutual relationships — harmonious or conflictual
- shared ways of engaging in doing things together
- the rapid flow of information and propagation of innovation
- very quick setup of a problem to be discussed
- local lore, shared stories, inside jokes
- the ability to assess the appropriateness of actions and products
- jargon and shortcuts to communication as well as the ease of producing new ones
- certain styles recognized as displaying membership

Examples and analyses throughout this thesis will demonstrate many ways in which MetaFilter displays all of these indicators. This further establishes MetaFilter as a CoP, as the term is defined by the educational researchers Lave and Wenger (1991) who originally created and codified the model. As such, my perspective on CoP and its application in this research differs somewhat from the definition and usage that is commonly used in post-variationist and interactional linguistics research, where the model is often applied using ethnographic methods to smaller groups of people who form their practice through FtF interactions (Bucholtz, 1999; Eckert, 2000; Moore, 2006; Lawson, 2011). The approach I’ve taken is less focused on ethnographic methods (although includes extensive participatory research over a number of years) and involves more quantitative methods, applied to a larger, globally-

dispersed group of individuals who do not necessarily meet FtF. As a result, MetaFilter may be more aptly characterized as CoP that is comprised of a *constellation* of interconnected sub-practices (Wenger, 1998).

2.2.1.2 *Constellations of Practices*

There is an important distinction to be made between a community itself being a single practice versus it being a *constellation of practices* (Wenger, 1998, pp. 126–131). This distinction exists because, as Wenger explains, some communities are too broad, diverse, or diffuse to be meaningfully characterized as a single, unified CoP. These configurations consist of interconnected practices, i.e., a constellation. The features that connect these smaller practices can vary, but Wenger (1998, p. 127) has identified several key ones (not a comprehensive list):

- sharing historical roots
- having related enterprises
- having members in common
- sharing artifacts
- having overlapping styles or discourses

The various subsites of MetaFilter, and the different ways that MeFites can participate in off-site MetaFilter-related activities (e.g., meetups, MeFight Club (the gaming-focused spin-off site), holiday gift exchanges, and music or cookie swaps), meet these constellation criteria and therefore provide compelling evidence for MetaFilter as a constellation of related practices.

Wenger (1998, p. 129) elaborates on the constellation idea further by explaining that viewing a practice as a constellation rather than singularly requires that continuity among all of the sub-practices must be viewed through the interactions that occur among those practices. Most relevant to this study is the type of interaction that includes discourse elements that travel across the boundaries of the practices and are used for various purposes such as negotiating, reconciling perspectives, and taking stances on topics (Wenger, 1998, p. 129). The M-Set is just one of many discourse elements that are used as resources in this way, and serve to help unify the constellation.

Furthermore, it is important to stress the idea that agreement and harmony are not required to achieve this unifying effect. In the case of the M-Set, the disagreements about pronunciation motivate increased discourse across subsites and various communication modalities and channels (e.g., text and speech through onsite and offline means). This idea of disharmony having benefits can hold true for any community of practice, as having contrasting views helps reinforce and bring to light what it is that the participants value; the disagreement is something they all share together (Wenger, 1998, p. 45).

2.2.1.3 *Learning, Meaning, Participation, and Reification*

Various aspects of the CoP model were expanded more fully in Wenger's (1998) book, *Communities of Practice: Learning, Meaning, and Identity*. The central ideas of his detailed vision of the CoP model focuses on *learning* and *participation*. Learning in this context is a social phenomenon, and not the acquisition of facts in a didactic sense. This view of learning sees the acquisition of knowledge as a matter of competence in social settings, and knowing as a matter of engaging with those social environments. As it relates to linguistics, learning involves understanding the language norms of the community, the community register, and the linguistic elements of the shared repertoire.

Learning produces *meaning*; it creates purpose in our behaviors and goals (Wenger, 1998, p. 4). Meaning in this sense is part of an ongoing process; it is something we negotiate within ourselves and with others (Wenger, 1998, p. 52). As individuals, we negotiate the meaning of things, experiences, ideas, behaviors, and events repeatedly. We revisit discussions and arguments and we engage in repeated behaviors. In doing so, we "*reinterpret, modify, and confirm them*" (Wenger, 1998, p. 52). This is part of our ongoing interaction with the world, and it provides us with new resources for continued learning, and new meanings for those resources.

Throughout this thesis, the CoP sense of meaning will often be referred to as 'social meaning'; this helps disambiguate the term from its dictionary definition, as well as connecting it to the understanding of 'meaning' that is referred to in studies of indexicality (see 2.5.3 *Indexicality*, p. 28).

Learning and the negotiation of meaning are key to a CoP, as participants rally around what it is that the group does. It is not language that is central in the minds of those who interact online, but what language allows participants to exchange or to do online. It is through language that they share ideas, learn, locate, and define each other and the group. Therefore, language is a means to participate in a practice, but it is not dominant over it in a way that ignores or minimizes the value of the practice itself.

Participation, as Wenger (1998, p. 55) defines it, describes the "*social experience of living in the world, in terms of membership in social communities and active involvement in social enterprises.*" We cannot escape participation; we are part of the places we visit and the people we interact with, in some way, however small or seemingly inconsequential it may be to us. Through our participation in the world around us, we come to learn and we may help negotiate the meanings of the things that matter to the communities that we are a part of (although negotiation of meaning is not a requirement of participation). As such, references to 'participation' in this research are defined by membership rather than by observable activity;

the latter of which will be referred to as ‘activity’ or ‘(social) engagement’.

Reification describes the process of transforming a thing — be it a behavior, a way of doing something, an idea, a tool, etc. — into having its own, less abstract existence in the minds of those who find it meaningful. Through reification — made possible by participation — a shared repertoire emerges, as participants learn and negotiate the meanings of the things they do and the resources they use to do them. As Wenger (1998, p. 66) pithily puts it, “*Participation and reification both require and enable each other.*” Furthermore, our sense of self — our identity — includes these reified things, as they allow our participation in practices to continue (1998, p. 70).

Throughout this thesis, I will provide descriptions and examples that demonstrate the learning that occurs on MetaFilter and the meaning that is derived from those experiences. This is deeply connected to MetaFilter members’ participation with various practices within the constellation, and results in the reifications that comprise their shared repertoire.

2.2.1.4 *Types of Participation*

There are many ways to participate within a community, and not all of these modes result in directly observable social interaction. This is important to consider, as the less visible or less easily-quantifiable modes should not be overlooked or be seen as less valuable toward the development of the identity of the community or of the individuals within it.

Wenger (1998, p. 173) has defined three broad categories of participation, which he describes as *modes of belonging*:

1. **engagement**: this includes the “*active involvement in mutual processes of negotiation of meaning.*” For this research on MetaFilter, engagement is considered to be the directly observable social interaction, and will be referred to as ‘social engagement’ throughout this thesis. This includes, but is not limited to, participants comments, posts, and other visible forms of activity on the site and in related practices within the constellation.
2. **imagination**: how participants view their own — and others’ — experiences and positions within the practice. Individual imaginations of who we are help to form the collective imagination of the group. Imagination is a less visible form of participation, but is assessed in this research through participants’ stances about the community, other participants’ identities, their own identity and place within the community, and their assertions about the M-Set variables.
3. **alignment**: the coordination of our energy, activities, and behavior to fit in with the community and beyond. In this research, alignment is viewed through metrics such as “favoriting” and quoting others’ posts and comments, as well as through similar demonstrations of agreement. The concept of *message chains* is an important mechanism of alignment, and will be defined later this chapter.

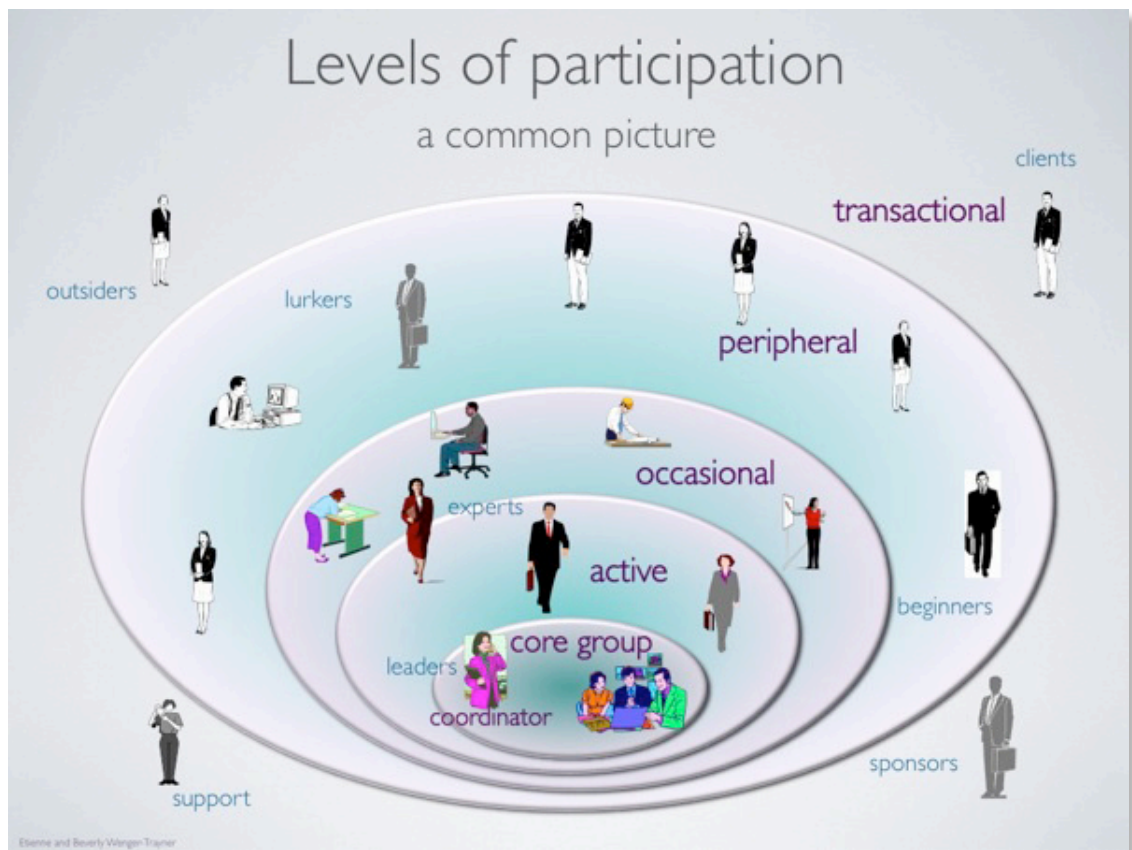
These modes of belonging are not found in every member of a practice, nor are they

in equal measure for any member at any given time. The practice itself also does necessitate that each mode be equally represented (Wenger, 1998, p. 183). In the MetaFilter constellation of practices, the sub-practices have distinct character, partly owing to their varying representations of these modes. For example, alignment is much more visible through favoriting posts and comments (each post or comment has a visible favorite count in its byline) on MetaFilter subsites than it is at meetups, where the FtF mutual engagement aspect takes higher precedence.

Lastly, both imagination and alignment do not require mutual social engagement to exist; MetaFilter members can envision themselves and the rest of their community without input or assistance from others. Similarly, they can behave in ways that align themselves with the practice or with certain stances, individuals, and norms without the influence or directive of others (Wenger, 1998, p. 178). That is to say, while authority figures on MetaFilter may influence alignment, they do not enforce alignment with their ways of doing things. Additionally, this would be difficult to enforce, as authority figures do not always agree — especially when it comes to linguistic matters, such as M-Set pronunciation.

2.2.1.5 *Levels of Membership and Trajectories*

Just as each mode of belonging may not be represented equally in each practice, each individual within a practice may not exemplify each mode in the same combination or to the same degree. This may inform the level of membership that each individual has within a practice, and how their contributions to that practice are perceived. Furthermore, each mode and level provides different potential for learning and shape the character of the meaning that may result from those experiences (Wenger, 1998, p. 183).

Figure 1. Wenger's (2011) 'Levels of Participation' in a CoP

Source: Wenger, 2011, <http://wenger-trayner.com/resources/slide-forms-of-participation/>

Levels of participation are defined as follows (as copied directly from Wenger, 2011):

- **Core group:** a relatively small group of people whose passion and engagement energize and nurture the community.
- **Active participants:** members who are recognized as practitioners and define the community (though they may not be of one mind as to what the community is about).
- **Occasional participants:** members who only participate when the topic is of special interest, when they have some specific to contribute, or when they are involved in a project related to the domain of the community.
- **Peripheral participants:** people who have a sustained connection to the community, but with less engagement and authority, either because they are still newcomers or because they do not have as much personal commitment to the practice. These people may be active elsewhere and carry the learning to these places. They may experience the community as a network.
- **Transactional participants:** outsiders who interact with the community occasionally without being members themselves, to receive or provide a service or to gain access to artifacts produced by the community, such as its publications, its website, or its tools.

(Wenger, 2011, <http://wenger-trayner.com/resources/slide-forms-of-participation/>)

Individuals may have differing *trajectories* with respect to their level of participation. A trajectory refers to the motion toward or within a membership category (Wenger, 1998, p. 154). For example, a participant with an inbound trajectory may be a newcomer who is currently at the periphery of the community, but wishes to become a core member. Their motivation and investment in learning the norms of the community and participating the community's activities may be higher than those who are on outbound or other types of trajectories. However, an insider trajectory is still possible for those who are already core members — as the community changes, these core participants must continually reassess and renegotiate their identity relative to the evolving practice (Wenger, 1998, p. 154).

As these modes of belonging and levels of participation directly relate to participants' identities within the community and how they learn and make meaning, these ideas will be revisited and addressed more fully in later chapters of this thesis. The ideas presented here will also be related to other concepts, such as those pertaining to influence, social capital, community involvement, and linguistic behavior and awareness.

2.2.2 Community of Practice versus The Speech Community

Since the introduction of the CoP model into linguistics by Penelope Eckert and Sally McConnell-Ginet in 1992, other sociolinguists have been applying the model to their ethnographies (Bucholtz, 1996, 1999; Eckert, 2000; Davies, 2005; Moore, 2006; Lawson, 2011; Holmes and Woodhams, 2013) to demonstrate how identities are not simply linked to linguistic features, but how those features are used as resources to construct identities within specific contexts (Moore, 2006).

The CoP approach has been contrasted with the speech community model, the latter of which was primarily used in sociolinguistic work until recent years. Several linguists have proposed slight variations on the definition of 'the speech community' (Chomsky, 1965; Gumperz, 1968; Labov, 1972). What these definitions have in common is a focus on a shared set of norms and linguistic forms that delineate one community from another. Chomsky's (and to some extent, Labov's) definitions also assume a level of homogeneity within the language of the group. This does not always account for variation or differences by social factors such as class, and therefore the model can quickly fall apart when applied to dynamic, heterogeneous groups of speakers that can comprise any single community.

Sociocultural linguist Mary Bucholtz (1999, p. 207) has identified six disadvantages of the speech community model, with the most relevant to this study being that the speech community views "*identity as a set of static categories*" and valorizes "*researchers' interpretations over participants' own understandings of their practices*". The CoP model is not subject to these pitfalls because it assumes that an identity (both individual and group

identity) is fluid and changing, and that those changes are defined by the participants in the community, not the researcher observing them (Bucholtz, 1999; Holmes and Meyerhoff, 1999). The CoP model therefore provides a way to link micro- and macro-levels of analyses (Holmes and Meyerhoff, 1999, p. 181). Additionally, Bucholtz (1999) points out how the boundaries of a practice in the speech community model are determined and defined by linguists. This is unlike a CoP, which necessitates that the community itself defines the boundaries. The CoP model takes the perspective of those within it; boundaries and categories are not imposed upon it. This is very much in the spirit of variationist methods and part of why the model has lent itself so well to sociolinguistic research. As such, the speech community model will not be elaborated on further in this thesis, although it may still have utility for other research endeavors.

2.2.3 Social Networks and Weak Ties

The concept of social networks was borrowed from sociology and brought to light in sociolinguistics through studies that explored the ties between participants in relation to the spread of linguistic innovations and social factors like class, gender and age (Milroy, 1980, 2002; Milroy and Milroy, 1985, 1992). Social networks can be defined in two dimensions: *density* and *multiplexity*. Network density refers to the amount of ties that exist connecting participants within the network. In dense networks, most or all of the participants are connected to each other; in loose networks most participants do not know each other. Multiplexity refers to the strength of the ties — participants who know each other in more than one role (e.g., participants who are both friends and coworkers). A uniplex tie refers to a single connection between participants.

Weak ties (high uniplexity) within a network are often the point where one can observe the spread of innovation, as the participants with weak ties often lie at the peripheries of networks (e.g., occasional, peripheral, or transactional participants) and can therefore act as brokers for information between groups (Wenger, 1998). The value of weak ties can often be observed in online settings. For example, Twitter users exploit this phenomenon often when participants send tweets to large groups of people that they could not otherwise reach to request or share information (Thompson, 2008). This is just one of many new ways that social media sources provide a particularly efficient infrastructure to exchange ideas and information between acquaintances and others outside of our immediate social circles (Gladwell, 2010). The level of motivation required to participate in these exchanges is low and the means to do so has been made possible by online communities specifically created for those purposes (e.g., Wikipedia or Twitter).

Additionally, these social spaces are typically not governed by a single authority figure

and so their social organization is non-hierarchical by design. Even though the ties are generally loose, negotiation occurs between individuals and decisions are made by group consensus with small commitment required of any one individual. This works surprisingly well for both low-risk decisions such as the emergence of social conventions, as well as bigger propositions such as site changes and redesigns, which effect community operations as a whole (Gladwell, 2010).

The effectiveness of a non-hierarchical, low-commitment, loose social network is also evident on the question-and-answer subsite of MetaFilter, called AskMetaFilter. On AskMetaFilter, participants seek solutions to all sorts of problems and reach a much wider audience than would be possible if they polled their friends or family, who may be too close to the issue or not have the experience necessary to help. Participants within the network who might not otherwise have awareness of each other are brought together through their connection to the issue needing to be resolved. It is yet another way that focusing on what it is that the community does becomes salient and central.

From this perspective, the social network and CoP models work in tandem to highlight communicative behavior. However, the social network model is limited in that it can only go so far as to highlight what the existing, overt connections are and stops short of accounting for ways that non-easily-quantifiable, emerging connections can be formed. For example, on the AskMetaFilter subsite, a participant who is interested in resolving a problem concerning a particular topic might form a social bond with another participant who is an expert on that topic. The network tie that could result from the problem-solving event would only be observable to linguists if the participants made each other contacts. Without these readily observable signs of connections (contact ties), tracking the spread of innovation requires the aggregation of qualitative analyses of interaction. The social network model is ill equipped to handle this task, as it is concerned with the quantifiable connections via explicitly defined roles between participants (Holmes and Meyerhoff, 1999; Davies, 2005). Put another way, social networks can more readily account for modes of belonging that are directly visible, such as engagement that results in observable interaction, and are less able to account for modes that may be less explicit, such as imagination and alignment (see **2.2.1.4 Types of Participation**, p. 13 for more on 'modes of belonging').

To summarize, the social network model involves a quantitative, structural view of a practice, whereas the CoP includes both qualitative and quantitative measures and shows how the structure influences learning (and vice versa). While the main focus will remain on the CoP model throughout this research, ideas from the social network model will be applied as they are relevant.

2.3 Capital

The concept of *capital* describes the ways in which members of society position themselves in a social hierarchy with respect to their access to and use of the resources available to them within their community (Bourdieu, 1986; Lin, 1999).

Three types of capital have been traditionally recognized in societies: *economic*, *cultural* and *social* (Bourdieu, 1986). A fourth type of capital — *virtual* capital — has been introduced in recent years, originating from a PhD thesis about capital on MetaFilter (Lawton, 2005).

Forms of capital vary by individual throughout a society or community, resulting in social stratification. Even in communities which pride themselves on having equal status for all participants, a social hierarchy will inevitably emerge. Participants may bring different levels of expertise to their interactions or may invest varying levels of time and energy into their involvement, resulting in recognition or influence in differing ways and amounts.

2.3.1 Economic Capital

Economic capital refers to the authority over and access to economic resources (e.g., money, means of production, property rights). In communities not based around commerce, such as MetaFilter, economic capital does not have much of a role (if any) and therefore will not be addressed further.

2.3.2 Cultural Capital

Cultural capital refers to privilege and advantages one has or can acquire to position themselves favorably within a society. Cultural capital can be observed on MetaFilter through the “*quality of writing, participation in threads and the ability to discuss different topics*” on the site (Lawton, 2005, p. 107).

Cultural capital exists in three states:

1. **the embodied state:** what each individual knows and is capable of, which can be increased through means of self-improvement. This state of cultural capital in the Metafilter community “*manifests through the prior accumulation of culture offline, and this accumulated culture becomes apparent to others when commentating in a wide variety of topics*” (Lawton, 2005, p. 110).
2. **the objectified state:** materials, goods that can be owned, worn, appropriated, etc. In this state, “*cultural capital exists on Metafilter in the form of a user’s body of postings and comments*” (Lawton, 2005, p. 110). Commodified or reified aspects of the community, such as MetaFilter t-shirts or elements of the community register, can be considered capital in the objectified state. Lawton (2005, p. 111) also points out how participants have equal access to the contributions that become reified as part of the shared repertoire of the community, but only to the extent that each

participant invests their time and energy in the community. For example, peripheral participants who do not follow discussions or read areas of the site where these norms are negotiated do not have the same access as more integrated (core/active) participants, who are exposed to this content and therefore acquire the cultural capital in the objectified state.

- 3. the institutionalized state:** certifications, qualifications, etc. that are recognized as meaningful, authoritative, or powerful. Lawton (2005, p. 111) argues that there is no direct correlate of this state on MetaFilter, but that participants may transfer their off-site institutionalized cultural capital into their comments and posts via the academic or other qualifications they've acquired, which inform those contributions.

2.3.3 Social Capital

Social capital concerns the benefits of investment in relationships and social networks (Lin, 1999). Social capital has been further defined as comprising two functionally different types, *bonding* social capital and *bridging* social capital (Gittell and Vidal, 1998; Putnam, 2000). Bonding social capital is centered around the trust that is formed and the resources that are exchanged between individuals with similar social identities, bringing those who know each other closer together (Gittell and Vidal, 1998). Bridging social capital connects people with dissimilar social identities or who do not already know each other. Bridging social capital is centered around weak ties, as they allow resources, information and new perspectives to be exchanged among differing social groups (Putnam, 2000).

Lawton (2005, p. 112) points out three avenues in which social capital on MetaFilter can be achieved: 1. through asynchronous conversations on the site, 2. through FtF meetups, 3. through other forms of media, e.g., email, MetaFilter-related spin-off sites, IRC channels. Crucial to these findings, the engagement that takes place must be recognized by others for the acquisition of social capital to occur. That is, the engagement must be mutual; it is not sufficient to invest in one-sided social interaction.

2.3.4 Virtual Capital

Emerging from Lawton's (2005) thesis research on MetaFilter, a fourth form of capital has been defined. *Virtual capital* concerns one's competence in online environments and can be measured by one's contributions to a community which demonstrate internet fluency and technical ability. Virtual capital, as it relates to MetaFilter, is discussed in greater detail in **3.3.2 Previous Study of MetaFilter and Capital, p. 65.**

2.3.5 Summary of Capital

While four types of capital have been described here, the examples and analysis in this thesis will primarily focus on social capital. This type is the most productive in the analysis of

MetaFilter and the M-Set, as social capital adds to a greater understanding of the relationship between types of community membership and the linguistic behavior that is correlated with these social positions.

2.4 Computer-Mediated Communication (CMC)

Computer-mediated communication (CMC) refers to social interaction that occurs through a computer connection or computer-based device (Baron, 2003). Aside from forms of CMC such as internet relay chat (IRC) and communication over the internet on forums and blogs, CMC also includes email, SMS (text) messaging, Skype and other forms of voice-over-IP, to name a few. CMC environments are places where humans interact for the purpose of social cohesion (Hutchby, 2001; Wilson and Peterson, 2002); this interaction is mediated by — not defined by — internet connectivity (LeBlanc, 2010).

CMC studies have had a major positive impact on sociolinguistic theory in the last decade. CMC-centered sociolinguistic investigations have included the definition and classification of genres and sub-genres of CMC (Herring et al., 2004; Herring et al., 2005; Herring, 2007), the perception of gender in weblogs (Miller & Arnold, 2001; Herring & Paolillo, 2006), youth and identity online (Androutsopoulos and Georgakopoulou, 2003; Herring, 2008b), methodological approaches to conversation analysis in CMC (Stommel, 2008; Herring, 2010), and ethnographic practices that are inclusive of internet environments (Androutsopoulos, 2008), to list a few.

2.4.1 Classification of CMC Research Areas

Research in CMC can be grouped into five major areas, presented below roughly in order of their emergence (Herring, 2008a):

1. **Classification.** In earlier years, classification was concerned with comparing genres of CMC to modalities of speech and writing as well as to a third, hybrid modality, coined 'netspeak' (Crystal, 2006).
2. **Structural features of internet language,** including orthography and *netologisms* (internet-derived neologisms; Witten, 2012).
3. **Discourse patterns,** including speech acts, conversation analysis and language varieties.
4. **General human behavioral studies** such as learning, maintaining social ties, argumentation and shopping online.
5. **Languages and language ecologies.** This area examines the effects of globalization, the use of English as a lingua franca, and other phenomenon associated with connectivity and language contact.

Some of the early classification work influenced researchers to examine how the

values that define the concept of ‘community’ have different weighting across domain and genre. Herring (2004) compared two CMC listserv environments, Linguist List and Inquiry Learning Forum (ILF) to test their validity as “virtual communities”. She found Linguist List to be more community-like, despite the fact that ILF initially seemed “richer” by providing a multimodal social experience (and Linguist List being text-only). One might have expected ILF to be more akin to a traditional community, since multimodality mirrors FtF interaction more closely in this sense. However, this was a superficial similarity, as it was only the technology that was mirrored, and not the social structure that actually facilitates communities.

Herring (2004) explains this by attributing the unexpected results of her study to Linguist List users having more regular offline contact, as the participants are peers (the ILF group has a hierarchical structure; list owners hold higher academic positions than list participants). The non-hierarchical structure of Linguist List led to a greater sense of shared ownership in the practice and the freedom to act autonomously (Herring, 2004). Again, it is the practice and the social organization that mattered first and foremost, not the technological features that mediated that practice.

2.4.2 Communicative Benefits of CMC

The technological features that mediate CMC environments can facilitate community cohesion in ways that may not be available in most FtF and written genres (Herring, 1993, 1999, 2007; LeBlanc, 2010). These features include (but are not limited to) increased intertextuality, a more egalitarian landscape (where people have greater control over one’s identity, anonymity, and privacy while communicating), and the ability to plan and carefully construct one’s speech in asynchronous communication. Each of these benefits of CMC are explained more fully in the following sections.

2.4.2.1 CMC Communities: Fostering Intertextuality

While books and other forms of written narrative are intertextual, LeBlanc (2010) argues that the types of communications found through her internet ethnography (and applicable to other online communities) directly and indirectly encourage and foster greater intertextuality. This is achieved through the funneling of conversational properties otherwise expressed through gesture, prosody and paralinguistic cues into the language itself (Rheingold, 1993; Warnick 2010; Greiffenstern, 2010; LeBlanc, 2010). In this way, the linguistic signs are imbued with layers of social meaning that are conveyed through innovations such as emoticons, quoting abilities (including the pragmatics in the presentation of text contributions), a specialized register, in-jokes, online tools and features that provide meta data for the communication.

2.4.2.2 *CMC Communities: A More Egalitarian Landscape*

In an early paper on gender and democracy in CMC, Herring (1993) explains four commonly cited reasons for a more egalitarian landscape in online environments. The first democratizing aspect is accessibility of individuals to get online and participate. This gives all people access to others, as well as to information and social networks, at little cost.

Her second point addresses the social decontextualization of CMC environments. Identities can be obscured, anonymized or completely fabricated. This is especially common with ambiguous usernames, where it may be difficult to discern any personal characteristics of the participant. Herring (1993) points out some pros and cons of this ‘flattening’ of social hierarchy, stating that social status cues such as accents and appearances are neutralized, but that may make communication seem less personal or socially informative.

Her third observation is that CMC does not have set usage conventions, being a relatively new environment. This also has its pros and cons, seeing that it can lead to the organization and development of social norms, but also to outrageous, anarchist behavior. This point is somewhat outdated for internet usage as a whole, as the internet has evolved quite significantly in the twenty years since this article was published, but still applies to emerging online communities as they establish their norms and conventions at a more local level.

Herring’s final argument for the democratic nature of CMC is the rarity of censorship. While some communities today are heavily regulated and contributions sometimes edited after posting (especially with respect to swearing or inflammatory language), in a more general sense, the internet is a place where one is free to say or post whatever one wishes. This is often allowed or encouraged within communities and helps foster a sense of openness and freedom.

2.4.2.3 *CMC Communities: Planning Speech in Asynchronous Communication*

One of the most notable differences between text-based CMC and FtF interaction is CMC’s heavier reliance on language to carry both the semantic content and pragmatic load in the absence of visible paralinguistic signals and prosodic cues that the speech modality provides (Rheingold, 1993; Warnick 2010; Greiffenstern, 2010; LeBlanc, 2010). Asynchronous communication (conversation occurring in protracted turn-taking; not unfolding in real time) fosters metalinguistic awareness by allowing participants to carefully plan their contributions, reflect upon and reference previous turns of speech, and use the timing and structure of their responses as metapragmatic cues. This type of communication also lets participants take part in multiple conversations at once (Herring, 1999, 2007).

2.4.3 *MetaFilter as a Social Network Site, or ...?*

According to boyd and Ellison (2008) a social networking site (SNS) is defined as a web-

based service comprising all of the following three criteria:

1. Users can create profiles, public or semipublic, within the bounds of the system.
2. Users can display their contacts/connections with other members of the system.
3. Users can find each other through the system.

By this definition, the MetaFilter CoP could be considered a SNS. However, this is problematic for two reasons. First, how participants within a community perceive and classify their community relative to other communities matters. For example, MetaFilter members generally do not see themselves as belonging to ‘a social network’ in the colloquial sense that the general public understands the term, akin to sites like Facebook or LiveJournal. In fact, many participants on MetaFilter are frequently vocal about their disdain for Facebook and other typical SNSes. These participants may pride themselves on their membership to a community that they perceive as qualitatively different from mainstream places that are more focused on social aspects.

This notion of distinction comes from positions of authority as well. In a presentation given at GEL Conference (2010), MetaFilter founder Matt Haughey paraphrased a fellow moderator and characterized MetaFilter as “*kind of like a social network for not-friends.*” This further suggests that MetaFilter is somehow off the worn path of typical SNSes, which are designed to enhance social bonds between people who acknowledge each other as friends first and foremost. This leads to the second problem with the classification: Maintaining social connections is secondary to the main purpose of MetaFilter, which is to share and discuss interesting links from the web.

This classification issue may be somewhat resolved with the addition of another typology — one in which there is a distinction between sites like MetaFilter and Facebook, not based on the structure of the site allowing users to connect, but rather on the social basis for those connections. This inherently includes a CoP approach, in which the community or site is defined by what it is the participants do. In this typology by (Armstrong and Hagel, 2009), four common types of online communities were outlined:

1. Communities of interest (e.g., MetaFilter, Language Log, topic forums)
2. Communities of relationship (e.g., Facebook, which relies on a social network that pre-exists offline)
3. Communities of fantasy (e.g., MUDs and MMORPGs such as WoW)
4. Communities of transaction (e.g., Craigslist, eBay)

Under this classification system, MetaFilter is primarily a community of interest, with participants rallying around not any one fixed topic, but collectively around the idea of exploring topics in general, whatever they may be.

The typologies covered so far are not mutually exclusive. In the case of MetaFilter, the site functions as a CoP, a community of interest, arguably a social network site, and a community blog (self-defined as such).

2.4.4 Describing the Features of an Online Community

A model of Computer-Mediated Discourse Analysis (CMDA) has been developed to tackle the task of succinctly outlining the features of a particular CMC environment (Herring, 2004, 2007). This model accounts for and is able to describe both the macro-levels of the genre or practice (the type of community it is and the collective identity and behaviors that it contains) and the micro-levels of CMDA (linguistic phenomena, features of the genre) (Stommel, 2008).

The CMDA model was design-inspired by Dell Hymes' (1974) 'SPEAKING' grid and contains multiple "facets", a concept borrowed from classification theory in the field of library and information science. The facets included in the description of a genre are based on the features of that genre that shape it most crucially. Because of this, the layers of the classification scheme are self-selected by the sociolinguistic researcher from a range of possible choices. Facets can be added or expanded upon for a more thorough description of the online community under investigation.

It is assumed that CMDA classification scheme is influenced in two major ways: through its medium (the technology) and its situation (socially). These are unordered, non-hierarchical relationships, but all facets are organized by how they fit into these two categories. In sum, the CMDA grid provides a straightforward way to convey important features of an online community, self-selected by the researcher and tailored for the purpose of the study. A CMDA sketch of MetaFilter has been provided in 3.2.2 A CMDA Outline of MetaFilter, p. 55.

2.5 Registers and Enregisterment

This section introduces register, enregisterment, and indexicality as crucial concepts toward understanding how forms come to have social meaning and how people in communities use them as resources to construct identities.

2.5.1 Registers – Definition

First coined by T. B. W. Reid (1956), the term *register* has been further defined by sociolinguist Asif Agha (2004) as "*a linguistic repertoire that is associated, culture-internally, with particular social practices and with persons who engage in such practices.*" Agha has expanded on this in recent years with studies that explore the concept of register and the process of *enregisterment* — how forms become part of a register and are indexical of the

people and characteristics associated with them (Agha, 2003, 2004, 2005).

While registers are continually changing, expanding and narrowing, there are three aspects that are common to all of them (Agha, 2004, 2005):

1. Registers have an observable size (number of forms), grammatical range (classes of forms in which the forms can occur), and semiotic range (signs that co-occur).
2. Registers have a range of pragmatic values, meaning that stereotypes exist for users of the register, the domains in which the registers can be used, and the set of associations (positive or negative) that are associated with the register.
3. Registers circumscribe a social domain, including categories of people that can recognize forms of the register and are competent in that register's use.

A register cannot exist in isolation — it must be validated by the recognition and activities of others. As such, a register is the reified result of the negotiated meaning of the forms used by participants within the domain. Because speakers' levels of participation within the domain or community vary, along with their modes of belonging, a register's forms and values are not uniformly shared by speakers, including the ascriptions speakers have about them, both within a register's social domain and outside of it. All aspects of the register are part of a continual social negotiation of meaning and competence. As the social structures and people who support the register change over time, so does the size, range, values and domain of the register (Agha, 2004).

2.5.1.1 *Register Competence and Register Socialization*

How one acquires a register — that is, *register socialization* — is an ongoing process. A participant engaging in any social activity may need to be aware of changes in the register, as they reflect changes in the social hierarchy and power structures of that domain. The content of a register (features, forms, etc.) are nothing without the understanding of the pragmatic norms of their use, which are also part of the register and not accessible without engagement in the practice. From this, there are inevitably variations in register competence between participants of any group; some participants may be able to recognize and understand forms, but are not skilled in using them in socially appropriate ways. This can create divisions or barriers to more entrenched aspects within a group's practice (Agha, 2004).

Elements of registers are constantly changing, as participants with influence and capital continually renegotiate the meaning of the forms. Participants in a CoP must keep up with these changes, and sometimes contribute or motivate the changes to stay current in the eyes of their peers. This constant repositioning keeps the register alive and builds historicity and cultural knowledge that participants can reflect upon, further demonstrating their authority in knowing the progressions that have occurred. For CMC social endeavours, these evolutions

can be rapid and complex.

There is a cognitive limit to the number of registers a person can know and competently perform. The ability for any one person to correctly identify registers is much greater than this. Therefore, for any register, the competence to recognize the register (through its forms) will always be much greater than the number of people who can speak in that register (Agha, 2004, p. 35). For prestigious registers — even covertly prestigious ones — this disparity creates a power differential between those on the inside and those who might want access to the register or the competence to learn it. Therefore, the ability to recognize and use registers competently can also result in increases in social or virtual capital.

2.5.2 Enregisterment – Definition

“We cannot understand macro-level changes in registers without attending to micro-level processes of register use in interaction.” — Agha, 2005, p. 47

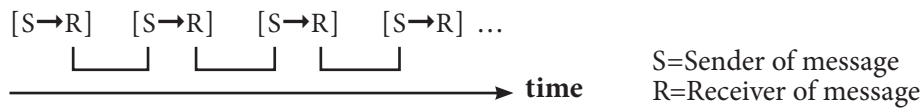
Developed by Asif Agha (2003, 2005), *enregisterment* is formally defined as the process by which a linguistic repertoire (or a linguistic form) becomes differentiable within a language or group as a socially recognized register (or as indexical of a speaker or speaker attributes). By this definition, enregisterment can apply as a macro-level process of dialect formation, or as a micro-level process of feature or form entrenchment.

2.5.2.1 Message Chains

For enregisterment to occur, there must be some observable mechanism by which forms and their social meanings are transmitted from speaker(s) to speaker(s). Agha (2003, p. 247) proposed the ‘speech chain’, *“a historical series of speech events linked together by the permutation of individuals across speech-act roles in the following way: the receiver of the message in the (n)th speech event is the sender of the message in the (n+1)th speech event.”* His illustration of this process is shown in **Figure 2**.

Owing to the flexibility of the medium in which chains can occur, it is more accurate to refer to ‘speech chains’¹ as ‘message chains’. An inclusive naming convention such as this is particularly apt for CMC research, where the items that are communicated are often propagated through many modalities, with only a few of them considered “speech” in the traditional sense of the term. This change to the chain model, as well as the expansion and further detail of the model, are covered in **6.3 Message Chains, p. 155**, as one of the contributions to sociolinguistic theory that this research provides.

1. The ‘speech chain network’ is also henceforth referred to as the ‘message chain network’.

Figure 2. The Message Chain Process

Source: Agha, 2003, p. 247

Any two links in the chain can be separated by time and space. The roles of sender or receiver are not necessarily relegated to single persons communicating FtF, but are conceptual entities that can loosely stand-in for any form(s) of message sender and audience. For example, in mass media, the 'sender' could be a company, a character, a news reporter, a map or even an animated toaster, to name a few. The audience could comprise millions of people, and in turn, it may be the case that just as many of them act as senders of the message later on in the process.

Participants in these chains do not need to be affiliated with each other, but they do need to have some shared discursive experience that links them together through an item of cultural value transmitted along the chain. Their participation is something they share even if it is not at the same time or in the same role. They need not even be aware of each other's shared experience — what is required is that co-members of the chain network have awareness of the values that are being conferred through the chain (Agha, 2003, p. 248).

The next step in this process is to explore how forms transmitted across message chains come to have social value and be recognized by message chain participants as such. Indexicality can account for this linkage between form and meaning.

2.5.3 Indexicality

“Indexicality is necessary for showing us how to relate the micro-social to the macro-social frames of analysis of any sociolinguistic phenomenon” — Silverstein, 2003, p. 193

For social meaning to be conveyed through forms in a message chain, a form must act as a sign, carrying not only the message, but also the message within some frame of reference related to social meaning. *Indexicality* can explain this relationship between signs and their referential meanings.

A sign that points to a property relating back to the sign is called an index. There are many types of indexes, including *referential*, *natural*, *socially-constructed* and *(socio)linguistic*, to name a few. A referential index could be the use of a name, indicating the person who goes by that name. A natural example of indexicality is the association between two related things or processes, e.g. a bullet hole in a wall indexes both a bullet and a gun fired. Socially-constructed examples of indexicality link forms such as objects, processes or behaviors to social meanings, e.g., a white dress being symbolic of brides or weddings. Linguistic examples

of indexicality involve word forms or features linked to social meanings, such as the use of 'like' as a quotative associated with youth culture. These types of indexical relationships do not hold for all speakers at all times; they change and are context-specific.

The type of indexicality that is of relevance to this research is non-referential and socially-constructed (sociolinguistic). These are linguistic forms linked to particular social meanings, such as when regional slang is associated with certain areas, cities, neighborhoods or even streets.

William Labov's (1972) taxonomy of the indexicality of socially-constructed linguistic signs includes three hierarchical categories: *indicators*, *markers* and *stereotypes*.

Indicators are sign-meaning relationships of variable features that are below the level of consciousness for speakers. Indicators can be identified by trained observers, but the associations they find are not obvious to members of the group to which the indicators belong. There is also no social meaning yet attached to the forms. Therefore, speakers do not use the indexical link to contextualize or interpret others' speech or construct identities (Johnstone, Andrus, and Danielson, 2006). However, at some point this indexical link must be made salient (i.e., noticeable) to speakers if an indicator is to move on and become a marker of speech.

Markers are context-sensitive variable features that are socially meaningful to speakers. They index social factors or characteristics of a speaker or community, independent of a community member's overt awareness of the variable or the social meaning attached to it. While speakers can recognize the sign-meaning relationship, those associations are not yet highly salient or stereotypical.

Stereotypes are variable features that are overtly recognized by speakers as linked to social meanings. Their use may increase or decrease or their meaning may change as a result of this awareness.

A slightly different view of indexicality was proposed by Silverstein (2003). He states that these links between form and social meaning can be categorized in one of three different, hierarchical *orders of indexicality*, depending on their level of abstractness in the minds of speakers.

The first of these orders is referred to as an *n-th order indexical* and defines a situation where the form can be deemed as appropriate for a context, or is otherwise linked to a person or situation. Like Labov's indicators, the sign-meaning relationships of n-th order indexicals are generally only noticeable by a trained observer; the association does not have meaning within a social space.

At the next level, *n+1-th-order indexical*, forms acquire "*an ethno-metapragmatically driven native interpretation*" (Silverstein, 2003, p. 212), meaning that these links come to have

social meaning attached to them. What was originally a neutral association between a form and a speaker or speaking context now has evaluation added. This makes a form pragmatically usable to convey the sign-meaning relationship that it contains, whether or not speakers are consciously aware of using the forms in those ways (Johnstone, Andrus, and Danielson, 2006).

At the final level, *(n+1)+1-th-order indexical*, the link between form and meaning has evolved to the point of acquiring additional evaluations, such as indirect associations to more nuanced schemas based on features of the original associations (features of the form, the identities it is linked to, or the context it is linked to).

Where Labov's model of indexicality focuses on the speaker's awareness of the sign-meaning relationship, Silverstein's narrows in on the direct versus indirectness of the social meaning that is associated with the form. Similar to the difference between CoP and Social Networks, Labov's indicators, markers and stereotypes are centered around how the indexicals are used and are perceived, whereas Silverstein's orders of indexicality are centered around the structure and the nature of the sign-meaning relationship.

2.5.3.1 *Form-Meaning Relationships*

The indexical relationships between variants and social categories are learned patterns, resulting from speakers' experiences with language and their social environments (Foulkes and Docherty, 2006). Therefore, speakers from differing cultural and experiential backgrounds will have dissimilar mental maps of which forms index which social meanings and categories.

Socially-constructed (learned) form-meaning relationships do not account for all types of variation. In many cases, internal factors such as physical characteristics have influence on the form-meaning relationships and how they evolve or change. For example, internal factors such as physiological differences in speech acoustics owing to a speaker's age, sex or health can sometimes be entirely independent from social constructions.

As indexical relationships between variant forms and internal factors are more or less experienced by everyone, the form-meaning associations that result are more consistent and universal than learned culturally- and context-bound patterns. Variation that can be entirely explained by factors such as universal acoustic principles (i.e., is not at least partly socially-constructed) is not to be considered within the domain of sociophonetics (Foulkes and Docherty, 2006), but must be considered in sociophonetic work, as it clearly has influence and cannot be ignored.

Therefore, in the examination of any variable's distribution of variants among a population, it is crucial to determine which types of factors (socially-constructed or internal) are at play for which variants and to what degree they influence the choice variant by a

speaker (or population of speakers).

The social meanings of phonetic variables are not fixed. They may be renegotiated through processes of participation in a CoP or otherwise change over time and across locales and social contexts (Foulkes and Docherty, 2006; Munson, 2010). They can refer to mutable characteristics (such as the stance or attitude of an individual) or fixed categories (such as where one was raised or their native language). The associations that are evoked with the use of any phonetic variable can index different things to different people. At any given point in time, this range of associations forms a constellation of indexical values, called an *indexical field* (Eckert, 2008). For variables such as the M-Set, the constellation can be vast and mercurial, with differing patterns for each of the variants.

Additionally, these variables and their associative values do not exist in isolation from other language features or non-linguistic signs (Eckert, 1996, 2008). They co-occur with features of language, appearance, behavior and style. For example, correlations have been demonstrated in sociolinguistic studies of Chicana gangs, where realizations of /i/ co-occurred with social position in the gang membership (e.g., core versus periphery), and even the length of eyeliner worn (Mendoza-Denton, 2008). Eckert's (1988) sociolinguistic research of high school girls in Detroit showed similar patterns, with vowel variants correlating with preferred jeans style. These studies show that details — even subtle ones — contribute to an overall projection of a persona. Their success is dependent on the stylistic construction and expression of features and how those features are received (i.e., perceived) by others. What this means for the M-Set is that, as a variable, it cannot be singularly representative of the linguistic style of MetaFilter, nor can it be understood without its context and co-occurring features and forms. As it pertains to this study, the M-Set is a starting point, and can lead to further, related investigations of variables and their indexicalities, which can then be compared or built upon.

It is important to also pay attention to speakers' trajectories along with their stylistic use of sociophonetic variables. This focus falls in the realm of audience design (Bell, 2001), where it is shown that people construct their utterances with awareness of the audience they are addressing and in the style in which they perceive their messages will be received best, according to the social goals they have in delivering those messages. If a positive interaction is desired, speakers will tend to accommodate their speech style to that of their interlocutors, in an attempt to evoke a sense of familiarity, camaraderie and group membership (Giles, Taylor, and Bourhis, 1973). To do this, speakers must recognize the available and salient resources necessary to accommodate, such as phonetic features their audience might be familiar with and likely to use, and demonstrate their competence in adopting them in appropriate ways. These

style performances allow speakers to position themselves as part of a group, project a desired identity and take effective stances in interaction (Johnstone, 2007; Hay and Drager, 2007).

Novel phonetic features, new forms, or new ways of using old features or forms may be particularly crucial in the performance of style. They are an opportunity for a speaker to demonstrate greater social competence through their creativity and innovation. A speaker who uses a new phonetic feature or form may (unwittingly) become a linguistic leader in the use of that form. This can be the start of a new message chain, or a major link in a pre-existing one. Regardless, any of these new forms (or new ways of using old forms) start out as a minority in their use (Foulkes and Docherty, 2006), and can eventually diffuse through a community, usually in a S-shaped curved pattern of use over time which is typical of the spread of innovation.

2.5.4 Sociolinguistic Approaches to Variation and Enregisterment

From the perspective of the sociolinguistic history of a given population, the form-meaning relationships that arise are motivated by the experiences of those speakers within their linguistic contexts. These form-meaning pairings have historical and cultural explanations. Therefore, any investigation into the social meaning of phonetic variables must include some attention paid to the specific history that allowed the form-meaning pairing to evolve as such. Exploring the community that finds this form-meaning relationship of relative social significance is the entry point into understanding the sociophonetic variation that results.

Participatory approaches allow researchers to analyze sociophonetic variation in depth, with social categories that are local and relevant to the community under investigation (Hay and Drager, 2007). This is especially important for phonetic variables, which can be highly context dependent and localized. The important or salient values or categories in an indexical field for any phonetic variable will vary from community to community, and so a participatory approach is necessary to discern exactly which values and categories are socially meaningful for any particular community of speakers.

Many studies of enregisterment have taken the approaches as outlined previously, which has resulted in a rich analysis for the motivations behind language change, situated in specific sociohistorical contexts. Selected studies using these methods will be presented here.

2.5.5 Enregisterment of a Language Variety

Several papers give examples of variety enregisterment occurring in geographically-bounded speech communities, where FtF communication is the dominant mode of interaction (Agha, 2003, 2005; Johnstone, Andrus and Danielson, 2006; Johnstone 2009; Beal, 2009). While this thesis will focus on the process of enregisterment occurring in the text-

based interaction of a thriving internet community, reviewing this earlier work is crucial to understanding the approach taken here. A comprehensive review of relevant studies on the enregisterment of varieties of speech will be outlined below.

2.5.5.1 *Enregisterment of RP – Associations with Socioeconomic Class*

The process of enregisterment was first introduced in a paper examining the evolution and spread of a prestige register of British English known as *Received Pronunciation*, or “RP” (Agha, 2003). In the 17th century, no English variety was considered standard, but over the next 200 years, a standard emerged (RP) and became well established and regarded as having high cultural value.

For RP, the enregisterment of the accent went through roughly four stages, demarcated by the dissemination of various types of texts and the audiences that they were available to. This process started with a small circulation of early prescriptivist works catering to aristocracy and allowing only them access to the knowledge to gain competence in the RP register. Later, popular handbooks were available to the upper middle classes as well, thereby expanding the range of people who had access to the register, as well as to those who might be able to recognize it as such. Over time, the middle class had access to novels and literary works which contained examples of RP speech (containing respellings that mimicked the RP accent). Lastly, smaller, cheaper publications such as Penny Weeklies, were distributed to a very large audience, consisting of lower middle and upper working classes.

These publications had become well established by the mid-19th century. By this time, all class levels had awareness of some of the forms of RP enough to recognize it as a prestige accent, but still only a very small percentage of the population had competence to use the register appropriately. However, the ability to recognize the RP accent was its own form of competence, and incidentally gave more value to RP as a social commodity.

In this example of variety enregisterment, a new accent emerged and new ways of speaking associated with evaluations of social class had contrasted with previous awareness of accent in Britain. While the dissemination of texts was not the only mechanism through which RP became enregistered, printed sources tailored to social class categories were a major component in the enregisterment of the variety.

These texts were not only important to note not for their impact on the enregisterment of RP, but in their ability to facilitate the process as a form of messaging that is consumptive and focuses on the imagination and alignment modes of belonging, rather than the engagement mode (see 2.2.1.4 **Types of Participation**, p. 13, earlier this chapter). In addition, this mechanism of message propagation provided an easy way for linguists to track the stages in this

process. However, as the next study demonstrates, other less-straightforward mechanisms can come into play, making the tracking of enregisterment slightly less transparent.

2.5.5.2 *Enregisterment of Pittsburghese – Associations with Place*

While Agha's exploration of RP focused on enregisterment of a prestige variety linked to socioeconomic class, Johnstone, Andrus, and Danielson (2006) showed how linguistic forms within a community were first associated with the speech of a socioeconomic class and later to place, as a socially recognized dialect called "Pittsburghese" (Johnstone, Andrus, and Danielson, 2006; Johnstone, 2009, 2010b). The establishment of the variety has been likened to that of dialect formation, both having the same economic conditions motivating speaker accommodation (Johnstone, Andrus, and Danielson, 2006).

Pittsburgh residents' working class identity has long been tied to the steelmaking industry of the city (Johnstone, 2010b). The dense, multiplex social networks that residents belonged to precluded them from awareness of other ways of speaking. However, the social trajectories of many of these Pittsburgh residents changed during WWII, when many of them left the city for military travel and the union workers who remained in the city for the years after the war were able to vacation to other locales (Johnstone, Andrus, and Danielson, 2006).

Social changes during this time within the city also had an effect on residents' exposure to other ways of speaking. Social mobility increased with the growth of the medical and university economic sectors. Formerly industrial areas became places where students and professionals lived cheaply and later settled. As a result of these changes, residents started hearing their own speech as contrastive with that of newcomers. Forms became noticeable and evaluations of these residents' "nonstandard" or "different" ways of speaking had become increasingly common knowledge, with evaluative stances linked to them. Additionally, the newcomers picked up the local forms they heard as part of their claim to local identity.

This specific social context allowed two groups of people within the community — residents (core participants with insider trajectories) and newcomers (peripheral participants with inbound trajectories) — to notice speech in new ways and use that social knowledge to construct their identities. In this case, the efforts of individuals from each group were aligned (both wanting to achieve the same social position of "local"), but their motivations for doing so arose from different starting points — the newcomers with outsider backgrounds sought new community membership and the residents sought to hold onto their authentic, historical roots. This reinforced enregisterment from two different perspectives, allowing both newcomers and residents to participate in the process alongside each other, for different social goals, but aligned in the claim of an identity and membership tied to place. Through this, both groups had an invested stake and an important say in the enregisterment of Pittsburghese.

2.5.5.3 *Enregisterment of Geordie & Sheffieldish – Mobility and Networks*

In this next study of variety enregisterment, two dialects of Northern England were compared, with one dialect becoming enregistered much more rapidly than the other, despite both of them having relatively comparable histories (Beal, 2009). The dialects of Newcastle and Sheffield each have enregistered features, but only Newcastle’s “Geordie” dialect had gone on to further stages of commodification.

Even though both cities had similar historical orientations to industry and growth, the demographics of each city led to the formation of different types of social networks, which crucially shaped the identity of the city and the people who live there. Additionally, Sheffield was seen as part of the larger identity of Yorkshire, while Geordie was seen as independent and distinct. These differences also facilitated more tourism in Newcastle than in Sheffield, and therefore commodification and the spread of innovation occurred in Newcastle more rapidly (Beal, 2009).

This comparison showed how the perception of a particular place, by both insiders and outsiders, had influence on the rate of enregisterment for that locale’s dialect. Social evaluations about distinctiveness (as part of a larger region versus having an independent identity) and local demographics (influencing the habitus or ethos of a place) led to favorable conditions for mobility and tourism, which facilitated linguistic exposure and transmission of innovation by both insiders and outsiders alike.

Similar to the positive effects of weak ties discussed earlier, the social network structure of any community will have a net effect on language change. With both Sheffield and Newcastle being relatively equal in all other aspects, the differences in the social networks between those two towns (driven by the demographics of the populations) led to a more rapid enregisterment of one dialect over the other.

2.5.5.4 *Enregisterment of Internet Language — Technological Determinism*

In an article investigating the enregisterment of the variety known as “internet language”, Squires (2010) explored various domains of metadiscourse from academic works to print media and online commentary. Two main themes emerged which were crucial to the enregisterment of the internet language variety. First, enregisterment was motivated by a contrast with what is considered Standard English. Second, views of technological determinism helped associate internet language with values of informality and nonstandardness; therefore in direct opposition to Standard English.

This first theme supports the idea that enregistering any language variety requires the previous enregisterment of another variety, so that the emerging variety can be hearable in contrast to a standard and differentiable from it along various possible social dimensions.

Standard English provided this lens through which to look at new internet forms and it has been used to see internet language as subversive and threatening. As a result, internet language provided a good target for the convergence of fears about the state of English, the habits of youth today, and the rapid rise of technology (Thurlow, 2003, 2006).

The other primary influence in the enregisterment of internet language was views about technological determinism — the idea that technological innovation determines social structure. Technological determinism has seen much criticism (Herring, 2008b; Jurgenson, 2012; Banks, 2013); many innovations have both positive and negative aspects, but no outright autonomy to wield over social interaction in such an absolute way. That said, technology does impose some structure in the sense that discourse is framed by context (Squires, 2010).

Standard language ideology also contains a prescriptivist imperative to create and maintain a divide between “nonstandard writing” of CMC and “standard speech” of FtF communication (Squires, 2010). Both of these divisions are inaccurate; as we are increasingly participating in CMC and FtF domains, often concurrently, it becomes difficult to keep up illusions about strictly separate domains for standard and nonstandard speech.

Squires (2010) demonstrates findings from reports of appropriate and inappropriate contexts for internet language and Standard English. Subjects stated that internet language should not be used in “formal” online settings, as well as in any FtF contexts, except in talk with friends. Internet language was acceptable in short emails, but not longer ones. Participants reported that Standard English should be used in all FtF contexts (especially formal ones), except in talking with friends. The only CMC genre where Standard English was not to be used was instant messaging.

These results are interesting for two reasons. First, they imply what responders perceived internet language to be (e.g., nonstandard, almost always informal). That is, these perceptions were based on awareness of Standard English and where it is and should be used, with deviations from this forming an undesirable “other”. Second, that the divisions of usage were decided by social function more than medium. Squires (2010, p. 482) explained this by saying, *“It is not the case that features are distributed along “internet/noninternet” lines and come to be perceived as such, but rather than an “internet/noninternet” line is perceived, and features are categorized as belonging on either side of the line.”*

Additionally, the findings support the notion of reductionism inherent in technological determinism (Herring, 2008b; Jurgenson, 2012; Banks, 2013). For example, how can internet language be unacceptable in some internet contexts, such as emails or professional forums and websites, when the very technology that created and perpetuates it would dictate (in technological determinism) that it be used in all CMC genres?

These results also make further case for CMC being a speech-writing hybrid (Crystal, 2006), as the two are not clearly universally separable in terms of formality or how they are actually used by people. Internet language features are not universal and online speech environments display variation both within and between genres (Squires, 2010).

2.5.5.5 *Enregisterment of Speech Varieties – A Summary*

All of the examples of variety enregisterment reviewed thus far occurred within a specific social context, with participants motivating the process from within the community (British upper class, Pittsburghese residents, Northern Englanders, internet users) or from the outside those communities (some motivated by an inward trajectory to be participants in those communities; others with an imperative to distance themselves from those communities or what they represent). While each linguistic situation has been unique with respect to its situational and historical context, weak ties amongst the networks have been crucial in propagating the linguistic innovations that have occurred throughout all of them. The commonality in social network structure facilitating the establishment of identifiable registers is an important component in a model of the enregisterment process. It is also key in tracking the enregisterment of individual forms within a variety, and will be explored further in the following sections.

2.5.6 *Enregisterment of Forms Within a Variety*

It is important to make a distinction between the enregisterment of a language variety and the enregisterment of linguistic features within that variety. While the process is the same, the contexts differ. Enregistering a variety occurs in contrast to other varieties. Enregistering features occurs within the context of the variety and in contrast to other features (Squires, 2010).

Like a variety, a form or feature becomes noticeable once there is a recognizable contrast (i.e., where there is variation). For example, the linguistic forms that led to the enregisterment of “Pittsburghese” only became noticeable to Pittsburgh locals once they were heard alongside other forms that came from elsewhere. Once that occurred, the Pittsburgh forms became available for negotiation of meaning and indexical associations were subsequently attached to them (Johnstone, Andrus, and Danielson, 2006).

Any form or feature that can be linked with a social identity — individual or group — is available for enregisterment. This includes phonetic features, words, syntactic constructions, pragmatic norms, politeness strategies, and the like (Johnstone, 2010a). Even though speakers are not always explicitly aware of social meanings associated with their local forms, they often respond to them with some implicit understanding of connotations that underlie their use.

Sometimes explicit talk about the social meanings of forms does arise and this metapragmatic activity can quickly further the standardization process along. In the case of Pittsburghese, this metacommentary was achieved through handbooks, websites, and cartoons containing evaluations of speech with the use of local forms. Pittsburgh speech was also negatively valorized through the use of respellings of common words in newspapers (e.g., ‘dahtahn’ for ‘downtown’), thereby reinforcing the links between forms, features and lower social class (Johnstone, Andrus, and Danielson, 2006).

2.5.6.1 *Enregisterment of Forms Within a Variety — ‘netspeak’ and ‘chatspeak’*

Mechanisms similar to those found in the enregisterment of Pittsburghese were at work for propagating negative evaluations of ‘netspeak’ and ‘chatspeak’ — terms used to describe internet language and associated with bad grammar, nonstandardness, and informality (Squires, 2010). This is despite the fact that ‘netspeak’, in use since 1993, was originally enregistered as having associations with technically savvy internet users, marking those who had competence using internet language as specialized insiders to a new subculture (Bacon, 1993, as cited in Squires, 2010). The term lent authority to those who could speak it, even if what “it” is was not clearly defined at the time.

As internet participation increased in the 90’s, ‘netspeak’ included both linguistic forms and features such as abbreviations, emoticons and respellings, in addition to specific lexical terms. The emergence of these new forms and features was attributed to the need for efficiency and speed in communication, owing to the constraints of the medium (Squires, 2010). This association of internet language and acronyms and abbreviations persists, despite the fact that several studies show that they are relatively rare in the particular genre of CMC known for being most representative of ‘netspeak’ — instant messaging (Baron, 2004; Tagliamonte and Denis, 2008, Squires, 2010).

The term ‘chatspeak’ came into existence and, over time, differed from ‘netspeak’ in a crucial way. While netspeak showed shifts in meaning having to do with the language used in specific internet contexts, chatspeak showed shifts in meaning having to do with contexts themselves and the features that are associated with them (Squires, 2010). This also makes sense semantically, as the terms netspeak and chatspeak denote different aspects of CMC with their use ‘net’ and ‘chat’ modifying ‘speak’. Meanwhile, both of these terms were converging with respect to their overall enregisterment as being linked to nonstandard language features, youth and technologically-driven change (Squires, 2010).

2.5.6.2 *Summary of Enregisterment of Forms Within a Variety*

The enregisterment of forms presupposes that variation exists and that speakers can

recognize some contrast between the variants. From this, social meaning can be applied to the forms and allow them to become resources in creating and expressing cultural values.

Even though some forms or features can be stereotypical of the register or community, no one form is responsible for the enregisterment of a variety. It is the collections of enregistered forms that delineate enregistered varieties, much like how bundles of isoglosses demarcate dialects or how groups of participants make up a community, define its boundaries, and co-create its identity.

In all of these cases, individual agents are micro examples of their macro-level processes. Some agents are more stereotypical than others, but they cluster together in some recognizably cohesive way — contrastive with other ways — to form dialects, build communities or enregister varieties. That is to say, the parts that make up the whole are reflective of the larger process.

2.5.7 Summary of Enregisterment

The indexical relationships that are integral to the enregisterment of forms do not necessarily need to be associated with geographically-bound places. Agha (2003) showed this in his examination of RP, where the enregisterment of the dialect concerned markers of social class, rather than being identifiable with a particular region of England. Furthermore, visible social categories are not necessarily required, as Squires (2010) has shown in her investigation of internet language. The form-meaning correlations seen there were outcomes of the shared cultural context and historicity of the participants in the genre of CMC, and not shared geography or social indices such as class, age or gender. This was also true for Beal's (2009) study of the enregisterment of Sheffieldish vs. Geordie, having different rates of enregisterment owing to historical factors and the structure of social networks. Lastly, enregistered values are not permanent, as demonstrated through Pittsburghese, which was once primarily associated with socioeconomic class and then later with place (Johnstone, Andrus, and Danielson, 2006; Johnstone, 2009, 2010b).

Enregisterments are two things at once: a means to reproduce forms and values, as well as a result of that process (Squires, 2010, p. 460). This cannot occur in isolation; it must occur relative to something else. In most cases, it is the contrast between the new form and what is already considered the standard in that domain, genre or locality. Furthermore, the contrast reflects back onto the thing that is being contrasted with. Enregistering new forms cannot occur without the presence of some preexisting form or variety, and that form or variety is ultimately affected by the emerging existence of the new enregisterment.

CMC environments provide an interesting, new way to look at the process of

enregisterment. Squires (2010) noted that the population of speakers in CMC is not clearly definable, either by geography or by local, place-distributed features (social factors or variables); both are not directly observable. Additionally, she pointed out that the ways in which enregisterment was accounted for in studies of regional dialect enregisterment do not apply in text-based mediums, as those mechanisms relied on hearable instances of language use, also directly observable even if below the level of consciousness for hearers. Therefore, other factors must account for how contrast is “heard” in CMC, and how ideas and values about features are disseminated and evaluated.

2.6 Onomastics

All names are indexical of various social factors at multiple levels of linguistic structure. At the denotative level of linguistic structure, a name identifies a specific referent in the world (be it a place, person, brand, or other onomastic category). At the semantic level, the name will have specific associations, based on its origin, history, and the words or morphological components that comprise it. At the phonological level, the name will consist of a specific set of phonetic elements which may be the locus of variation and change. These indexical associations can be inconsistent and involved in various processes of change.

Furthermore, a name can display ambiguity of referents (‘Paris’ can refer to a town in Texas or a city in France), differences in morphological components, or variation in other ways such as orthography (spelling), orthographic features (hyphenation, letter case, etc.), abbreviated elements (sometimes in the form of nicknames, diminutives or other shortenings) and stylistic affectations (e.g., the US shopping outlet ‘Target’, reappropriated with “French” features and the pronunciation [tɑʒeɪ], usually to humorously elevate its status to hearers).

The indexicality of names (and the phonetic variation they may display) will be the focus of the onomastic literature reviewed here. These studies showcase some of the various ways that names form and are formed by associations that people have with them.

2.6.1 Overview of Naming Categories

Place names, personal names and product names are major categories in the taxonomy of onomastic studies. Each has particular relevance to the research presented here, especially concerning the process of how names come to be enregistered. Descriptions of these naming categories are as follows:

- *Place names*, or *toponyms*, are representative of groups of people linked to a particular place, usually defined in geographical terms.
- *Personal names*, are usually representative of particular people (or groups of

people). They can also become the source material for the development of nicknames to refer to those people (or groups) more colloquially.

- *Product or brand names* differentiate companies or product identities within a commercial landscape. More recently, the concept of branding has been co-opted from marketing and applied to personal identities and social enterprises.

It is necessary to understand place, personal and product naming as onomastic categories that are crucial to the process of enregisterment in which the terms of the M-Set take part. Each major category in the taxonomy (place, personal, and product/brand names) will be reviewed separately in this section.

2.6.2 Place Names

“Place names are not only linguistic signs; they also represent social and historical values. They are created as a result of human’s need to name their environment, and they reflect a person’s relationship with a particular place.” — David, 2011, p. 215

Few academic articles have been written about the pronunciation of place names. To date, none have been written about the pronunciation of CMC place names. This is likely owing to the fact that CMC environments are often not viewed as ‘places’ in the traditional, geographic sense of the word. However, there is something to be gained from previous studies of place names, as there may be some important parallels between the relationships people may have with the places they visit online and those they visit in person. Additionally, considering that place name pronunciations can serve to differentiate insiders from outsiders, locals from foreigners, or index points in history, one’s experience, etc., it is worthwhile exploring these indexicalities, as they may also serve equivalent functions in online spaces.

2.6.2.1 Identity and Ideology in Place Names

“Now comes The Nomad in the Boston Transcript and tells us that we native Missourians do not correctly pronounce the name of our state [...] The Nomad has just about as much propriety in telling us how to pronounce Missouri as we Missourians have in telling him how to bake beans.” — quoted in Read, 1933, p. 28

In an early paper in onomastics, Allen Walker Read (1933) gave one of the first accounts of a widely recognized place name dispute. The US state ‘Missouri’ is phonetically ambiguous at every syllable. Read explored the etymologies of the variations, but did not directly delve into sociolinguistic perceptions or indexicalities; observations of that nature would come decades after the establishment of the field of sociolinguistics in the 1960’s. However, many of his examples included evaluations that would be of interest to sociolinguists, such as the spelling variant ‘Mizzoura’ (and its corresponding [z] variant pronunciation) as “*expressing*

uncouthness and suggestive of barbarism” (as cited in Read, 1933, p. 23).

Examples such as ‘Mizzoura’, ‘Missouri’ and other variants were found in plays, newspapers (both national and local) and through direct attempts at standardization. One such attempt pitted the local majority — ‘Mizzoury’ residents, supported by a Missouri school board — against ‘Missouri’ outsiders, consisting of schoolteachers, dictionary advocates and Eastern, textbook writers. These spelling variations were passionately argued, with the explicit understanding that the orthographic choices were reflective of pronunciation.

This debate over pronunciation led to a great controversy, with arguments supported by local identity on one side of the ideological fence line and prescriptivist appeals of outsider scholars on the other. Authoritative voices in newspapers described Missouri residents as “*plain people*” who were “*careless in pronunciation*”. They further supported the ‘Missouri’ spelling (aligned with a [s] pronunciation) as an “*alphabetic dignity*”, avoiding the dreadful “*tail*” of the alphabet — the late addition of the ‘z’ by the Roman schools at the time of Cicero. On the ‘Mizzoury’ local majority, several judges represented the views of many by supporting a “*good old fashioned*” pronunciation and an unwillingness to be dominated by Easterners.

For Missouri locals, knowledge about what was common was knowledge about what was correct in that context. The defense of this from attacks by outsiders who wanted to impose their own ideology about what is correct was more than a nuisance; it was an attack on their very identity, as well as a clash between two different measures of authority. This authority clashes might be a contributing factor to the onomastic state of ‘Missouri’ today; the local pronunciation won out, but the prescriptivist spelling remained.

The debate over the pronunciation and spelling of Missouri was partly waged and won through text-based communication. This is not unlike CMC onomastic debates of today, where words like ‘.gif’, company names like ‘Linux’ or sites such as ‘MeFi’ face similar battles over ideological factions in ongoing prescriptivist wars.

2.6.2.2 *Identity, Ideology, and Indexicality in Place Names*

Identity is carefully and consistently constructed, even at the phonetic level of speech. In recent sociophonetic work, it was shown that the pronunciation of ‘Iraq’ can be a linguistic resource for American politicians to index their sociopolitical persuasions and assert their identity (Hall-Lew, Coppock and Starr, 2010).

Hall-Lew and colleagues (2010) showed the multiple indexicalities of the second vowel of ‘Iraq’, as used by politicians from the US House of Representatives during their political speeches. Their results revealed that, out of two likely pronunciation variants for the second vowel in Iraq, the [æ] pronunciation indexed conservative ideology, whereas [ɑ] indexed

a liberal ideology. The results were significant even when controlling for regional accents, gender, age and ethnicity. In fact, political party affiliation was the only factor out of the long list of factors examined that had a significant correlation with pronunciation.

Their sociophonetic research of Iraq variation was supported by previous work on ‘foreign (a)’ realizations (Shapiro, 1997; Boberg, 1997), where words borrowed into English with an <a> spelling in a CV- syllable can either be expressed with the Americanized variant [æ] or with a foreign variant [ɑ]. The [ɑ] variant is associated with prestige indices such as higher social class, education, correctness/prescriptivism, and respect or sophistication, stereotypical of RP and more locally, the Boston-area “Brahmins” (Boberg, 1997). More importantly, speaker attitudes towards the people and language associated with Iraq may be a factor. Positive associations with the source language and culture of Iraq would lend prestige to using the foreign variant, allowing speakers to align themselves with the values the variant represents and show respect for the culture that is referenced by the name (Weinreich et. al., 1968).

Continued study on the pronunciation of ‘Iraq’ by American speakers outside of politics showed significant positive correlations between the foreign (a) variant and acquisition of a second language, whereas the Americanized variant [æ] was correlated with time spent in military service (Silva, et al., 2011). The researchers explained their results as owing to speakers’ choices being governed by experiential factors, in the absence of the goals to construct a specific political identity (where sociopolitical indexicalities associated with the variants would be more relevant). In other words, the motivations for military personnel versus politicians in choosing the [æ] variant may be entirely different; the military personnel could be signaling their belonging to a particular CoP, whereas the politician is actively constructing a specific political persona.

These outcomes from two studies on the pronunciation of ‘Iraq’ highlight the complexity of indexicality. Pronunciation variants were chosen based on differing and non-contrastive indexical values held by distinct groups of people. The pronunciation of ‘Iraq’ was an outcome based on differing identities, ideologies, and social goals, despite all the participants belonging to the same general speech community (i.e., America). From this, we can see why it is important not to assume that participants come from the same (or even similar) perspectives on such matters, even in cases where they make the same linguistic choices (e.g., choosing the same variant for ‘Iraq’), as they may be doing so for completely different reasons.

2.6.2.3 *Summary of Place Name Studies*

In these studies of place names, identity is constructed in several ways. In Read (1933), the name ‘Missouri’ was at the heart of a clash between local versus prescriptivist ideology.

In Hall-Lew et al. (2010) and Silva et al. (2011), the perceptions about ‘Iraq’ allowed the pronunciation variants to be used as linguistic resources in asserting sociopolitical ideologies or experiential backgrounds.

Shaping an identity was the central goal in the use of the names by participants in all these studies (whether participants were aware of this or not). This could be personal and ideological (defending one’s own place), or be transitory and self-motivated (using the pronunciation of a name as a carrier to index something about oneself). The range of possibility here is worth noting, and has been taken into consideration when analyzing the potential motivations participants may have for selecting variants of the variables under investigation here.

2.6.3 Personal Names

Like place names, a personal name, by virtue of having the intention of representing an individual, will inherently contain indexical associations with foreignness (or nativeness), typicality (or uniqueness), age, gender, and other social or experiential factors (Lipski, 1976; Kasof, 1993). This happens regardless if whether the hearer can single out a particular referent to which the name belongs. These associations have real-world implications affecting the assessment of individuals or groups of individuals, and in common but extreme cases, their upward mobility or access to particular social spheres (Kasof, 1993; Aura and Hess, 2004). Oftentimes these assumptions are made long before the associations can be verified by a FtF conversation between interlocutors.

However, the vast inventory of possible personal (and user) names means that consistent indexical associations between names and social factors cannot be created or maintained. Any name must be examined within a specific context or identifiable boundaries, so that the range of social factors and characteristics associated with it can be located and explained for that particular context. This becomes particularly important in a group naming event, such as a community name, where a number of factors and potential associations must be agreed upon. It also makes the process of tracking indexicality much more challenging, since names are bound by context and contain varying associations for different people.

2.6.3.1 Nicknaming

“In social terms, we need nicknames, just as we need scapegoats and both cooperative and uncooperative nicknamed and nicknamers — they are all essential tropes of social organization.” — Adams, 2009

At a local scale, nicknames are demonstrative of social power struggles. Adams (2009) argues that these negotiations are political in this sense, where the nickname is a token in the

re-territorialization process of a particular identity (be it a place, person or community). He describes these ‘naming contracts’ as agreements between the named and the namer, which are ultimately “*politically focusing social objects.*”

Adams’ study explored the social world of the characters from the Peanuts comic strip. In this fictional community of practice, group solidarity and power relationships are reified through nicknaming practices. Adams’ study, therefore, provides an example of naming which has some important parallels for an investigation of an online nickname. In both cases, the stakes are low, yet the issue is still meaningful to participants. Like the names of Peanuts characters, pronunciations of ‘MeFi’ can be ‘politically focusing’, reflective of ideologies contained within a particular social world.

2.6.4 Product and Brand Names

“...the marketer and manufacturer has virtually become a kind of word magician, creating codes of meanings that are, literally, their own reward.” — Danesi, 2011, p. 178

Branding as we know it today had its start in the late medieval period. Tradesmen and guild hall owners adorned their shop fronts with signs containing symbols related to their trade, signaling to others their “trademark” or specialty. During the industrial progress of the next several centuries, society began to view these trademarks as more than simply referential tools — they became resources for positioning brands within a (social) marketplace. Over time, this new shift in perspective contributed to today’s societal orientation towards brands as cultural symbols in and of themselves (Danesi, 2011).

Similar to knowledge about registers and their uses, awareness of the social landscape of brand images demonstrates a particular type of social competence. Danesi (2011, p. 178) characterizes this as “*a subtle form of imaginary social status climbing, replacing membership in organizations that sustain such climbing in reality.*” Nowhere might this be more relevant than in CMC environments, where recognition of and competent engagement with social brands (and the forms or registers associated with those brands) from Academia.edu to Ziggs and beyond can earn internet explorers credibility and capital.

2.6.4.1 Branding as Strategic Identity Construction

Branding is one of the most conscious and carefully-constructed forms of naming practices today. The main goal of marketers in this endeavour is to evoke a specific set of indexicalities through a strategically designed name that the consumer can easily access, consciously or unconsciously (Danesi, 2011). The strategies used to achieve these goals are varied and many of them today rely on using linguistic features that are popular or otherwise

favorably perceived (usually carrying some form of (covert) prestige or distinction), such as is found in rapidly changing speech genres like CMC.

More recently, naming strategies have focused on symbolic aspects, incorporating elements perceived to originate from CMC genres and specific internet registers. Features of acronyms, abbreviations and letter-number combinations can confer a modern, edgy feel to the product to which they refer, often indexing youth, techno-savvy and social competence. One currently popular symbolic feature is the lower-case ‘i’ prefix, indexing “*imagination*”, “*internet*”, “*ingenuity*”, and “*intelligence*” (Danesi, 2011, p. 183). Also popular are letter substitutions, seen in products like ‘DataViz[®]’ and ‘Krispie Kreme[®]’. These are interesting examples in that along with inviting spelling variation, these names sometimes introduce phonetic ambiguity for their spoken counterparts, such as in ‘Imgur’ and ‘Num3ers’. Speakers may resolve these ambiguities automatically, unaware that the pronunciation they have chosen is one of several possibilities.

Other, similar brand names may demonstrate pronunciation variation for reasons having little to do with their orthographic form. For example, names that have been co-opted from other cultures (e.g., ‘Adidas’, ‘Nike’) may be only partially phonetically assimilated into the new environment or speakers may be unaware of pronunciation particulars that are linked to the name’s origin. At other times, these borrowed brand names are fully nativized and popularized as such.

For these co-opted names, particular variants can index different levels of categorical linguistic “belonging”. For example, the American sandal brand ‘Teva’, derived from the Hebrew word for ‘nature’ is often pronounced “tee-vah”, [tivə], despite the company’s claim² that it should be pronounced “teh-vah” (which is itself ambiguous; should the first vowel be more like [eɪ] or [ɛ]?). Speakers using the “teh-vah” pronunciation may be implicitly indexing their alignment with the brand, their knowledge of Hebrew (and perhaps a Jewish ethnicity), or both. Returning to the earlier discussion of ‘Missouri’, this becomes another ideological struggle between local associations versus etymological or prescriptivist ones (Read, 1933). Similar to Silva, et. al.’s (2011) study of ‘Iraq’ pronunciation variation, it also signifies outcomes dependent on experiential factors (e.g., whether one speaks Hebrew, is Jewish).

In his study of brand naming strategies, Danesi (2011) refers to the creation of a “meaning code”, which serves to create the identity for the brand through the implicit expression of its function or characteristics. This identity performance exists at two basic levels. First, the name must referentially identify the brand amongst its competitors. This is its denotative function. The second level of brand identity creation through naming consists of

2. <http://www.teva.com/Our-History/history,default,pg.html?>

its connotative function, the meaning code of indexical references that the name is associated with. This meaning code can convey the quality or prestige of a brand, the audience it is created for, or the social status ownership affords, among other things. Examples of the most successful meaning codes can be seen in designer clothing and luxury cars, especially where the prestige of the name is greater than the function or quality of the product itself. This is perceived as akin to works of art, where value and social capital are attained by proxy of ownership of a piece crafted or authored by a highly regarded name.

In such cases, pronunciation ambiguity may even work as an asset to the brand, where only particular clientele may be aware of the “correct” or in-group pronunciation. Examples of high-end brands with prestige pronunciations are ‘Porsche’, ‘Guerlain’, ‘Hermès’, ‘Courvoisier’, etc. For these names, pronunciation acts as a sort of secret code, allowing those “in the know” to recognize each other or simply demonstrate their brand awareness. While this sort of pronunciation prestige is unlikely to be occurring with the M-Set, some in-group versus out-group positioning may be unintentionally happening, as more involved participants become aware of pronunciations preferred by fellow MeFites.

2.7 Summary of Literature Review

In this section, several concepts were introduced, spanning many research disciplines. An overview of community models was given, with a focus on the CoP model. Following this, a review of CMC concepts and previous research was presented. In-depth analysis of registers, indexicality, and enregisterment provided a basis for understanding this research. Lastly, an exploration of onomastics gave some insight and context about the variables under investigation here.

This literature review has provided the general background necessary for contextualizing this research. In the next chapter, a specific background for MetaFilter and the M-Set variables will be given, so that the unique social environment in which this study takes place can be fully understood.

Chapter 3: MetaFilter and the M-Set Variables

3.1 Introduction to MetaFilter and the M-Set Variables

The picture that I present here of MetaFilter, its participants, and their language practices are based on over eight years of almost daily social engagement with the MetaFilter community, in addition to focused participatory research on the site. More days than not, I spent several hours on MetaFilter conversing with others, reading and writing posts, learning about and researching various topics, and building a vast database of relevant articles and bits of useful information. On days busy with offline activities, I would stop by the site to see if I had missed anything of interest, and to perhaps mark a post or two as a “favorite” (thereby allowing me to easily return to it later). Regardless, my investment in MetaFilter has been consistent and enduring. I have benefitted personally, academically, and professionally from the time I’ve spent there over the years. I feel that MetaFilter is a wondrously fascinating place that both researchers and laypersons can continually learn from.

3.2 The MetaFilter Community — Culture and Context

Established in July 1999, the online community blog MetaFilter began as a place for website creator Matt Haughey and his friends to share and discuss interesting links they had found online. Over the years, MetaFilter has maintained this primary focus, but has also expanded with a highly successful Q&A subsite (AskMetaFilter), a site-related discussion area (MetaTalk), and six other smaller subsites: Projects, Music, Jobs, Podcast, IRL (“In Real Life”, a place where MeFites can plan events or MetaFilter meetups, to gather and socialize offline), and FanFare.

3.2.1 The MetaFilter Subsites

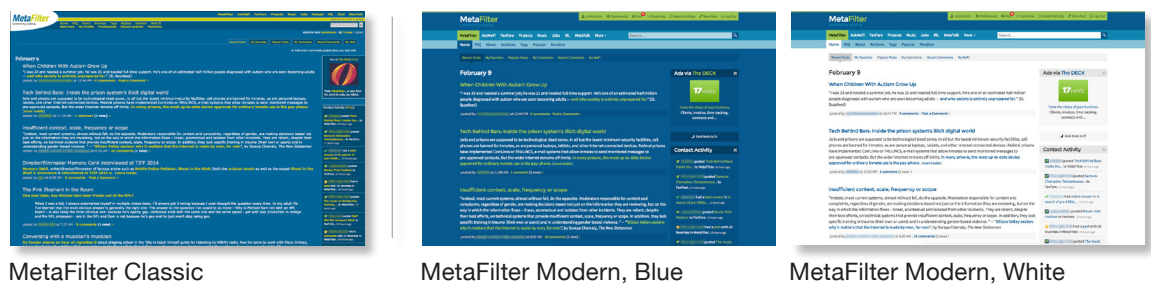
Understanding the culture and the linguistic context of the community in this investigation requires an exploration of the distinct social spaces on MetaFilter. The three main subsites — namely, MetaFilter (also the name of the site itself), AskMetaFilter, and MetaTalk — differ in their focus and quantity of content. These differences dictate the type of communication that occurs on the subsites and therefore also influence the participation styles and social norms that emerge there.

However, some features are universal across the subsites, such as the use of bold yellow text to signify linked content, or the option to click on posts or comments to save them as “favorites”. These features help to establish a cohesive theme for the MetaFilter community across all of the subsites, while at the same time preserving each subsite’s distinct focus and style. For example, the layout and format of all of the MetaFilter subsites and the

corresponding post pages are virtually identical, but the color schemes and taglines vary, allowing each subsite to be uniquely identifiable. This duality between common features versus location-specific ones establishes each subsite as a distinct way to experience and interact with the community, while at the same time still being an integral part of a broader MetaFilter identity.

It should be noted that toward the end of writing up this thesis, MetaFilter implemented its first official site-wide redesign. While the colors and themes remained, the overall aesthetic was modernized to fit in line with current internet styles (e.g., updated fonts, more white space, modern tab and menu bar styles, etc.). MetaFilter members can choose which view they would like to adopt in their profile settings. As all of the research here took place before the redesign, the screenshots, examples and visual descriptions of the site in this thesis will be referencing the classic theme.

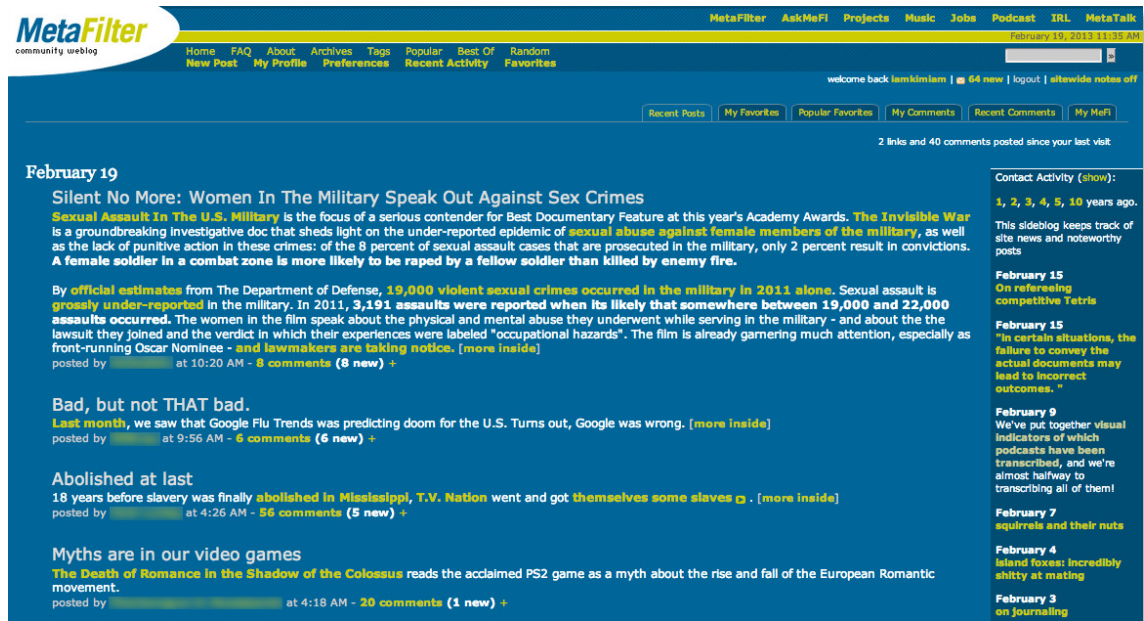
Figure 3. Screenshots of the MetaFilter Redesign, Before (Classic) and After (Modern)



3.2.1.1 MetaFilter, “The Blue”

The MetaFilter front page (shown in **Figure 4**, sometimes referred to simply as “MetaFilter” or “MeFi”; also nicknamed “the Blue”) is the most informational and outwardly-focused area of the site. Self-promotional comments are highly discouraged, as is directing the discussion toward oneself or other specific members of the site. Generally, personal stories are only shared here when they are relevant to the topic of the post. However, the moderators and much of the community recognize that there is a lot of leniency within this framework. Knowing when and how to flout such norms is part of understanding the site culture, navigating the register and demonstrating competence in knowing how to do so as an in-group member. Many participants are taken to task by other community members when norm violations occur, and yet many others consider it a rite-of-passage to make such errors (and learn from them).

Figure 4. Screenshot of the Front Page of MetaFilter, “the Blue”



3.2.1.2 MetaTalk, “The Gray/Grey”

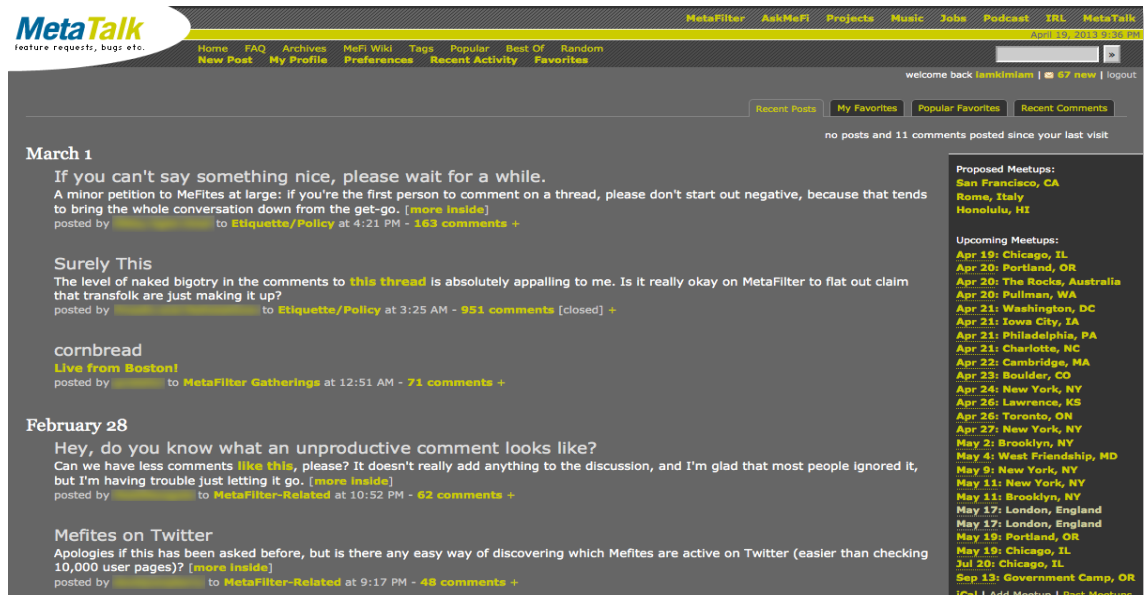
MetaTalk (i.e., “the Gray/Grey” or “MeTa”), shown in Figure 5, was the first subsite added to MetaFilter. It was launched in March, 2000. MetaTalk was created to allow site-related issues to be discussed in a separate, dedicated space created specifically for that purpose. Issues that participants may have with posts or with other site members can be debated on MetaTalk, rather than on other areas of the site, where such conversations or arguments may distract attention from the post topic. As a result, MetaTalk provides a way for the community to self-regulate, to establish norms and to create policies.

MetaTalk is the most internally-focused area of MetaFilter; the discussions that arise are either directly focused on the site or on other members (or both). The posts on MetaFilter may range from feature requests to policy and moderation debates, notable achievements by community members, or lighthearted enquiries through which participants can get to know each other better (e.g., “Show us your desks” whereby participants shared links to pictures of their workspaces and discussed the ways in which they experience MetaFilter).

The inward-facing and introspective ethos of MetaTalk naturally lends itself to being one of the more exclusive, in-group areas of MetaFilter. Those who participate on MetaTalk are likely to be more interested in the inner workings and development of the community than those who solely read or comment on the other subsites. Participants involved in MetaTalk may be helping to effect changes in site culture and behavior, regardless of whether or not they are aware of this or are intending to do so. Frequent MetaTalk participants (and readers) might also be more knowledgeable about the site history, as well as about the identities of fellow participants, especially as those participants are taken to task or are otherwise having attention

drawn to them on MetaTalk. Knowledge about the site and its userbase may be a form of capital for such participants, bringing with it possible benefits such as higher social status within the community and greater positive reception of their site contributions.

Figure 5. Screenshot of the Front Page of MetaTalk, “the Gray/Grey”



3.2.1.3 AskMetaFilter, “The Green”

The AskMetaFilter subsite, nicknamed “the Green” or “AskMe” (shown in Figure 6) was started in 2003 as a way for participants to crowdsource solutions to problems they are experiencing. The types of questions that get posted on AskMetaFilter range from deeply personal (they can be asked anonymously if need be) to complex, obscure, or lighthearted ones. Topics can range from recipes and dietary concerns to how to do repairs, personal development, or seeking understanding on a subject.

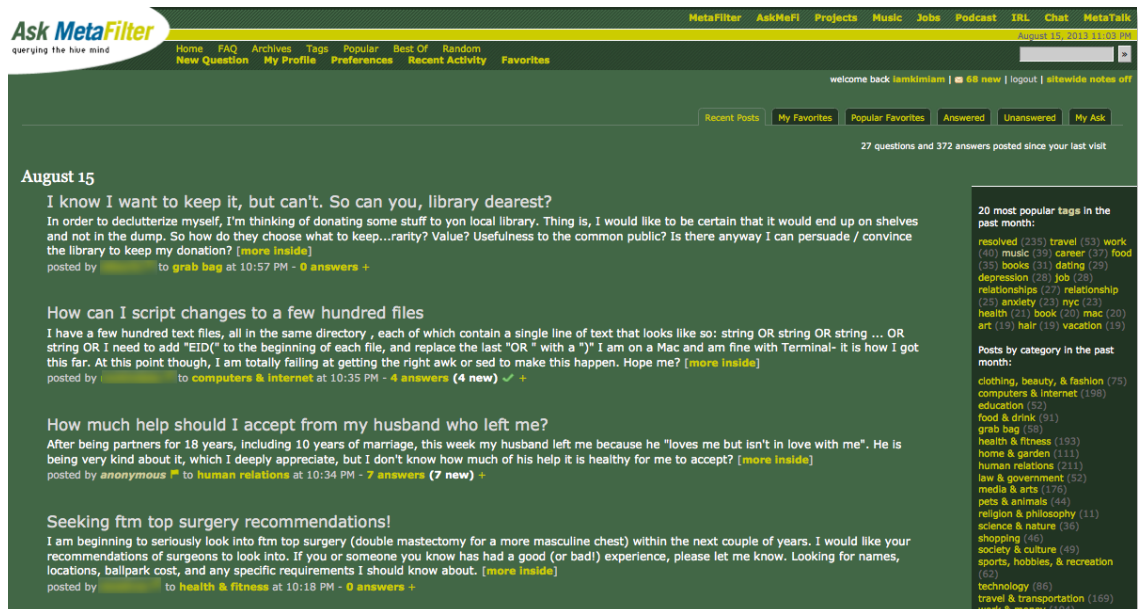
The types of responses to AskMetaFilter questions vary in both quantity and diversity. Questions regarding interpersonal matters often garner a lot of responses, commensurate with the range of ways people across the world and of varying backgrounds might address such concerns. This sometimes leads to contentious discussions, which may need to be addressed in MetaTalk. Other question types may require specific expertise, which can result in fewer responses overall, but also tend to receive more targeted answers from participants who are knowledgeable in those areas.

The purpose of AskMetaFilter is therefore to lead to discussions which are focused on the personal concerns of the community members, even though the topics themselves are external to the community (i.e., they are not about MetaFilter itself or what the community of practice does). As a result, some personal information about participants may be given, as well

as stories or opinions, within the context of solving the issue at hand. Therefore, participants reading or commenting on AskMetaFilter may learn more about the identities of their fellow community members than they would on the main MetaFilter site.

However, AskMetaFilter is much more heavily moderated than MetaFilter and comments which do not help to answer the question at hand are deleted. On MetaFilter posts (i.e., on “the Blue”) comments are often more like a conversation, where participants may joke, debate or share stories, so long as the comments match the tone and seriousness of the topic and do not derail the discussion. It can therefore be argued that both MetaFilter and AskMetaFilter provide opportunities for participants to learn personal details about other members of the MetaFilter community, albeit in different ways.

Figure 6. Screenshot of the Front Page of AskMetaFilter, “the Green”



3.2.1.4 IRL and Meetups

Other MetaFilter subsites have developed over the years, but none of these has achieved popularity or participation levels comparable to those of the three main subsites. However, a more recently-added subsite, in June 2010, named IRL (a netologism meaning “In Real Life”), is slowly but surely gaining equal footing. In this area of MetaFilter, participants can plan MetaFilter “meetups” and organize offline events that involve MetaFilter members.

MetaFilter meetups are a popular activity for many MetaFilter members. These gatherings can occur anywhere in the world, but are most frequent and well-attended in big cities such as New York City or San Francisco. A typical meetup occurs at a restaurant or bar and generally consists of two or more people conversing for several hours about the site, their lives, and whatever topics may interest them. Family and friends are usually welcome at

meetups, and pets have been known to make appearances as well.

Meetup participants often don't already know each other in a FtF context, which is a frequent topic of discussion in planning the meetup online. Other common topics of discussion in posts on IRL concern how participants will be able to recognize each other as MeFites, what is expected at a meetup, transportation and travel concerns, as well as any details specific to the meetup event.

While meetups are commonly held at restaurants, other past meetup occasions have included concerts, protests, conferences, beaches, and campsites, to name a few. Attendance at meetups has ranged from one person (e.g., MeFite eotvos at the South Pole for the Tenth Anniversary celebration³) to over thirty attendees (e.g., the Guy Fieri restaurant event⁴).

I have personally benefitted by planning meetups at almost every travel opportunity I've undertaken in the last several years. I've found lasting friendships and made valuable professional and research contacts with people whom I'd only previously conversed with online. Additionally, these MetaFilter events have provided a way to familiarize myself with new surroundings — meetups sometimes turn into guided tours, as I've found that MeFites often like to show fellow online friends around the cities they live in. These positive outcomes of meetups that I've described are frequently echoed by other MetaFilter members, both on the site and in FtF discussions (at meetups or elsewhere).

Another common discussion topic at meetups is the pronunciation of the M-Set. This topic arises frequently, as the site is often referred to by nickname in casual discussions during a meetup. Therefore, differing pronunciations are noticed and commented upon; for those new to meetups, this may be the first time they have heard the terms spoken aloud. As a result, the discovery of pronunciation variation in the M-Set (and in many MeFites' usernames) and their ensuing discussions is bound to recur, as attendees and locations vary over time.

As the IRL subsite is designed to allow participants to coordinate a place to meet and to get to know each other in non-CMC spaces, it can be the most participant-focused area of the site. Those who are active in IRL may acquire in-group knowledge about the goings-on of others, but not necessarily about the site itself. In this way, IRL participation is an in-group activity whereby participants can form stronger bonds with each other (increasing bonding social capital), possibly in lieu of continued participation with the community that initially brought them together (which may decrease bridging social capital). This trade-off of social capital types relating to FtF community involvement was explored by Sessions (2010) in a study focusing on MetaFilter participants' offline versus online engagement (described in greater detail in **3.3.2.1 Online Community, Offline Meetups, and Bridging Social Capital, p. 67**).

3. <http://ten.metafilter.com/127/Antarctic>

4. <http://metatalk.metafilter.com/22250/META-FILTER-RUN-FLAVORTOWN>

3.2.1.5 *Other Subsites — Podcast, Music, Projects, Jobs, FanFare*

Prior to the establishment of the IRL subsite, meetups were posted and planned on MetaTalk. Similarly, the first few MetaFilter podcasts (starting in February 2007) were posted only to MetaTalk as well; there was no subsite dedicated to that purpose at that time. The Podcast subsite was launched in mid-2007 (exact month unknown), and mainly serves as an archive for the podcast posts; podcasts are cross-posted to MetaTalk as well, where they reach a wider audience.

The approximately one-hour-long podcast has consisted of two or more moderators (and sometimes podcast guests) discussing recent events on the site, including commentary on notable posts from each of the subsites. The Podcast airs approximately once a month and is available on the site as well as in Apple's iTunes media player.

Since October, 2012, the crowdsourced monthly transcription of the podcast has become another welcomed and appreciated aspect of the community. This effort to transcribe the podcasts has succeeded in a nearly complete set of transcriptions for all of the episodes to date. Several regular podcast transcribers have embraced the task and some have even gone so far to learn IPA so that pronunciations of usernames and the M-Set can be accurately transcribed. In other cases, eye-dialect spellings are used. Regardless, pronunciation variation has been captured and replicated in this new medium, highlighting MeFites' attention to detail and the importance of linguistic matters as a community value that participants attune to.

The Podcast is a popular resource for more invested (e.g., core) MetaFilter participants to learn about the site and other members of the community. With the addition of the transcripts, the Podcast is now accessible to all participants as well. However, the infrequency of the Podcast means that while this resource is information-rich, it is ultimately a rare commodity, and may be unheard by the majority of the userbase.

MetaFilter Music began in 2003 as a place for musically-inclined participants to showcase, share and discuss songs they have created themselves and/or in collaboration with others. Occasionally, music tracks will be highlighted on other areas of MetaFilter (e.g., the MetaFilter Podcast), or even on other areas of the web and beyond. Irrespective of these spikes in notoriety and attention, MetaFilter Music carries on with a very small but strong following, dedicated to discussing issues related to the music-making process.

Growing out of an increasing need for participants to be able showcase their own web-based creative endeavors, MetaFilter Projects was established in November 2005. In addition to the commenting feature available to all subsites, Projects includes the ability to vote on projects, as well as a link for MeFites to post the project to the front page of MetaFilter, if they feel it is worthy of broader attention. Projects hosts approximately one to three new posts per day, and the average post receives fewer than five comments.

MetaFilter Jobs was launched in 2006, allowing participants to post job openings for paid positions or projects. Occasionally, a job posting will receive attention through a mention on the Podcast (e.g., if the posting is particularly amusing or otherwise unique). However, with fewer than 10 posts per month on average, Jobs shows the smallest amount of engagement of all the MetaFilter subsites.

Launched in April, 2014, FanFare is the most recently added subsite, where participants can post about and discuss TV shows, movies, and podcasts. This subsite will eventually be expanded to cover books, video games and other forms of entertainment. Even though FanFare is still in beta at the time of this publication, the subsite has a devoted following and several posts are made daily (more accurate statistics are not yet available).

3.2.2 A CMDA Outline of MetaFilter

The overview of the various MetaFilter subsites in the previous section provides a general picture of the constellation of MetaFilter, e.g., the areas one can visit and their relative size and popularity. This will be supplemented with a description of the important situational (facets of situation) and structural (facets of medium) features of the community, using Herring’s (2004, 2007) Computer-Mediated Discourse Analysis (CMDA) classification model (see 2.4.4 **Describing the Features of an Online Community**, p. 25). Each facet is described in detail following Table 1. Insights from my personal experiences with MetaFilter, based on over eight years of community participation, will be included where applicable.

Table 1. CMDA Etic Grid Classification of Medium and Situational Facets of MetaFilter

Facets of Situation	MetaFilter’s Orientation
Purpose	Informative and social
Theme	Variable
Tone	Informative, intellectual, casual, cooperative, friendly, nerdy
Community visibility	Public but bounded
Participant visibility	Anonymity at the participant’s discretion
Participation structure	Messages are always one-to-many
Participant characteristics	Diversity of age, gender, occupation, etc.
Facets of Medium	MetaFilter’s Orientation
Synchronicity	Asynchronous
Persistence of transcript	Persistent
Channel(s) of communication	Primarily text, but other means are available
Anonymous messaging	Yes
Private messaging	Yes
Quoting	Manual
Message format	Variable

3.2.2.1 *Facets of Situation — MetaFilter’s Orientation*

Purpose: The main purpose of MetaFilter is to share and discuss interesting things found on the web, making the general environment one that is both informative and social. As shown in Figure 4. Screenshot of the Front Page of MetaFilter, “the Blue”, p. 50, this is typically achieved by a participant researching a topic, gathering relevant links surrounding the topic, and framing them with text in a paragraph-like structure. This allows the reader to understand the gist of the post, while also allowing them to click on various links to learn more about different aspects of things shared in the post. These links and their content are discussed in the comments following the post. An example of a typical medium-sized post and discussion is shown below.

Figure 7. A MetaFilter Post Page and Comments



Theme: There is no overarching theme on MetaFilter; members post on any topic they find interesting. Mostly owing to the demographic makeup of its userbase (and those participants’ interests), frequent topics discussed on MetaFilter are technology, science, popular culture, politics, art, music and history (Lawton, 2005, pp. 36-37).

Some topics are known to be contentious on MetaFilter and are necessarily moderated much more heavily than others. Topics that fall into this category are the Israel/Palestine conflict, cat declawing, (male) circumcision, as well as various issues surrounding obesity, gun laws, gender identity, sex, and religion.

Tone: The atmosphere of MetaFilter can be generally described as informative, intellectual, casual, cooperative, friendly, controversial, and “nerdy”. The tone of MetaFilter

posts vary wildly, based on the post topic/content and current events (both outside of MetaFilter and within the community). For example, during times of disaster or tragedy, community members may be deeply involved in highly-contextualized, nuanced discussions about various aspects of the event.⁵ Comments in these types of posts may be very technical, personal, compassionate, and tend to be longer than comments in other, less emotionally-charged threads. Conversely, posts which are more lighthearted in nature, such as those about humor or containing short, amusing videos, may allow MeFites' sillier and more joyful qualities to shine.⁶ These types of posts tend to generate many shorter comments, which often perpetuate the tone of the post through puns, additional links, and personal stories (where it is applicable and would not be considered self-promoting to do so).

Community visibility: MetaFilter is a public but closed community; even though anybody can read most sections of the site, there is a one-off \$5 fee to join and contribute content. This creates an important boundary between members and non-members and serves to reduce random, "drive-by" comments from those who do not have a vested interest in being a part of the community. From a CoP perspective, this boundary not only serves to define who is outside the community, but also works to "*keeps insiders in*" (Wenger, 1998, p. 113).

Participant visibility: Some meta-data is built into the site automatically and cannot be hidden, while other data is at the discretion of the user to provide and make visible. Visible public profile data includes each user's online name, their join date, contributions made (divided into number of posts and number of comments, for each subsite), frequent tags used in posts made, number of favorites received and given (for any post or comment on any subsite), and contact data (whom the user makes a contact, and who makes that user a contact).

Meta-data which is at the discretion of the user to provide includes, but is not limited to: the user's real name, profile photo, birth date, geographical location, occupation, gender, relationship status, links to profiles on other social network sites, relationship categories with those they designated as contacts, nearby MetaFilter users (according to geographical location, if provided), and a space for writing a biography or sharing additional information in text. Some of this user-provided meta-data can also be set so as to be visible only to logged-in MetaFilter members, again at the user's discretion.

Personally, my profile on MetaFilter has become less anonymous over the years. This has been partly motivated by the research I've undertaken here. Prior to that, however, I had gradually disclosed more information about myself over time (e.g., personal websites, location, gender, and age). This was natural progression; I have become closer to the community

5. <http://www.metafilter.com/137298/MH370-missing> provides a good example of lengthy, technical and nuanced discussion concerning a serious topic (the Malaysian Airlines Flight MH370 going missing).

6. <http://www.metafilter.com/122342/I-is-all-you-can-say-is-purple-p-yes>, about bad jokes created by children, is a good example of one of the sillier extremes in which MeFites can participate together online.

through posting comments and responding to others, as well as by learning about my fellow community members. Owing to this perceived closeness, I have shared more about myself and have felt more comfortable and open to do so. This general sentiment has been corroborated by other MeFites in discussions about their feelings of affinity with the community and their comfort with disclosing information on their profiles. This is not always the case with MetaFilter members, as some have found the site off-putting in various ways. In general, however, the MetaFilter culture is one that values privacy and anonymity; the community places less importance on finding out personal details of fellow MeFites, than say, on reading about what those MeFites have to share about a given topic.

Participation structure: Messages are always one-to-many, as all communications are made by one individual to the entire public (including non-members of the site). All messages are marked with user names and timestamps.

Participant characteristics: Members of MetaFilter span all ages, genders, occupations, etc. There is a US-centric liberal bias and a high prevalence of technically-minded individuals who are interested in computers and related subjects. The site has been male-dominated, but less so in recent years. Gender and other participant characteristics are covered in greater detail in **Chapter 5: Data Results, p. 110**.

3.2.2.2 *Facets of Medium — MetaFilter's Orientation*

Synchronicity: Messages on MetaFilter are asynchronous; they are stored on the site until they can be read. This allows participants to read and to participate at their own leisure, and is therefore inclusive of those in differing time zones from all over the world.

Owing to the majority of the MetaFilter userbase residing in North America, posts and comments are more frequent during North America's daylight hours. However, this does not negatively impact the overall quality or quantity of responses on any given post on any subsite — there are always participants engaging with the site in some part of the world, and many participants enjoy catching up on the posts and comments they missed during their time away from the site. Additionally, in recent years there has been 24-hour moderation by MetaFilter staff members. This has meant that issues that have arisen during the North American nighttime on the site are dealt with in a similar time frame and manner as is done during North American waking hours.

Persistence of transcript: Posts and contributions remain on the site indefinitely, resulting in a persistent transcript. Persistency increases meta-linguistic awareness, allowing participants to re-visit, track, or reflect on language (Herring, 1999, 2007).

MetaFilter members refer to past posts and comments regularly, often pointing out their own or others' earlier previous stances on topics. This becomes especially relevant in

situations such as when a MetaFilter member is being discussed in MetaTalk (oftentimes owing to inappropriate behavior, but sometimes for recognition as well), or when discussing matters of policy or style on MetaTalk. Therefore, the persistent transcript becomes a tool for accountability, precedence, and argumentation. Referring to past discussion helps establish what the community cares about, allowing the community to refer to their own history as proof of prior investment in a topic.

Channels of communication: The main channels of communication on MetaFilter are text-based. However, other MetaFilter-related discourse contexts exist and are available to members, such as IRC, specific topic-based discussion spin-off sites, and MetaFilter meetups at physical locations. The MetaFilter Podcast and MeFi Music are spoken-audio format.

While the majority of participants read MetaFilter and its subsites, and may not engage in all or any other aspects of the MetaFilter community, they may hear about the discussions or events that occur through these other channels, as participants who do engage more widely return to discuss those events in text. In this way, the text-based pages of MetaFilter serve as a central hub around which all other activity is connected to in the constellation.

Anonymous messaging: Messaging on MetaFilter is anonymous, in that members create and control their own profiles as they wish, disclosing or not disclosing information as they see fit. They can also post anonymously in specialized instances on AskMetaFilter, where no references to their profiles or usernames are given.

Private messaging: Launched on October 19th, 2007, MeFi Mail is a feature of MetaFilter which allows members to email each other directly; participants have the option of disabling MeFi Mail if they wish. As this feature is built into the site, social interaction and bonding are warranted and encouraged from the top down (as opposed to the onus being on the participants to seek out others' contact information, as well as to provide a means of being contacted by others).

It is difficult to assess the extent to which MeFi Mail is used and perceived favorably by other MeFites; without further research, I can only speak about my own personal experiences. During the times when I comment more frequently on the site (i.e., a more engaged mode of belonging), and am therefore more visible, I receive approximately one or two MeFi Mail messages per week. These messages are often from MeFites with whom I've never spoken (in text or FtF) and the content of the emails usually contains an enquiry or comment regarding a contribution I'd made on the site.

Quoting: Quoting other members is not automatic; quoted text must be copied and pasted at the member's discretion. The act of quoting allows comments to be reproduced and responded to more easily, but the site conventions for how this is done are formed organically

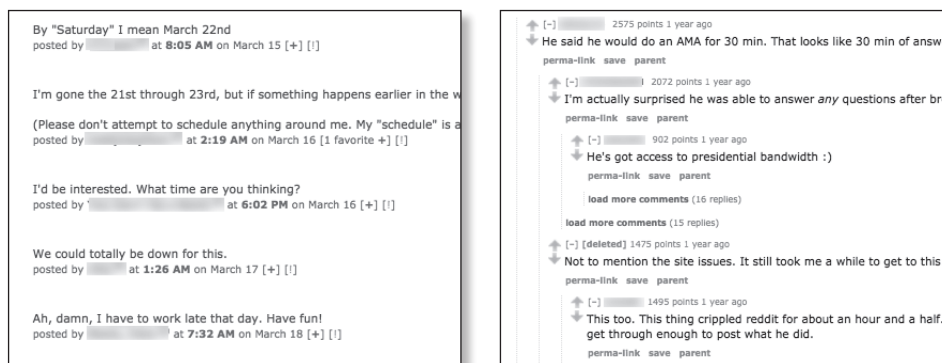
through use within the practice and are not imposed by the site itself (i.e., are not imposed by a built-in quoting feature).

However, several MetaFilter participants have contributed to writing browser scripts or miniature applications that allow quoting (and other tasks) to be done more easily. These features are usually designed and customized for a better MetaFilter experience. This is one way in which participants work together to contribute toward improving the community and their communication together online.

Additionally, when other participants make use of these creations, they implicitly endorse the betterment of the practice. Giving public feedback on these improvements helps make those endorsements visible, as do counters on the script pages (showing how many times the script has been downloaded and/or installed, as well as how many updates or comments the feature has associated with it).

Message format: On the main pages of the subsites, newest posts are shown at the top so that one can easily see the most recently-added content. On click-through to individual posts, newest comments are shown at the bottom so that the discussion can be read in chronological order. All messages are flat (not threaded), resulting in equal indentation and prominence for comments on the page.⁷

Figure 8. Flat Commenting Versus Threaded Commenting Structure



MetaFilter's flat commenting style

Reddit's threaded (e.g., nested) commenting style

Many discussions have ensued over the years regarding proposed changes to MetaFilter's flat commenting style (as well as the possibility of adding pagination features). These requests are addressed, but consistently overruled with concerns about the negative structural and social effects that these changes would impose. Many participants in these discussions state that threaded comments would fundamentally change their user experience for the worse, making the site feel less 'MetaFilter-like' and more in line with sites such as Reddit, which they'd prefer MetaFilter to remain visually and structurally distinct from.

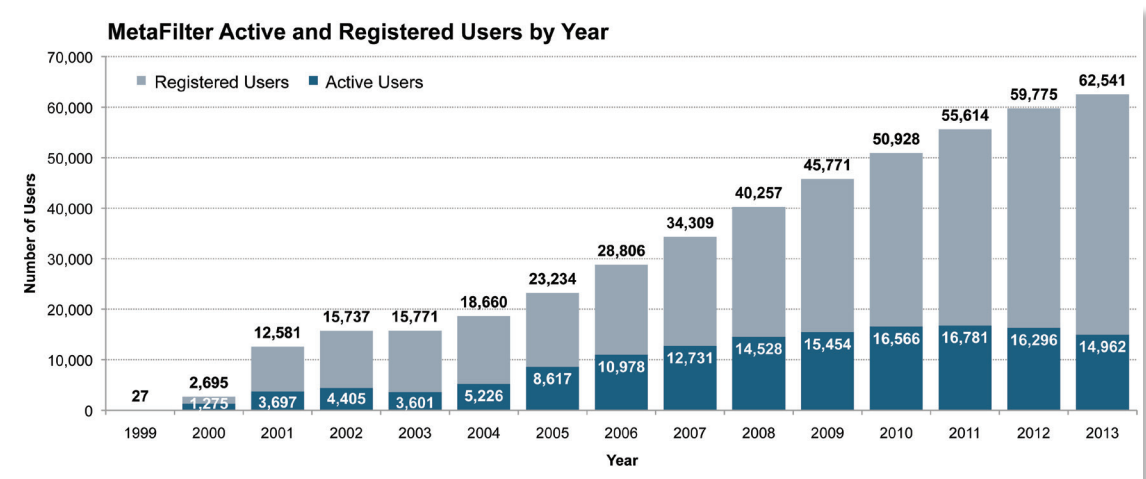
7. This is unlike similar sites such as Reddit, where comments can be "upvoted", with the most liked comments being moved to the top of the page and therefore given heightened visibility.

3.2.3 The MetaFilter Userbase

While MetaFilter is arguably the oldest community weblog in the world, it has not achieved the widespread popularity of similarly-focused discussion sites like Reddit or Digg. MetaFilter is relatively small in terms of its overall number of registered accounts, but the active accounts make up a substantial percentage of the registered userbase — on average, over 30% of registered users were active users in any given year.⁸ An active account/user is defined as any MeFite who made at least one comment or post in the year measured.⁹

Figure 9 illustrates the yearly number of active accounts as a portion of the total number of annually registered user accounts on MetaFilter.

Figure 9. The MetaFilter Userbase – All Registered Participants



There were over 190,000 registration attempts by the end of 2013 (as measured by existing user numbers, which are issued in a sequential order). However, only 62,541 users fully completed the registration process, resulting in a personal username, user number and profile page. Therefore, it can be said that MetaFilter had a total userbase of over 62,000 registered users as of the end of 2013.¹⁰

The overall number of registered users on MetaFilter has steadily increased since the site’s inception in 1999, but the number of active users on MetaFilter increased only up to 2011. In 2012 and 2013, a slight decrease in the number of active users was observed. The reason for the decrease in active users in recent years is not known and is beyond the scope of the analysis presented here.

8. The average was calculated excluding the first two years of MetaFilter, as they are outliers. This is typical of new communities, where growth is not stable until some time has passed.
 9. Many participants read the site regularly but do not comment or post. While these participants’ accounts are technically active (e.g., the participants have not abandoned or left the site), they are unable to be included as active users by this metric.
 10. This information was gathered from the MetaFilter InfoDump, which is a publicly available free download of various MetaFilter site statistics for all registered users in the community (see 4.2.4 Overview of the Types of Data Collected, p. 92 for more information).

It should also be noted that new user registrations were closed on November 9, 2002 and reopened on April 1, 2004 (with only 20 new registrations per day allowed at that time). On November 18th of that year, unlimited registrations began, but the \$5 one-time registration fee was implemented. These site policy changes account for the small number of new registrations in 2003.¹¹

3.2.3.1 *The MetaFilter Userbase and Enregisterment*

These basic statistics describing the MetaFilter userbase help build a general picture of the number of participants who may be actively shaping enregisterment, i.e., registered MetaFilter members who comment, post, or favorite items. These participants may be participating in message chains about the M-Set, either in an active mode of engagement or in more passive modes, such as imagination or alignment.

Over the last five years, the registered number of users increased from approximately 45,000 participants to 62,500 participants. However, only 25–35% of those participants were active, i.e., only approximately one quarter to one-third of registered users were visibly participating in the community in ways that might influence others' behaviors and understanding of the community and its norms, linguistic conventions, etc. (through their posts, comments, or alignment activities such as favoriting items). Of this smaller proportion, even fewer contribute with any frequency or consistency, e.g., are core participants. It is these highly active participants who may be more influential, owing to their heightened visibility and recognizability as core members. Additionally, these participants are more likely to be involved in community matters, and to participate in message chains relating to linguistic norms of use. This is important to note, because while the community is collectively the size of a small city, the immediately visible identity of that community has historically been shaped by a considerably smaller, self-selecting proportion of that group. It is through the activity of this smaller group that norms, features, etc. and their indexical associations are formed and propagated. These core participants' linguistic choices (and the resulting indexicalities and enregisterments) may not be reflective of the community as a whole, but this cannot be verified and assessed without methodologies such as those undertaken in this research.

3.2.4 *MetaFilter Comment and Post Frequency*

The next two figures illustrate the average daily contribution statistics from all active MetaFilter participants, separated by year and subsite. This gives an overview of the amount of participation that occurs on MetaFilter per day, as well as where this energy is directed (e.g., the most frequently updated areas of the site).

11. Some new members were allowed into the community at the discretion of the site owner.

Figure 10. Average Number of Posts Per Day on MetaFilter Subsites

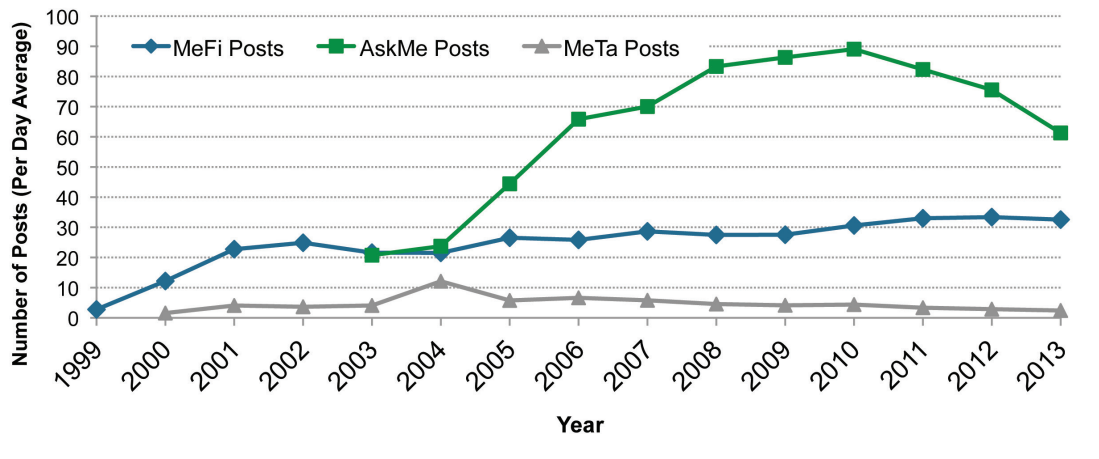
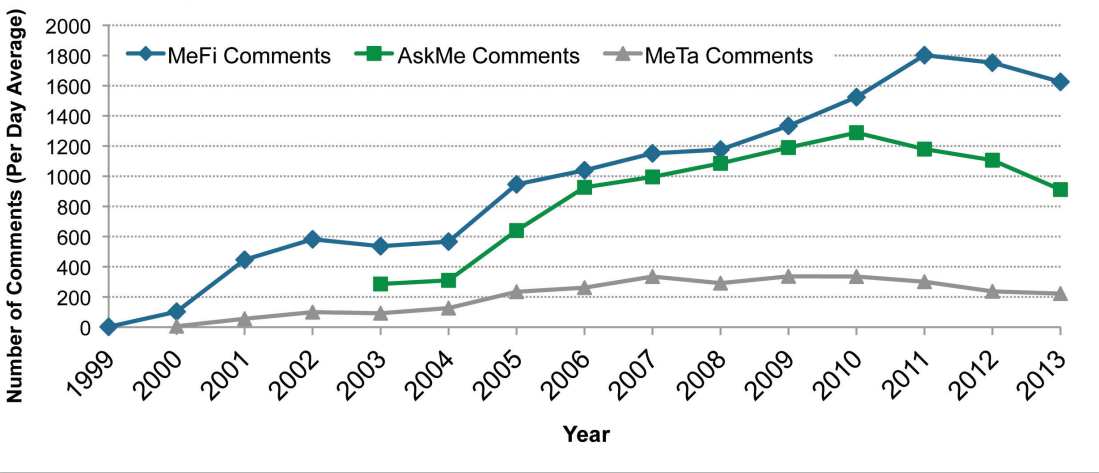


Figure 11. Average Number of Comments Per Day on MetaFilter Subsites



AskMetaFilter has the most daily content in the form of new posts, but receives fewer average daily comments than MetaFilter on those posts. In practical terms, more questions are being asked on AskMetaFilter than topics of interest are being posted on the MetaFilter front page, but the topics of interest generate more discussion than the questions. MetaTalk content (both posts and comments) is much less frequent than content on other subsites, which is to be expected given the focus of the MetaTalk subsite (i.e., site-related issues). The relationship between frequency and content of contributions on the main subsites and the enreregisterment of the M-Set will be thoroughly explored in **Chapter 6: Enreregisterment, p. 153**.

One final note about contribution frequency is that recent years show a departure from the trend of increasing quantity of content on MetaFilter over time. This correlates with a slight dip in number of completed registrations and active participants in recent years (see **Figure 9**). This decline in average number of contributions is most evident on AskMetaFilter, which has historically been largely driven by Google traffic. However, in November, 2012 Google made a change to their indexing algorithm, which resulted in lowering MetaFilter's rank in search

results and caused a 40% drop overnight in MetaFilter site traffic (Auerbach, 2014; Sullivan, 2014). Google's update might be partly responsible for both the decline in contributions and even new registrations, as the search engine's ranking equates to the visibility and perceived value of the search results (regardless of their actual value).

3.3 Previous Research on MetaFilter

Several previous CMC studies have involved MetaFilter, either tangentially or as their main focus. The research approaches of these investigations have ranged from models of community (Silva, Goel & Mousavidin, 2008, p. 55; Ali-Hasan, 2010) to social capital (Lawton, 2005; Sessions, 2010), and ethos and identity (Warnick, 2010). The findings from these past endeavors has laid a sturdy foundation for continuing research on contemporary perspectives on community and identity formation in CMC environments today.

3.3.1 Previous Study of MetaFilter as a CoP

It has only been within the last decade that researchers have applied the CoP model to online settings (Herring, 2004; Ali-Hasan, 2005; Stommel, 2008; Silva, Goel, and Mousavidin, 2008; see 2.2.1 Community of Practice — Definition, p. 9 for a review of the CoP model). This has helped to highlight particular CMC environments as meaningful and content-rich places for participants to carry out their social practices.

The three tenets of the CoP model — *mutual engagement*, *joint enterprise*, and a *shared repertoire* — are very present in the MetaFilter community. While there is no set theme or topic for MetaFilter posts, members are invested and engaged in sharing links and participating in discussions that adhere to the quality standards self-regulated by its userbase. The community is reified through participants' use of social conventions, the codification of norms (e.g., help pages, wikis), and the creation of site tools or addition of features to improve the experiences of community members.

These community-building activities are of continued interest to MetaFilter participants, but are not the main focus of their practice. First and foremost, MetaFilter participants are preoccupied with socializing, learning and being entertained online (Warnick, 2010). These objectives are often achieved through participation in ongoing community debates, ones often featuring wordplay and aggressive social banter. The stance-taking that occurs and the outcomes of these debates indirectly work toward improving the practice, co-creating a community identity and developing the shared repertoire. Participants' success in these endeavors confers legitimacy to the practice and positions MetaFilter as a thriving social environment that is recognized as such both online and offline.

In a 2005 study of MetaFilter, the social structure and participation properties of the

site and its participants were analyzed using the CoP model (Ali-Hasan, 2005). The results of this analysis showed how the non-hierarchical social organization of the site led to feelings of empowerment among its participants and an egalitarian ethos of the site in general. These factors have contributed to MetaFilter's long-standing success and exemplify why this online community has become a model for other, newer communities to emulate and achieve similar success. Through the application of the CoP model, this study also provided further evidence that online communities can and do share many of the same features of FtF communities, giving participants in either environment a sense of purpose and belonging.

In another CoP approach to analysis of MetaFilter, Silva, Goel and Mousavidin (2008) found that the success and cohesiveness of the community were attributable to six features: explicit ground rules regarding membership, the presence of moderators, the availability of profile information, "net etiquette", features for discerning pertinent posts, and the deployment of specific techniques of discipline (Silva, Goel & Mousavidin, 2008, p. 55). Their research on MetaFilter also resulted in four analytical generalizations, summarized as follows:

1. Site moderation is a necessary feature of community blogs.
2. The exercise of disciplinary power by "old-timers" (or other core members) is fundamental for cohesion. Displays of power are not obstructive, but rather facilitative, wielded to enforce norms, promote unity and keep discussions on track.
3. When participants' identities are connected to contributions through links to profiles attached to each post, reputations become both accessible and assessable.
4. Participation on MetaFilter is motivated by how well a participants' contributions are received by others.

These findings illustrate how knowledge of the site structure, moderation policies, profiles and identities are all integral aspects of accounting for the culture of language use and the establishment of linguistic conventions.

3.3.2 Previous Study of MetaFilter and Capital

The social hierarchy that results from the evaluations of contributions and their effect on the varying social status of community members means that influence to effect change is not uniformly distributed across the community. This power to influence others can be described in terms of social capital (see 2.3.3 Social Capital, p. 20 for a general overview).

Similar to the way Bourdieu's (1986) model of capital positions individuals within a FtF social hierarchy, Lawton (2005) showed that the three forms of capital on MetaFilter (cultural, social, and virtual) serve similar purposes, in that capital allows participants to position themselves within the social landscape of their CMC environment.

Cultural capital on MetaFilter — being conversant, having the ability to make good

posts, being able to comment on a wide range of topics, etc. — equates to a measure of status for a user through praise, recognition and visibility. Successful contributions of this sort presuppose knowledge about social norms, competence to use the register, and the shared perceptions of what is of value to other members. The demonstration and recognition of capital by others gives more credibility and weight to the contributions and the stances contained within them for such participants. As it relates to enregisterment, these participants have greater influence in the spread and adoption of ideas, features, and behaviors.

The second form of capital at work on MetaFilter is social capital, observable via the number and types of connections a participant has, as well as the amount of interaction they have had with other members. This can be seen on a user's profile page, where contact information is displayed (how many contacts the user has made and how many other users have contacted them) and the number of comments and posts that have been made in each MetaFilter subsite (reflecting the user's amount of engagement in discussions). Other metrics are evident on profile pages as well, such as number of favorites given and received and participant's user number (a proxy for how long the participant has been a registered member on MetaFilter) (Ali-Hasan, 2010).

The third form of capital that emerged from Lawton's (2005) exploration of MetaFilter, is "virtual" capital. This describes the user's ability to use CMC space effectively and to show competence in finding content, as well as formatting and posting it in line with the site norms and conventions. Virtual capital is observable through frequent participation and meaningful contributions to the community in ways that demonstrate internet fluency, including but not limited to providing links, technical solutions or improvements, and awareness of innovations and changes to CMC environments.

Lawton (2010) used these three forms of capital observed on MetaFilter to explore the ways, both positive and negative, in which participants can differentiate themselves. Two types of distinction were examined: celebrity (positive) and notoriety (negative), as well as the ways in which they can confer or diminish social capital.

Participants who were informally deemed "celebrities" within the site were seen to have acquired all three types of capital. Those who were similarly visible on the site, but less universally positively evaluated (the "notorious" participants) tended to have high levels of virtual capital, but low social capital. Regular participants on the site who were not highly visible (non-core or non-active levels of participation) or were known only in smaller circles were seen to have low virtual capital, but high social or cultural capital. The remainder of the participants (i.e., occasional or peripheral levels of participation) were a group found to have low levels of virtual, social and cultural capital. These users were less visible and made up the

majority of the site's participants (Lawton, 2005).

From these outcomes of Lawton's (2005) thesis, it stands to reason that participants with high levels of all three types of capital, such as found in "celebrity" MeFites, are not only more visible, but are likely more influential in the establishment of linguistic norms and conventions. Conversely, it is possible that "notorious" participants are influential in qualitatively different ways and/or their linguistic choices may be imbued with indexical values commensurate with how they are perceived. Regardless, the relationships between participants with high levels of capital and their influence in enregisterment processes (e.g., the spread of message chains) is of interest in this research, and therefore Lawton's (2005) study provides a groundwork for understanding these dynamics.

3.3.2.1 *Online Community, Offline Meetups, and Bridging Social Capital*

In a 2010 study using eight years' worth of MetaFilter activity data, the concept of online community was explored by examining the impact of participants' attendance at offline gatherings (MetaFilter meetups), both at the individual level as well as in terms of the effect on the community as a whole (Sessions, 2010). Data showed that these offline gatherings strengthened social ties between participants' online social networks and increased their social capital. Participants who attended meetups were more engaged with the MetaFilter community, but tended to prefer to maintain relationships with other MetaFilter participants who also attended meetups. As a result, these offline/online participants' ties with online-only participants were weakened, to the detriment of those participants' social capital (Sessions, 2010).

Sessions (2010) synthesizes her results to make an important point about the potential negative effects of offline gatherings, including the loss of weak ties, the decrease in online participation, and the dwindling of resources exchanged online. She warns that meetups should not be assumed always to be beneficial, and in some cases, they can have deleterious effects on community cohesion. For example, on the one hand, ties between members who meet up are strengthened (e.g., fostering bonding social capital) as the relationships become more multiplex (i.e., participants are connecting in a variety of ways). Therefore, the increase in engagement has direct positive effects on individuals within the community. On the other hand, this comes at the cost of sacrificing weak ties, which are crucial to the spread of information and innovation (e.g., fostering bridging social capital), and the health of the community as a whole. This often happens because members who meetup may go on to prefer exchanging support through offline interaction in place of their previous engagement with the entire community through text, which can have a negative effect on the majority of the community members, who do not attend meetups at all.

It therefore becomes interesting to see what effects, if any, offline interaction has on the establishment of linguistic conventions. Meetups are a way in which bonding social capital is fostered between FtF participants, often to the detriment of bridging social capital and weak ties between participants and the community as a whole. Therefore, it is likely that there would be more consensus on conventions between participants who frequently meetup (through both bonding social interaction as well as advantages that the speech modality of FtF interaction confers) than between those who do not meet up (and whose social interaction is more focused on bridging social capital, i.e., exchanging information and ideas).

3.3.2.2 *Summary of Previous Studies of MetaFilter and Social Capital*

These studies have demonstrated complementary ways in which forms of capital are integral components of community interaction online. An understanding of how individual users might be socially stratified within the community — correlating with different types and amounts of capital — is helpful toward assessing the influence and reach of their contributions. The social status of individual participants will no doubt have an impact on how their contributions might be received and how much influence those contributions will have on others. For the development of community norms and the establishment of language conventions it is necessary to explore this type of influence and power, as it can help explain the process of enregisterment via chains (see 2.5.2.1 **Message Chains**, p. 27). This is especially useful when analyzing the stances participants take as evidence for emerging trends, such as the pronunciation of ‘MeFi’.

3.3.3 **Previous Study of MetaFilter Ethos and Identity**

Another study of MetaFilter examined the concept of ethos and identity through a year-long ethnography (Warnick, 2010). Results showed that there was no one strict definition of ethos as it occurs on MetaFilter, but rather several, sometimes contradictory accounts of the concept coexisting on the site.

All communities co-create a collective identity and ethos, regardless of whether or not this is overtly or consciously done. In an online community such as MetaFilter, ethos is achieved through the text that is exchanged, which also reflect upon and influence each individual’s identity as expressed through their contributions. As such, all collaborations and gatherings are interpretable as identity- and ethos-building events (Warnick, 2010).

3.3.3.1 *Warnick’s (2010) Four Paradoxes*

Warnick also outlined four paradoxes about MetaFilter that researchers should be aware of when conducting their research on MetaFilter or other similar communities. Attention to

these paradoxes aids the researcher in understanding some of the seeming contradictions that are inherent in the way identity is constructed online.

The first paradox states “*Ethos resides in the text — except when it doesn’t*” (Warnick, 2010, p. 127). This concerns whether or not identities we come to know in online environments (including the identity of MetaFilter itself) are gleaned only from the text that participants read online (the Aristotelian view) or from a *vir bonus* notion of ethos (i.e., that identity is reliant on what people know about the author of the text). Warnick found evidence for both, while also illustrating a problem of locating such social information from outside of a text, when text is mainly all one has to refer to. He resolves this conundrum by pointing out that boundaries between CMC and FtF environments are increasingly blurred, and that online identities are not as anonymous as people might perceive them to be. Warnick concluded that the Aristotelian notion of ethos (i.e., that identity information is located within the text) and the *vir bonus* notion (i.e., that identity information comes from what we know about the text authors) are complementary and not exclusive to an online community. This is evident through looking at the participants on MetaFilter who are at the extreme end of the anonymity scale (e.g., having the most minimal profiles and providing few to no identifying details in their contributions), as well as the existence of some users who have extensive profiles, little to no anonymity, and cite their own credentials or authority when making contributions. These two oppositional contribution styles — and everything in between — coalesce to form a collective identity for MetaFilter itself.

This leads to the second paradox — that “*Collective ethos is the work of a few*” (Warnick, 2010, p. 128). This idea introduces another model for understanding identity: the concept of “a gathering place”, which is easily observable in online communities. The gathering place is predicated upon community members maintaining a collective identity. However, the results of Warnick’s (2010) study showed that this collective identity — including the policies and discussions of norms — are in fact shaped by a very small, self-selecting subset of the much larger MetaFilter community.

The third paradox states, “*The best dictators have no rules*” (Warnick, 2010, p. 130). While Matt Haughey is the sole arbiter of MetaFilter, his leadership style has consistently been less authoritative or corporate, and more personal and amiable. This has set the tone for the culture of MetaFilter, as the community ethos (by either the Aristotelian or the *vir bonus* definition) cannot be separated from its leader’s individual ethos — the two are inextricably tied. Therefore, Haughey (and the other moderators) have had substantial influence on participant behavior, even if that influence is in the form of an egalitarian, hands-off, non-authoritative approach.

Lastly, the fourth paradox asserts that “*Serious communities are powered by silliness*” (Warnick, 2010, p. 132). People communicating online often over-exaggerate their humor or sarcasm as a proxy for the inability to express vocal tone or other features found in FtF communication (Rheingold, 1993; Warnick 2010; Greiffenstern, 2010; LeBlanc, 2010). This may be misread by outside observers as uncultured, immature, or base humour. The paradox lies in the cleverness that is employed by many users in constructing such sentiments, which appear low-brow on the surface, but are actually often complex or intentionally hyperbolic. This type of humor on MetaFilter has become recognized as typical of the community style and therefore allows participants to index their in-group knowledge and insider status when expressing humor in that style. Additionally, letting others know that they are not a newcomer in this way provides credibility to their textual voice and allows them to be taken seriously when they are speaking as such. Jokes also increase solidarity and create opportunities for others to participate by extending or “riffing” on the joke. Lastly, jokes provide a way for community members to “lighten up” and release pent-up energy, clearing any bad air. These behaviors show that online communities do not always need to be serious to be credible. In fact, it is this silliness that allows the community to grow and thrive in a healthy way.

These four paradoxes inform possible outcomes regarding the negotiation of site conventions and innovative linguistic behavior on MetaFilter. Awareness of the ways in which ethos and identity are created on MetaFilter is knowledge about what ‘MeFi’ may mean to others, how stances about the nickname could be interpreted, and who has serious influence in an otherwise silly and humorous ongoing debate.

3.3.4 Summary of Previous Studies of MetaFilter

Over the years, there have been several studies of MetaFilter, from a range of theoretical perspectives. The CoP model has been applied to MetaFilter and found to be a useful way to describe a community that does not center around a single theme, but whose practice consists of several themes. Explorations of social capital showed how social status and hierarchy exist and are useful toward the healthy function of a community, even when such communities pride themselves on having a relatively flat social structure. Lastly, Warnick’s (2010) study of ethos on MetaFilter demonstrated how identity is created at multiple levels and involves both the content that is created on the site as well as the identities of the participants who create it.

3.4 The MetaFilter Register

All language communities display at least some explicit metapragmatic activities in the form of identifying terminology, register naming, identifying stereotypical or exemplary community members, codifying norms of language use and behavior, and discussing rites and

rituals (Agha, 2005). MetaFilter is extremely productive in this regard, with the [MeFi Wiki](#)¹², FAQ pages and archives of MetaTalk posts, in which participants have engaged in discussions about etiquette, standards and proper behaviors on the site.

It should be noted that the ongoing and repeated negotiation of language and behavior on MetaFilter occurs not despite the high turnover of participants, but because of it. The constant turnover of members is not necessarily a negative feature of online communities. New members bring with them new perspectives and ideas, as well as providing opportunities for existing procedures to be reviewed and renegotiated if necessary. These types of opportunities keep the community as a whole vibrant and interesting (Silva, Goel & Mousavidin, 2008).

Changes in the MetaFilter population occur much more frequently than in FtF communities. MetaFilter population shifts are concurrent with time zones, days of the week, holiday schedules, geographic-specific events and the waxing and waning of participants' interest in the community. These factors, and others, result in rapid progressions of online registers, meaning that the elements of the register and the societal evaluations the elements are imbued with are constantly changing. Many online participants — especially on MetaFilter — have come to expect these changes and embrace them. For these participants, online communities are places where content is perpetually new and interesting.

Despite the rates at which changes occur, online registers persist and the distinctive forms and their values are communicable to new participants in ways that are unavailable to FtF participants in physical speech communities. For example, MetaFilter's entire archive of commentary on any topic in its history is available for anyone with internet access to read and refer to. Participants' use of hyperlinks allow past discussions on topics — including specific instances of commentary or events — to be called upon and referenced in current discussions. Additionally, resources such as the [MetaFilter Wiki](#), [InfoDump](#), and [MetaFilter Corpus](#) create a record of prior events and data (see **4.2.4 Overview of the Types of Data Collected**, p. 92 for more information about these resources). All of these features contribute to a persistent means of maintaining historicity for the community in which newcomers may reference and learn from. These features are the mechanisms by which the register can persist over time, be communicated to new populations of participants, and be renegotiated if necessary (Agha, 2004, p. 27; Lawton, 2005).

3.4.1 Elements of the MetaFilter Register

The MetaFilter register consists of behaviors, conventions, memes, snowclones¹³, abbreviations and acronyms, and other netologisms. MetaFilter is extremely active in both

12. <http://mefiwiki.com>

13. Snowclones are linguistic tropes in the form of phrasal templates, usually employed in making jokes. The term 'snowclone' was invented by economics professor Glen Whitman (Pullum, 2004).

creating and discussing elements of the register, even if its participants are unaware that community- and register-building are one of many outcomes of their interactions. A few examples from the MetaFilter register are given in **Table 2**.

It should be noted that register elements do not always strictly fall into a single category in **Table 2**. In fact, category overlap is a feature of some of the most productive and creatively emulated elements of the register. Regardless, MetaFilter is continually updating and reifying its practice with behaviors and netologisms that personify the culture and identity of the community and its participants.

Table 2. Examples of Elements from the MetaFilter Register

Register Element	Explanation & Meaning	Variants (if applicable)
Behaviors and conventions		
the obituary dot	A comment which consists of a single period, placed in an obituary post, to signify a moment of moment of silence for the deceased.	An asterisk is sometimes seen instead of the period, to convey disapproval for the deceased. Other variants of the dot exist.
“threadsitting”	Refers to posting a thread (usually in AskMetaFilter) and then repeatedly commenting in it. This behavior is highly discouraged and usually prompts mods to step in and kindly advise to poster to stop.	“threadshitting” refers to leaving an unpleasant comment in a thread which serves to derail the discussion.
Memes and snowclones		
MetaFilter: taglines	The act of taking a previous comment and post-pending it to “MetaFilter: [comment here]” as a new comment, to form a tagline which serves to epitomize MetaFilter for humorous effect.	A catalogue of over 115,000 taglines can be found at: http://metachat.org/wiki/Taglines
Abbreviations and acronyms		
“FIAMO”	An acronym for “Flag It And Move On.” A directive for other participants to not engage in an unproductive discussion, but rather to flag the comment or post and move on.	Few variants for “FIAMO” exist, but a list of other acronyms can be found here: http://mefiwiki.com/wiki/Acronyms
“eponysterical”	A blend of “eponym” and “hysterical”, to express the idea that a participant’s username is humorous in light of their post or comment.	
Other netologisms		
“beanplating”	A lengthy discussion in an Alanis Morissette music video post prompted the comment: “HI I’M ON METAFILTER AND I COULD OVERTHINK A PLATE OF BEANS.” References to beans are now understood as a humorous way to express the idea of intense but ultimately pointless and possibly self-indulgent intellectual pursuit of a topic.	beanplates, beansing, etc.

Several additional examples of register elements can be found in **Appendix E, page 239**. Each item in the list has undergone its own individual process of enregisterment. Together, they form a constellation of linguistic forms that contribute to the register and represent MetaFilter style and identity.

3.5 A Linguistic Overview of the M-Set Variables

Arguably the best-known MetaFilter register elements are the netologisms ‘MeFi’ and ‘MeFite’. ‘MeFi’ is the nickname for the site and ‘MeFites’ is the name for the people who are members of the site.¹⁴ Like ‘MeFi’, ‘MeFite’ is a term that has been the subject of a pronunciation debate that has continued for at least thirteen years on the site. The two terms together have been labeled the ‘M-Set’, which are used throughout this thesis where it is necessary to refer to ‘MeFi’, ‘MeFite’, and all of their pronunciation variants.¹⁵

Although the pronunciations of the terms of the M-Set no doubt have an influence on each other, the pronunciation of ‘MeFi’ has generally been the main focus of the debate over the pronunciation of these terms on MetaFilter. Additionally, it is the first-syllable of ‘MeFi’ that garners the most attention in community discussions. Usually this concerns whether the first vowel (henceforth, V1) in ‘MeFi’ is pronounced as [i] or [ɛ], although other variants are possible and are debated as well.

The M-Set variables are high-frequency and essentially non-variable in text (barring variations in case or “eye-dialect” spellings for specific purposes), yet low-frequency and highly-variable in speech. This linguistic situation is not unique to the M-Set or MetaFilter — many new word forms are entering the language through text-based media such as CMC, and display orthographic/phonetic ambiguity, such as ‘LOL’, ‘doge’, ‘WiFi’, ‘SciFi’, ‘.gif’, and ‘Linux’. Many of these netologisms become shibboleths within their own social spheres as well, potentially separating people along sociocultural lines of geography, age, gender and expertise.

The M-Set variables are representative examples of an entire category of new word forms that are participating in language change. The M-Set and many similar netologisms often display interesting allographic features, such as CamelCase (mixed-case letters), and some even use non-alphabet characters, such as those found in the netologisms ‘.gif’ and ‘L33TSP34K’. Additionally, as a particular type of abbreviation (a blend), the M-Set variables further embody several of the possible features that a netologism can have.

The M-Set (and ‘MetaFilter’) also references major categories in the onomastic taxonomy. While ‘MetaFilter’ has no geographic boundaries, it has an identifiable place in CMC, complete with a locatable address (www.MetaFilter.com). ‘MetaFilter’ is also a proper name and the referential source for the abbreviated nickname ‘MeFi’. Both ‘MetaFilter’ and ‘MeFi’ are crucial in the designation of the people who belong there, i.e., ‘MeFites’.

From a social marketing perspective, the collection of the terms ‘MeFi’ and ‘MeFite(s)’

14. ‘MeFi’ was first coined by internet pioneer and content-creator Jason Kottke (username jkottke) in January, 2000. ‘MeFite’ was first coined by MetaFilter user aaron in March, 2001.

15. The M-Set is a term created for the purpose of this research, to be able to refer to the variables together and to avoid biasing others with a particular pronunciation when saying ‘MeFi’ or ‘MeFite’ aloud. The ‘M-Set’ is not known or used by MeFites, perhaps apart from those who are aware of the current research.

(the M-Set) form a product family, which constitute the brand that is MetaFilter. Additionally, elements of the shared repertoire of the community, such as resources or tools (MeMail/MeFiMail), spin-off sites ([MetaChat](#), [MeFight Club](#), [MonkeyFilter](#)), related projects ([MeFiMag](#), [mefi infodumpster](#)) and commodifications (MeFi shirts, MeFiSwap (CD) collections) also indirectly support and promote the MetaFilter brand. These reifications often adopt features from the enregistered MetaFilter naming conventions as a way to demonstrate their relation to the MetaFilter social enterprise.

‘MeFi’ can be further defined as a proper name, a nickname, and an online place name. Even though all names are indexical of the specific histories, identities and ideologies of the people and places they represent, place names in particular have a special indexical relationship because they are the culmination of political and power structures and reflective of a co-created group identity (Horsman, 2006; David, 2011; Tucker, 2011). Previous research looking into the socioindexical information that place names convey has revealed some ways in which names are used as linguistic resources to express the personal stances (Hall-Lew, Coppock and Starr, 2010) or affiliations (Silva et al., 2011) of speakers. Pronunciation variants of place names may be used to define and develop in-group and out-group identities as well (Read, 1933). For more detail on these and other ways place names are used as linguistic resources in identity construction, please see **2.6 Onomastics, p. 40**.

The M-Set variables are many things at once — they are both netologisms and abbreviations (specifically, blends), as well as belonging to several types of naming categories: personal names (and nicknames), (online) place names, and product names (as a social brand). The M-Set terms display sociophonetic variation and are used as linguistic resources in taking stances and constructing identities (both individual and group identities). All of these features make the M-Set particularly interesting to study, as speakers may be influenced by some or all of these properties, to varying degrees.

Returning to matters of sociophonetic variation, since it is difficult or unnecessary to convey pronunciation particulars in writing, and many MeFites read the site without commenting, it is not easy to establish a consensus on pronunciation trends. It is reasonable to assume that MeFites do not frequently hear variants of the M-Set other than instances from more prominent members of the community who are creating podcasts, attending meetups, etc. These differences in exposure to variants, involvement in the community, recognition of the M-Set as in-group identity markers, and other factors open up an additional range of possible rationales for why participants would opt for some pronunciations over others.

Predictions about the more dominant pronunciations can be made based on preliminary research on internal factors (e.g., English phonotactics, grammatical and grapheme-phoneme correspondence (GPC) rules), and to a limited extent, some social factors (e.g., general dialect

differences, what MetaFilter participants say about their own pronunciations of ‘MeFi’ in online discussions, etc.). However, until this research involving large-scale surveys was carried out, those predictions could not be empirically verified.

A brief grammatical sketch of the M-Set variables begins with an overview of the community name ‘MetaFilter’. An analysis of the M-Set variables is given, starting with orthographic features, then focusing on grapheme-phoneme correspondence (GPC) rules for the letters <e> and <i>. Types of phonetic differences which may lead to pronunciation variation are discussed, followed by an analysis of how possible vowel phonemes for the M-Set terms may be realized differently across dialects. Prosodic features are also covered. This preliminary research, focusing on internal factors, provides a basis for the chosen methodology and for interpreting the results of the data collected.

3.5.1 The Structure & Composition of ‘MetaFilter’

The name ‘MetaFilter’ is clearly a compound consisting of two distinct, free morphemes which retain their original core semantics (i.e., the definitions of ‘meta’ and ‘filter’, as recognized by the OED). ‘MetaFilter’ is a typical right-headed compound, where the left constituent ‘meta’ modifies ‘filter’.

The name ‘MetaFilter’ is usually represented in CamelCase, whereby the elements of the compound are fused together without spaces, and the initial letters of the individual elements comprising the compounded form are in capitals. This style convention is consistent across the framing of the site, the FAQ, wiki and related pages. MetaFilter participants generally adhere to the convention as well. A participant might leave the ‘F’ of ‘filter’ in lowercase, possibly due to ease of typing, but would not typically insert a space between the compounded terms.

The pronunciation of ‘MetaFilter’ is straightforward and fairly consistent¹⁶, with all major aspects of its constituent parts unaltered by the compounding process. Neither the orthographic representation nor the pronunciation change by the fusion of these two terms. The only noticeable phonetic difference between the compound term ‘MetaFilter’ and the phrase ‘meta filter’ is a stress re-assignment typical of English compounds — a noun phrase would typically carry stress on both elements, but compound nouns tend to shift to single stress on the first element (Ladefoged, 2006, p. 111). These features of ‘MetaFilter’ are important to note; the term is a stable, relatively unambiguous form. It is not until additional morphological processes are applied that the pronunciation ambiguity is introduced. Consequently, ‘MeFi’ and ‘MeFite’ display variation, whereas ‘MetaFilter’ does not.

16. In rare instances, the pronunciation of the V1 in ‘MetaFilter’ as [e] or [eɪ] has been observed (as opposed to the more commonly heard [ɛ] V1 in ‘MetaFilter’).

3.5.2 Orthographic Features

Table 3. Orthographic Features of ‘MeFi’ and ‘MeFite’

Variable:	Most common orthographic representation(s):	Allographic variants:
MeFi	MeFi	Mefi, mefi, #meFi
MeFite(s)	MeFite, MeFites	Mefite, mefite, #meFite

As with the name ‘MetaFilter’, the M-Set is most commonly represented orthographically using CamelCase (mixed-case letters). Lowercase variants are sometimes used, however, with the most common case variant retaining the capital ‘M’, but employing a lowercase ‘f’ (i.e., not in CamelCase). Variants with lowercase ‘m’ but capital ‘F’ are rare.

3.5.3 Grapheme-Phoneme Correspondence (GPC) Rules

As a starting point toward understanding likely and/or possible pronunciation outcomes for the M-Set, it is necessary to look to common spelling-to-sound mappings in English for the letters that comprise the M-Set variables. The consonant letters in ‘MeFi’ and ‘MeFite(s)’ (specifically, <m, f, t, (s)>) are unambiguously pronounced as /m, f, t, s/ in all major dialects of English.¹⁷ The vowel letters, <e, i>, however, may be realized as several different phonemes. There are some dominant patterns among the possible phonetic realizations of <e> and <i>, but even these may show variation by the letter’s position in the word or morpheme, and other internal factors, in addition to variation by dialect.

The main reference text used for exploring the common mappings of the letters <e> and <i> to phonemes in English is Carney’s (1994) *Survey of English Spelling*. While Carney’s book focuses primarily on American and British English, it is by far the most comprehensive reference for understanding the common relationships between letters and sounds in English. The GPC mappings outlined in these following sections provide a baseline for assessing the possible and likely outcomes for the M-Set, according to internal linguistic factors (and prior to assessing the influence of social factors such as community engagement, which are the focus of later sections of this thesis). The example words used in these sections show the letters which are relevant to the feature being discussed in boldface font wherever possible.

3.5.3.1 The Pronunciation of <e> in English

The default pronunciation for the letter <e> is /ɛ/, such as in ‘bed’, ‘credit’, ‘enter’ (Carney, 1994, pp. 318-321). One might therefore conclude that the initial syllables of the M-Set would likely be pronounced as /mɛ(f)-/. However, this GPC mapping (<e> => /ɛ/) has

17. In other word positions, these letters may be realized differently. However, for the positions in which these particular letters appear in ‘MeFi’ and ‘MeFite(s)’, the phonemes corresponding with the letters <m, f, t, (s)> are unambiguously /m, f, t, (s)/.

several rule exceptions, idiosyncrasies, and alternations according to context and dialect.

The first and probably most relevant rule exception is formally written as:

<e> ≡ /i/ | _<C>{<a>, <i>, <o>}#

(Carney, 1994, p. 318)

This rule can be reformulated as: the letter <e> is underlyingly represented as /i/ in an environment where it precedes a consonant letter and either <a, i, o>, such as in the words ‘edema’, ‘Levi’, ‘veto’. Exceptions to this rule occur, mostly in names, such as ‘Greta’ (/e/), ‘Modena’, ‘Helena’ (/ɪ/). While this rule has therefore been described by Carney as a marginal rule, applying consistently 49% of the time in words and 56% in names (1994, p. 318), it is possible to see from this a GPC-based justification of the M-Set variables beginning with the phoneme segments /mi-/ and not the dominant mapping outcome, /mɛ-/.

Other rule exceptions of <e> realized as /i/, concern morphological endings, such as:

<-al> in ‘legal’, ‘penal’, ‘lethal’ (but not ‘medal’, ‘metal’, ‘pedal’),

<-is> in ‘penis’, ‘thesis’ (but not ‘debris’),

<-ine> in ‘feline’ (but not ‘refine’).

(Carney, 1994, p. 319)

Examples with other morphological endings were reported on in Carney’s (1994) reference, but the <-ite> ending, most relevant to this research, was not included in Carney’s list. Regardless, he stated that this rule exception was both not very predictable as well as considered marginal, occurring in words 31% and in names 47% of the time (Carney, 1994, p. 319).

Lastly, a similar rule exception was given, whereby <e> is underlyingly realized as /i/ in the environment of C<i>V, such as in ‘abbreviate’, ‘comedian’, ‘mediocre’, and many other Latinate forms (Carney, 1994, p. 320). Exceptions to this occur in forms such as ‘discretion’, ‘geriatric’, ‘special’. Even so, these marginal rule exceptions provide further linguistic justification for <me> → /mi/ in some cases.

Further idiosyncratic dialect issues occur with the phonetic realization of <e>, especially between American English (AmE) and British English (BrE). Many words are usually pronounced in AmE with a lax vowel, /ɛ/, but are usually pronounced with a tense vowel, /i/ in BrE, such as ‘lever’, ‘methane’, ‘systemic’, and ‘penalize’. Conversely, several words typically display the opposite pattern (/i/ in AmE and /ɛ/ or /e/ in BrE), such as ‘zebra’, ‘premier’, ‘zenith’, and ‘crematorium’ (Carney, 1994, p. 141). Therefore, it is possible to say that the mapping of the letter <e> in ‘MeFi’ and ‘MeFite’ to phonemes might show GPC-based variation for reasons owing to rule exceptions, dialect-based exceptions, or both. Even when the dialect of the speaker is known, it becomes impossible at this point to disambiguate from the pronunciation outcomes alone which factors influenced the choice of variant.

For pronunciations of <e> as the English phoneme /eɪ/, similar patterns of lexical distribution occur between AmE and BrE (Carney, 1994, p. 164). However, AmE and BrE show exceptions where they might differ in words like ‘beta’, ‘devotee’, ‘theta’; AmE will typically have /eɪ/ where BrE typically has /i:/. Many words borrowed from French also show variation between Englishes, typically with AmE /eɪ/ and BrE /ɛ/, such as in ‘crepe’, ‘ballet’, ‘filet’, ‘crochet’, ‘melee’.¹⁸ These examples provide linguistic support for participants who might choose /meɪ-/ pronunciations of ‘MeFi’ or ‘MeFite(s)’, especially in AmE, since BrE speakers show more exceptions resulting in phonemes other than /eɪ/ for the letter <e>.

Despite these many rule exceptions (and others not included here) and dialect differences, the general rule for <e> realized as /ɛ/ is quite productive, with 82% efficiency in words and 76% efficiency in names (Carney, 1994, p. 321). Going by the overall frequency of the realization of the letter <e> in words, we might expect that the first vowels of the M-Set would be pronounced as /ɛ/. However, even if GPC mappings were the only factors (which they are most certainly not), there are enough exceptions to mapping rules alone to warrant pronunciation variation for <e> in the M-Set.

3.5.3.2 *The Pronunciation of <i> in English*

Analyzing the common phonetic realizations of the letter <i> in English words informs predictions about what the popular pronunciations of the final syllables of ‘MeFi’ and ‘MeFite(s)’ may be. Carney’s Survey of English Spelling (1994) directly addresses the occurrence of the letter <i> in word-final position, which is commonly realized as /i/, with examples found in the words ‘bikini’, ‘Fiji’, ‘graffiti’, ‘Israeli’, ‘kiwi’, ‘mini’, ‘Nazi’, ‘potpourri’, ‘salami’, ‘ski’, ‘Tahiti’, ‘taxi’, ‘wiki’. The many exceptions to this general rule, however, result in its classification as marginal, with 57% frequency in words and 86% in names (and strong bias towards words with low token frequency). These exceptions — all which usually take /aɪ/ — include words like ‘alibi’, ‘alkali’, ‘quasi’, ‘rabbi’, ‘Hi-Fi’, ‘Wi-Fi’, ‘fungi’, and anglicized Latin plurals such as ‘alumni’, ‘cacti’ (Carney, 1994, p. 333). Furthermore, word-final <i> displays some variation in many words, e.g., ‘foci’, ‘sci-fi’, ‘stimuli’.

Words ending in <ite> are almost always pronounced /aɪt/ (e.g., ‘white’, ‘despite’, ‘quite’, ‘site’). Additional support for this is found in Carney’s (1994, p. 331) rule for the spelling <igh>≡/aɪ/, e.g., ‘bright’, ‘high’, ‘night’, as many of these forms have colloquial spellings, such as ‘brite’, ‘hi’, ‘nite’, reinforcing the <-ite>≡/aɪt/ mapping. Lastly, a rule stating¹⁹ <i> ≡/aɪ/ | <C>*<e>*# accounts for <-ite> endings implicitly, even if none of the examples Carney (1994, p. 334) provided for it include <t> as the intervening consonant, e.g., ‘alkaline’, ‘bike’, ‘ice’,

18. Many of these words take the typical French word-final stress pattern in US English.

19. Asterisks refer to rule modifications: <C> is not <x>, but may be a cluster containing <l> or <r> so long as they are not <ll> or <rr>; <e> may be elided before initial suffix vowel.

'Nile', 'Viking', and many others. Regardless, some rare exceptions to this rule do occur, e.g., 'elite', 'petite', 'suite', 'favorite', 'granite', 'composite', 'Yosemite'.

Dialect variation between AmE and BrE with respect to /aɪ/ and /ɪ/ or /i/ can complicate the assessment of the phonetic realizations of these forms. For example, some prefixes ending with <i> are typically pronounced with /aɪ/ in AmE, but can be sometimes pronounced with /ɪ/ in BrE, such as 'anti', 'multi', and 'semi'. Conversely, unstressed <-ile> endings commonly show the opposite pattern (/ɪ/ in AmE and /aɪ/ in BrE), such as 'agile', 'fragile', 'futile', 'reptile', 'volatile', 'hostile' (Carney, 1994, p. 151). Furthermore, alternations between AmE /aɪ/ and BrE /i/ may exist in words like 'albino', 'migraine', 'quinine' and its opposing pattern (AmE /i/ and BrE /aɪ/), such as 'philistine', 'bovine', 'serpentine', 'neither'. Alternations are also possible in words like 'vitamin', 'simultaneous', 'endive', where AmE tends to map <i> → /aɪ/ and BrE tends to map <i> → /ɪ/ (Carney, 1994, p. 152). It is unclear whether any words ending in <-ite> are subject to dialect variation, but it seems unlikely, as <-ite> ≡ /aɪ/ as a suffix is so consistent (e.g., the exceptions are few, and are idiosyncratic rather than rule-based) and neither the rule nor exceptions were noted in Carney's survey.

Considering only grapheme-phoneme correspondences in English words (and not other linguistic or sociolinguistic influences), at this point it can be said that ambiguity and phonetic variation in the M-Set terms is highly justified, given the numerous conflicting rule exceptions for both <e> and <i>. The range of possible phonemes for <e> is greater than the range for <i>, but the ambiguity involved in mapping <i> onto phonemes is greater than the ambiguity in mapping <e>. That is, <e> → /ɛ/ with higher lexical incidence than word-final <i> → /i/. This is notable because it has traditionally been the first vowel of the M-Set variables that has been the focus of the debate, and not the second vowel. Review of previous discussions on MetaFilter may also lead readers to assume that there is a substantial (if not majority) number of participants who prefer the [mifaɪ(t)] variants (despite GPC rules that would suggest /ɛ/ would be the dominant choice for this word position). These two points suggest that there is more ambiguity about the first vowel of the M-Set variables than one might assume from looking only at GPC rules. That is to say, there are other factors to consider, many of which introduce additional complications in making a pronunciation choice.

3.5.4 Phonetic Realizations of the M-Set

Moving away from differences in GPC rules and onto matters of differences in production, it is understood that speakers of various English varieties may realize phonemes in unique but broadly identifiable ways. Variation in phonetic output can be described as belonging to one or more of four categories of phonetic differences (Wells, 1982, §1.3; Carney, 1994, pp. 53–59). These categories help explain possible biases, and spelling-related

pronunciation issues in general. The categories involve differences in phonotactic distribution, phonetic realization, phonological system, and lexical distribution. These four categories will be addressed separately, as each pertains to the M-Set in a unique way.

3.5.4.1 Differences in Phonotactic Distribution

Differences in phonotactic distribution concern the likelihood of the presence of phonemes in certain contexts. As covered previously, some of these differences have consequences for pronunciation outcomes when based on orthographic input. For example, regional differences in phonotactic rules dictate that /ɪ/ is not allowed in final, open, unstressed syllables for speakers of Standard American English and some other varieties (Ladefoged, 2006, p. 96). Therefore, the possible options for pronouncing word-final <i> in ‘MeFi’ may be limited to [i] and [aɪ] for these speakers. As such, any geographic differences in distributions of M-Set pronunciations may at least be partly a result of differences in phonotactic distributions.

Using Standard American English phonotactic rules and grapheme-phoneme correspondences as reference, a list of possible pronunciations of ‘MeFi’ is given in **Table 4**. As this is only a starting point, and ‘MeFi’ is the subject of more discussion than ‘MeFite’, only the former variable is tabled for now.

Table 4. Phonotactically Viable Phonetic Realizations of ‘MeFi’

‘MeFi’ Variant	M[i]Fi	M[ɪ]Fi	M[eɪ]Fi	M[ɛ]Fi	M[aɪ]Fi
MeF[i]	[mifi]	[mɪfi]	[meɪfi]	[mɛfi]	[maɪfi]
MeF[ɪ]	[mifɪ]	[mɪfɪ]	[meɪfɪ]	[mɛfɪ]	[maɪfɪ]
MeF[eɪ]	[mifeɪ]	[mɪfeɪ]	[meɪfeɪ]	[mɛfeɪ]	[maɪfeɪ]
MeF[ɛ]	[mifɛ]	[mɪfɛ]	[meɪfɛ]	[mɛfɛ]	[maɪfɛ]
MeF[aɪ]	[mifaɪ]	[mɪfaɪ]	[meɪfaɪ]	[mɛfaɪ]	[maɪfaɪ]

A matrix displays phonotactically viable combinations of first and second vowels for ‘MeFi’, with non-viable outcomes for Standard American English in dark gray. Two problems immediately arise from this process of elimination. First, speakers of other varieties of English may allow /ɪ/ in unstressed, word-final position. For MetaFilter participants who opt to derive their pronunciation of ‘MeFi’ from a strict derivation of the initial syllables of ‘MetaFilter’, a final [ɪ] would be preferable, and may in fact override any phonotactic and/or other grammatical rules dictating otherwise. Secondly, similar exceptions to phonotactic rules may be employed for pronunciations of ‘MeFi’ as [maɪfaɪ]. Preliminary background research on the pronunciation debate reveals that this seems to be the case for a very small proportion of MetaFilter participants. The rationales given for these pronunciation choices

usually concern analogy by rhyming with forms such as ‘HiFi’ and ‘WiFi’, despite the fact that these forms contain an <i> instead of <e> and that there are no attested words in English demonstrating a mapping of <e> → [aɪ].

Therefore, two additions have been added to the possible pronunciation outcomes for ‘MeFi’ (and ‘MeFite’), shown in light gray in Table 4. The [mɛfɪ] variant is included as a possibility, stemming as it does from a strict clipping of ‘MetaFilter’. The [maɪfaɪ] variant is also included as a possibility, arising by analogy from ‘HiFi’ and ‘WiFi’, as claimed by a small number of MeFites in online discussions. There were no substantive reasons to include other phonotactically non-viable variants into the set of possible pronunciation outcomes.

A new table outlining the possible pronunciations of the M-Set has been created. Each variant has been given a number-letter label (i.e., 1a - 3a, 1b - 4b, 3c), which have been consistently used throughout the data collection and analysis processes. The number refers to the first vowel of the M-Set and the letter refers to the second vowel of the set. Therefore, a 1b pronunciation can refer to both [mifaɪ] and [mifart] variants, providing an efficient way to describe both of them as a like grouping, while also distinguishing their first and second vowels from other groupings.

Table 5. M-Set Variants and Their Vowel Codes

1st vowel code (numbers 1-4)	2nd vowel code (letters a-c)	Possible variants of the M-Set	Common eye-dialect spellings used by MeFites
1 = [mi-]	a = [-fi], [-fit]	1a = [mifi], [mifit]	me-fee(t), me-feat
	b = [-faɪ], [-fart]	1b = [mifaɪ], [mifart]	me-fai, me-fye, me-figh(t)
2 = [meɪ-]	a = [-fi], [-fit]	2a = [meɪfi], [meɪfit]	may-fee(t), meh-fee(t)
	b = [-faɪ], [-fart]	2b = [meɪfaɪ], [meɪfart]	may-fai(t), may-fye, meh-fye, may-figh(t), meh-figh(t)
3 = [mɛ-]	a = [-fi], [-fit]	3a = [mɛfi], [mɛfit]	meh-fee(t), meffy, mef-fee(t)
	b = [-faɪ], [-fart]	3b = [mɛfaɪ], [mɛfart]	meh-fai(t), meffai(t), meh-fye, meh-figh(t), mef-figh(t)
	c = [-fɪ], [-fit]	3c = [mɛfɪ], [mɛfit]	meh-fih, meffih, meh-fit, mef-fit
4 = [maɪ-]	b = [-faɪ], [-fart]	4b = [maɪfaɪ], [maɪfart]	my-fai(t), myfy, my-fye, my-figh(t)

It should be noted that the “eye-dialect” spellings are sometimes ambiguous, e.g., ‘meh’ can refer to either a category 2 or category 3 pronunciation. Other features, such as more granular variations in vowel length or quality, will be discussed in the following sections.

3.5.4.2 *Differences in Phonetic Realizations*

Differences in phonetic realizations between speakers are the central component of accents. In writing, these differences are sometimes conveyed by recasting conventional spellings into “eye-dialect” forms. This is often done as a means to overtly describe phonetic particulars for stylistic purposes (e.g., a popular device used by Charles Dickens in his characters’ speech; Dickens, 2004, p. 263).

Differences in phonetic realizations of the M-Set can occur in dialect areas in which there are merged or overlapping lexical sets for the FLEECE, FACE, DRESS, or PRICE vowels. This may result in a difference in distribution of pronunciation choices for the M-Set (in addition to other linguistic and/or social factors which may influence pronunciations) in those areas.

Examples of dialect particulars which may explain outcomes or may bias certain individuals toward particular pronunciations are given below. The dialect particulars presented here are not a comprehensive list, but they do highlight some of the better known differences between dialects which specifically pertain to the M-Set.

The high, front vowel /i/, has high lexical incidence in stressed syllables and is found in both checked and unchecked positions (Wells, 1982, p. 140). In most English varieties, /i/ is realized as [i:]. However, in many dialects of England, /i/ can also be realized as [eɪ ~ ɪ ~ əɪ] (Beal, 2008, p. 130; Clark, 2008, p. 160; Upton, 2008, pp. 271-272). Both Australian and Newfoundland English show variation as well, with some realizations displaying an onglide [eɪ] or being expressed as a diphthong [eɪ] (Horvath, 2008, p. 91; Schneider, 2008, p. 386).

The overall variation in /i/ accounts for a substantial area within the possible vowel space. For the M-Set, the variation in /i/ is complicated, as it is difficult to determine whether surface representations of the vowel are underlyingly /i/, or if speakers are truly realizing <e> as underlyingly /ɛ/ or /eɪ/. Without the ability to measure the vowel spaces of the participants in this study, we can only go so far as to be concerned with surface representations, and to save the perceptual and production issues for potential further research.

The front, open-mid vowel /ɛ/, most often represented by the letter <e> (Wells, 1982, p. 128), is usually expressed as [ɛ], but is often realized in some English varieties as a more raised variant, [e], such as found in Southeast England (e.g., London, Milton-Keynes and Reading) (Altendorf and Watt, 2008, p. 203), Traditional Received Pronunciation (Upton, 2008, p. 241), Australian English (Horvath, 2008, p. 91), and New Zealand English (Bauer and Warren, 2008, p. 41). For some New Zealand English speakers, /ɛ/ overlaps with /i/; /ɛ/, in this dialect, is qualitatively the closest short vowel to /i/ (Bauer and Warren, 2008, p. 47).

Conversely, in Belfast English, /ɛ/ may be lowered to [æ], where words like ‘set’ are realized as [sæt] (Hickey, 2008, p. 93). The same holds true in some North American English

varieties, most notably as part of the Northern Cities and Canadian Shifts, where /ɛ/ is further back and/or lowered (Gordon, 2008, p. 82; Boberg, 2008a, p. 155; Schneider, 2008, p. 385).²⁰

The English phoneme /eɪ/, most often represented with the letter <a>, but sometimes <e>, can appear in both checked and unchecked syllables (Wells, 1982, p. 141). /eɪ/ is pronounced with a diphthong [eɪ] in most English dialects (most notably AmE and RP), but can be realized as monophthongal [e] in Scotland, or [e:] in Ireland, Wales and Northern England (Upton, 2008, p. 274), as well as in Cajun and Jamaican English, Gullah, and other varieties (Schneider, 2008, p. 388).

Diphthong variants of /eɪ/, such as [ɛɪ], are found in Southwest England, [ʌɪ] in Southeast England, [ɛɪ/ʌɪ/æɪ] in West Midlands (Upton, 2008, p. 274), and [æe], [øe] or [øi] in New Zealand English (Bauer and Warren, 2008, p. 41). In the American South, /eɪ/ is highly variable, more so than in any other part of North America (Thomas, 2008, p. 98). In the Low Country (South Carolina, Georgia) and other parts of the American South, /eɪ/ can be heard as lowered, [ɛɪ], or even [æi ~ ɜi]. The latter, more extreme forms usually co-occur among speakers who also have monophthongal /aɪ/ in all contexts, as is often found in southern Appalachia, The Ozarks, Texas, the Piney Woods belt, and parts of North Carolina (Thomas, 2008, p. 98).

The English phoneme /aɪ/ is usually found in stressed syllables, and can be in both checked and unchecked positions (Wells, 1982, p. 149). /aɪ/ is a diphthong in almost all varieties of English²¹, but shows some range as to its start and end points, as well as the general trajectory of change. The basis for comparison is usually recognized as [aɪ], as exemplified by the Standard American English pronunciation (Kretzschmar, 2008, p. 44). In RP and many other British Englishes, /aɪ/ also starts from the central position, [a], but moves toward [ɪ] or [ʌ] (Upton, 2008, p. 274). Variations in starting position are often observed in Australian English, where /aɪ/ is often a “*low central vowel with a closing glide*” (Horvath, 2008, p. 92), represented as [æe] or [ɔe]. Additionally, New Zealand English speakers may start /aɪ/ considerably further back than do speakers of most other varieties (Bauer and Warren, 2008, p. 41).

In Canada, /aɪ/ is one of two vowels involved in Canadian Raising — that is, the pronunciation of /aɪ/ and /aʊ/ with higher nuclei before voiceless consonants. Therefore, pronunciations of ‘MeFi’ with either a V1 or V2 of /aɪ/ might actually use the realizations [ʌɪ ~ ɜɪ ~ øɪ], which could possibly be perceived as more like /ɛ/ or /eɪ/ by some hearers. In such cases, the intention of mapping the <e> onto an /aɪ/ is probably done for other reasons

20. Lowering of /ɛ/ is also noted as an emerging feature of California dialects; this sound change is primarily led by young, urban women. In Canada, lowering of /ɛ/ was also found to be a feature of mostly younger speakers (Gordon, 2008, p. 139).

21. For example, /aɪ/ often displays glide weakening (i.e., monophthongization) in some phonetic contexts in parts of the American South (Thomas, 2008, p. 100).

(e.g., analogy with ‘HiFi’) which happen to defy GPC rules.

As with the /i:/ English phoneme, /ɛ/, /eɪ/ and /aɪ/ display enough range and ambiguity across English varieties to justify the need for further investigation of the perceptual and production biases at play in pronunciation outcomes. This topic is beyond the scope of this investigation, but will be readdressed in the Discussion Chapter. For now, it is sufficient to be aware of these inherent biases, as they may account for distributional differences in the M-Set by geographic region (as a proxy and correlate of the linguistic varieties spoken in that region). What are of utmost importance — as a starting point in this research enquiry — are the surface realizations of the M-Set. That is, whether or not there are statistically significant biases by geographic region, and if those differences can be at least partially accounted for by dialect-based variation.

3.5.4.3 *Differences in Phonological Systems*

The third of the four types of phonetic difference defined by Wells (1982) concerns differences in the phonological systems of the speakers. Non-native English speakers, multilingual speakers, native speakers of different English dialects, and even people style-shifting or adopting a different register, can vary as to the number of phoneme contrasts they exhibit (as the phoneme inventory shifts or changes). This may restrict or expand the options available to them in selecting grapheme-to-phoneme mappings, as compared to others. Experience with other language varieties may also influence a speaker’s tendency to select particular mappings, for various reasons both internally and socially motivated.

From this, it is important to consider to what degree MetaFilter participants’ experiences with other languages are correlated with their pronunciations of the M-Set. The survey methodology employed in this research targets this information and the results in the data analysis identify whether such links exist, and to what degree of statistical significance.

3.5.4.4 *Differences in Lexical Distribution*

The fourth and final category defined by Wells (1982) involves differences of lexical distribution. This refers to the occurrence of phonemes in words and is therefore of great concern to this research. However, recent published work based on speech (and not text or dictionary) corpora which outlines the frequencies and transitional probabilities (e.g., which segments are likely to follow a given segment) of phonemes is unavailable and difficult to produce in any representative way. Therefore, it is more efficient for the purposes of this study to focus on survey responses from MetaFilter participants (see 4.2 **Research Design**, p. 89), in which these participants may reveal the words that are of (phonological) importance to them in making a pronunciation choice (e.g., words that are analogous to their

pronunciation of the M-Set), rather than to refer to a general set of analogous forms derived from spoken corpora.

3.5.4.5 *Summary of Phonetic Realizations of the M-Set*

Wells (1982) outlined four distinct ways to categorize phonetic differences. These four categories have been useful toward further understanding the different ways that pronunciation variation can occur in the M-Set. In the following chapters of this thesis, these categories will be revisited as necessary, so as to give a full account of pronunciation variation of the M-Set on MetaFilter. For now, other features to consider in completing the grammatical sketch of the M-Set will be addressed.

3.5.5 **Stress Assignment and Vowel Length**

The primary stress in ‘MeFi’ and ‘MeFite(s)’ is on the first syllable, which is typical of English nouns and compound nouns (and is the stress pattern of the compositional parts of ‘MetaFilter’, i.e., ‘meta’ and ‘filter’). While second-syllable primary stress is possible — and indeed more likely for speakers of languages where second-syllable stress is fixed (e.g., Polish) or more common (e.g., French) — it is not a notable variation to discuss further, as differences in stress assignment have not historically been a part of the debate over the pronunciation of the M-Set on MetaFilter; the reassignment of stress is overwhelmingly recognized by MetaFilter participants as influences from those other languages, and not arising from ambiguities in English pronunciation.

That said, more subtle differences between stress patterns across English varieties could potentially alter the vowel quality of pronunciations. For example, AmE speakers tend to preserve secondary stress more than do BrE speakers, resulting in more fully-realized vowels in syllables with secondary stress for Americans (Kretzschmar, 2008, p. 49). In AmE, one of the more marked features of Southern varieties is the tendency toward strong initial syllable stress, which results in longer and more fully-realized vowels in the first syllable than in subsequent ones. This could result in AmE pronunciations of the M-Set with longer and tenser vowels, and British varieties with comparatively shorter and laxer realizations. These subtle differences in stress patterns are not the sole cause of categorical differences between variants, but it should be noted that the effects of increased or decreased syllable stress may result in more or less ambiguity in the perception of variants as belonging to a particular pronunciation category.

Differences in vowel length are also possible, but are not a notable part of this pronunciation debate. Some dialects of English have grammatical rules (e.g., the Scottish

Vowel Length Rule²²) which could result in differences in the length of the vowels in the M-Set. These differences, however, would not typically result in differences in vowel quality, nor would they be phonemically contrastive. Preliminary research into online discussions about the M-Set show that MetaFilter participants generally do not recognize these subtleties as meaningful differences between pronunciation outcomes.

3.5.6 Syllabification

Issues of syllabification have an impact on how to predict and explain pronunciation outcomes of the M-Set. Linguists do not always agree on theories of syllabification, especially concerning intervocalic consonants following stressed vowels (Carney, 1994, p. 77). This makes assessing the likely syllabification of ‘MeFi’ and ‘MeFite’ difficult. Three possibilities arise, resulting in different outcomes for the vowel quality of <e>:

1. Syllabification is CV.CV — employing the Maximal Onset Principle (MOP; Selkirk, 1981), whereby a consonant attaches to the following syllable wherever possible. This results in an unchecked V1 in ‘MeFi’.
2. Syllabification is CVC.V — Wells’ (1990) view, whereby consonants attach to the stressed syllable (and to the leftward one in the case of equally stressed syllables). This results in a checked V1 in ‘MeFi’.
3. Syllabification is CVC.CV — the ambisyllabic view (Kahn, 1976), whereby the consonant is shared by both syllables, even if it is not pronounced as geminate. The results in a checked V1 in ‘MeFi’.

‘MeFi’ (and ‘MeFite’) is a disyllabic word, with primary stress on its first syllable.

Assuming vowel selection occurs first and then words are syllabified accordingly, pronunciations with [i] or the diphthong [eɪ] as the first vowel can be assumed to undergo syllabification in an uncomplicated manner. The MOP can be easily met and no phonotactic constraints are violated. For these pronunciations, syllabification is most likely CV.CV.

However, where the lax vowel [ɛ] appears, the MOP is met, but the phonotactic constraint stipulating “no lax vowels in open, stressed syllables” is violated (Ladefoged, 2006, p. 96). This is unresolvable, unless ambisyllabicity is accepted, and the second consonant of ‘MeFi’ is shared across both syllables (but is not necessarily pronounced as a geminate). The result is CVC.CV, as in [mɛf.f(V)]. Furthermore, speakers with phonetic outputs of [mɛf.fi] or [mɛf.fɑɪ] are likely prioritizing this phonotactic constraint above the visual cue that is created by CamelCase, which would suggest a syllable break.

22. SVLR, or “Aitken’s Law” stipulates that vowels are lengthened before voiced fricatives, before /r/, and before a syllable or word boundary (Aitken, 1984; Watt and Ingham, 2000, p. 207; Stuart-Smith, 2008, p. 58). For example, if ‘MeFi’ were syllabified as CV.CV by a Scottish English (ScE) speaker, both vowels could be potentially realized as phonetically long. However, the SVLR rule may not even apply to both vowels in ‘MeFi’ if the ScE speaker perceived the syllabification to be CVC.V. In any case, this would likely not prompt noticeable differences in vowel qualities for the segments involved in the M-Set variables.

However, the possibilities change if it is assumed that syllabification takes precedence and motivates the vowel quality. If this is the case, the V1 in the first syllabification possibility (i.e., CV.CV) would be unchecked and therefore subject to phonological rules which constrain how that vowel would be pronounced in some dialects. For example, in Standard American English, lax vowels are not allowed in stressed, unchecked positions (Ladefoged, 2006, p. 96). Therefore, a [mɛ-] pronunciation would not be a possible outcome for those prioritizing this syllabification rule. These speakers might additionally choose this syllabification based upon the visual cue provided by the CamelCase.

Syllabifying first also precludes the possibility of a tense vowel in ‘MeFi’ if Wells’ (1990) or Kahn’s (1976) syllabification principles are adopted. Both of their views result in a checked first vowel for ‘MeFi’, and therefore a lax vowel would be expected.

The scope of this study does not allow for a deeper exploration of syllabification, but this is an important topic for future study, as it can lend new insights to theories on ambisyllabicity and priming of stressed syllable-based phonetic outputs. What can be said at this point is that the existence of variation is evidence for either competing theories of syllabification being adopted by different speakers, different processing order by different speakers (vowel choice first, or syllabification first), or both. For now, it is sufficient to be aware of these possible internal influences on the pronunciation of the M-Set and to be open to any evidence of speaker awareness of these issues during the data analysis process.

3.5.7 Morphological Processes

Another area that warrants investigation, but owing to space limitations cannot be covered here, is the abbreviation classification for ‘MeFi’, including analysis of the different possible morphological routes that can be taken to derive ‘MeFi’ from ‘MetaFilter’. Understanding this process can help provide additional clues towards pronunciation. For example, ‘MeFi’ can aptly be categorized as a syllabic clipping (i.e., abbreviations which take syllables of words as units, e.g., ‘flu’ from ‘influenza’), an acronymic clipping (i.e., abbreviations which take the initial letters of words as units, e.g., ‘scuba’), or both (Bauer, 2008, p. 197).

Some speakers may tend to realize the form as more like a syllabic clipping, and reanalyze the constituent parts as a new whole according to a set of rules which apply to syllabic clippings, whereas others might realize the form as more analogous to an acronym, and treat it as conforming to a different set of grammatical rules (e.g., reanalysis in order to fit English phonotactics). These varying perspectives can result in different pronunciations. As with syllabic theories, this is worth keeping in mind during data analysis, for reasons pertaining to this study as well as to address issues concerning current theories about

morphological clipping processes (Gries, 2004; Bauer, 2008), the taxonomy of abbreviations (Cannon, 1987; Bauer, 1998), and non-standard words (Sproat et. al., 2001).

3.6 Summary of MetaFilter and the M-Set

While it remains to be seen which pronunciation variants are actually preferred by MetaFilter members, it has been possible thus far to identify some linguistic constraints which might bias speakers toward particular variants. The two most influential factors revealed in this chapter have concerned the mapping of orthographic segments onto phonemes and the phonetic realizations of phonemes. Both of these factors display variability, and both influence each other to some degree. Additionally, an overview of MetaFilter, its subsites, and its userbase provided sociohistorical context for understanding the community and its participants.

The methodology proposed in the next chapter will outline a way to address the three research questions stated at the beginning of this thesis (see 1.2 **Research Questions**, p. 5), such as which sociolinguistic factors may be involved in the pronunciation M-Set. The methodology has been designed to also identify further influences on the pronunciation of the M-Set, including the ways in which these variables have come to be enregistered with the indexical values that they have been imbued with by the participants who use and recognize the variables.

Chapter 4: Methodology

4.1 Introduction to Methodology

The methods used in this research were selected to give a comprehensive account of the distribution and enregisterment of the M-Set variables. Data were collected through research methods involving community participation and large-scale surveys. Quantitative and qualitative methods were employed to analyze these data.

Methodology will be discussed in this chapter as follows: the **Research Design** section provides an explanation of the research approach taken and the types of data collected. The **Sampling Design** will describe the population of participants included in this study. Next, the **Measures** used to track the enregisterment of the M-Set will be explained, followed by the **Data Collection Procedures**. Lastly, the **Data Analysis Procedures** will be outlined.

4.2 Research Design

The mixed methods research design undertaken for this thesis work involved eight years of regular, daily participation in the MetaFilter community, large-scale online surveys conducted at two points in time (2010 and 2012), and analysis of online discussions occurring among MeFites. The purpose of this design was to collect multiple types of data from several related sources, which would reveal the distribution of M-Set pronunciations preferred by a representative sample of MeFites, including their attitudes about those variables over time. This allowed the process of enregisterment for the M-Set variables to be observed and tracked.

4.2.1 The Research Approach

Variationist methods in sociolinguistic research aim to show the distribution of a variable under investigation, and how its variants are used in different contexts, by speakers who can be located by their position within a social space. Understanding this social space — that is, the communicative context in which the variable occurs — is fundamental to explaining the linguistic variation that arises there (Labov, 1963, p. 275). Therefore, this approach requires that the researcher invest a sufficient amount of time and effort into being a socially-engaged participant within their chosen community under investigation. This allows the researcher to observe the community from an emic perspective, with a true understanding of the social organization, activities, resources, and practices that define it (Duranti, 1997, p. 85; Levon, 2013).

One drawback of this approach is that it can sometimes result in work that is highly subjective (Blommaert, 2007, p. 684). However, the addition of quantitative methods can support potentially subjective interpretations with more concrete, objective results. This is one reason that the research undertaken here is a mixed-methods design, allowing the variables to

be interpreted as part of a larger social system, with conclusions about their meaning and use in the practice empirically supported by statistical outcomes of quantitative data analysis.

Starting with Labov's 1963 study of variation on Martha's Vineyard, sociolinguistic research involving both qualitative and quantitative methods have been combined in many sociolinguistic studies to date; see Eckert's (1989) Detroit high school investigation or Mendoza-Denton's (2008) study of Latina Youth in California for other successful exemplars of this particular type of mixed-methods research. These research designs provided a comprehensive picture of a community with respect to specific linguistic behavior. The approach was required to contextualize the sociolinguistic constraint system that the quantitative methods revealed, thereby allowing researchers to view the results as reflective of a linguistic situation that is beyond the concept of 'place' to one of membership to a shared social space that is defined and constructed by the participants themselves (Llamas, 2007, p. 582). A mixed-methods approach is even more necessary in online environments (Warnick, 2010, p. 45), where 'place' begins and ends with the shared social space and the texts that result need to be connected to practices, so that language choices can be understood from the user's perspective (Barton and Lee, 2013, p. 167).

Mixed-methods research is especially relevant to the study of enregisterment. Immersion in a community alone cannot fully explain the spread of innovation or the indexical relationships between variables and social factors within a community. Some quantitative methods need to be employed, so that linguistic behavior may be observed, quantified and correlated with other behavior (linguistic or otherwise) in ways that empirically demonstrate a linguistic progression (i.e., the steps in the enregisterment process). For example, a mixed-methods design can show quantification and rationale for how highly-participatory subgroups within a community, such as super-users or long-time members, are more actively influencing the process of enregisterment than 'lurkers' or otherwise less participatory members. Equally important to the process of enregisterment, participating in the M-Set debate is one way that the less participatory MeFites may move from the edge of a community to a more core position within a practice, such as when 'lurkers' shift to public stance-takers in what Lave and Wenger (1991) consider "*legitimate peripheral participation*" (Wenger, 1998, pp. 100–101). Accounting for the range of participation levels — from the peripheral participants to the super-users, using mixed-methods research, gives a description of the entire system (i.e., community), rather than focusing on one perspective, trajectory or sub-practice involved in the enregisterment process, and doing so from a purely descriptive, qualitative perspective.

4.2.2 The Researcher's Perspective

The variables chosen for investigation and the mixed-methods employed to study them

were borne out of the variationist approach, and have borrowed from traditional ethnographic methods where applicable and appropriate. I have been a participant-observer of the MetaFilter community since 2006. Over this time, I have posted and commented with regular and consistent frequency, which has afforded me status as an active (or possibly core) participant in the community. One important aspect of this — for both personal and academic interest — has been to regularly plan and attend meetups at every possible opportunity (e.g., when travelling to a new city). As a result, I have made several long-term friendships across the world and have gained valuable insights about the MetaFilter community and what it means to be a part of it. These personal interactions were crucial to the research undertaken and meaningfully informed the interpretation of the data collected (Llamas, 2007, p. 602; Barton and Lee, 2013, p. 176; Stanford, 2013). This extended to every step in the research process. For example, experience-based insider knowledge about the community allowed me to make reasoned decisions about which measures were likely to correlate with increased use of variants, which informed the survey design and its implementation.

4.2.3 The Researcher's Influence on Enregisterment

As Squires (2010) and Johnstone (2009) have pointed out previously, linguists implicitly take part in the process of enregisterment through identifying features and linking them to speech varieties and registers, as well as by raising linguistic awareness through publications about those features and varieties, which may be read and distributed among other academics or even participants in the communities which use the features or varieties. This can have both positive and negative results for the enregisterment of the features or variety, as well as for speakers (even for speakers who do not use the features or speak the variety being enregistered). In the case of MetaFilter, the research presented here fundamentally alters the enregisterment of the M-Set in several crucial ways. First, the survey methodology reveals information about the distribution of the M-Set which was not easily observable or quantifiable previously, such as which variants were the most or the least preferred, as well as the range of variation for the M-Set variables. Secondly, the research draws focused attention to the debate about the M-Set and the influences on pronunciation. This raises metalinguistic awareness about the M-Set, MeFites, and sociolinguistic phenomena in general, which in turn may satisfy the curiosity of some but may cause others to tire of the subject. Lastly, this research draws attention toward specific members of the community (myself included) through examples of commentary or references to social categories (and their relationship to the M-Set). This focus unintentionally and unavoidably creates an imbalance by making individuals or groups salient for the purpose of highlighting their influence on the enregisterment of the M-Set. The eventual outcomes of this are not clear,

but equal treatment and representation of all members of MetaFilter has been an overarching value of the community since its inception, and so it is conceivable that some participants may feel that the selection of individuals or groups to exemplify linguistic behavior may flaunt or flout those values in some way. It is not the intention to bias the research in this way, but rather to demonstrate an awareness of my role and unavoidable influence as a linguist, MeFite, and researcher on the enregisterment of forms.

It is my hope therefore, that these potential drawbacks are counterbalanced by the positive impact that this research may have in the field of sociolinguistics, CMC studies, and the MetaFilter community.

4.2.4 Overview of the Types of Data Collected

Data collected and analyzed for this thesis consisted of four main types, each from the MetaFilter community, collected using different means. Data consisted of online surveys, participation and usage meta-data, corpora, and qualitative data from community discourse.

4.2.5 Online Surveys

The primary source of data for this thesis come from two online surveys, conducted for five days each at two different points in time (two-and-a-half years separated the 2010 survey from the 2012 survey). The implementation of the surveys and their content are discussed more fully in **4.5 Data Collection Procedures, p. 100**.

4.2.5.1 Rationale for Survey Data Collection

Surveys are especially well-suited for collecting data in a text-based community. Surveys are often traditionally administered through a printed medium, allowing participants to choose from options or describe their choices in text. In recent years, many of these surveys have been conducted online, with recruitment for participants employed through listservs, email or other signage. For participants who are used to text-based interactions and participate in CMC environments with any regularity, the online survey may be a natural context that they are familiar with, even if they have never participated in a survey prior.²³

There are many advantages to online surveys. Researchers may be capable of collecting a large amount of data in a relatively short space of time, from participants who are physically remote (Boberg, 2013, p. 3117). Large data samples allow for rigorous statistical analysis, leading to robust conclusions about social patterns in the data, which can then be generalized to the population as a whole (Boberg, 2013, p. 3151).

Additionally, online surveys have ‘inter-participant comparability’, meaning that

23. Many site registrations or online stores have similar, paginated survey-like formats that internet users may be experienced with navigating.

participants are responding to the same stimuli and therefore differences in responses cannot be due to the data collection techniques (Boberg, 2013, p. 3161).

Ease of analysis is another advantage to this method. Data are collected digitally and can be quickly and efficiently converted to other formats or data types for tabulation, statistical analyses and visualizations (Boberg, 2013, p. 3163).

One last notable advantage of online surveys — specific to this particular study — is that they can allow participants to select preferred pronunciations from an array of choices. While this may disadvantageous in studies where the researcher wishes their participants to be unaware of the variables being targeted for investigation, for the linguistic situation that occurs with the M-Set, direct enquiries do not pose a research problem. In fact, MeFites frequently engage in metalinguistic discussions as an enjoyable part of their practice — the M-Set has been central to many of those discussions. As such, the appearance of a survey focused on participants' opinions about MetaFilter-related issues is seen by many as an opportunity to contribute to a fun discussion, rather than a source of linguistic insecurity or researcher scrutiny. A direct elicitation of pronunciation preference is also fitting with the ethos of MetaFilter, where being passionately interested in minutiae and obscure intellectual pursuits is generally viewed favorably, to the extent that the phrase “overthinking a plate of beans” has been enregistered within the community to express the very idea of meta-analyses of this ilk (see 3.4.1 **Elements of the MetaFilter Register**, p. 71).

It would be remiss to not discuss some drawbacks to online surveys as well. The biggest disadvantage being that there is no easy way to verify whether responses were selected at random, were errors, or were unintentionally left blank. Additionally, the likelihood of the participant experiencing other types of technical problems is increased with online surveys, where computer setups and internet access differ among participants.²⁴

Survey fatigue/boredom is another potential hurdle when administering questionnaires to online participants. I intentionally designed the survey to be as brief as possible, so that participants would be more likely to complete it and answer the few fill-in questions at greater length. While this turned out to be a fruitful strategy, it unfortunately limited the number of questions I could ask of the participants and later analysis revealed some minor gaps in the data collection.²⁵ Fortunately, interest in the study (and especially the topic of the study) was high and the majority of participants responded to all of the questions asked. This created a robust data set to work with.

24. I was alerted to a few technical problems during the 2010 survey. These caused minimal loss of data and all issues were able to be quickly resolved. These issues were prevented in the 2012 survey.

25. For example, in hindsight I realize that this research could have benefitted from knowing how participants access MetaFilter (e.g., laptop, mobile phone, tablet), resulting in an understanding of how different types of access may correlate with increased engagement and therefore pronunciation preferences.

Online surveys are also limited by their inability to inform the researcher about the frequency and acoustic detail of actual spoken use of the variables. As such, this research does not intend to make generalizations about how frequently forms are actually spoken aloud (or about fine-grained phonetic detail), beyond what has been observed through participatory research and what participants reported to me about their frequency of use (which was not directly elicited in the surveys). What has been of greater interest and reliability is which variants participants claimed they do (or would) use, and how strongly they felt they (would) use those variants exclusively (as opposed to using other variants). This provided a reliable distribution of variant selection, but not frequency or fine-grained phonetic detail beyond categorical vowel differences.

In this way, these surveys were not intended to be a substitute for natural speech data, but rather an insight to what occurred (or what may occur) in the minds of MeFites in an oral speech environment. Fortunately, the differences between variants are perceived by MeFites as categorical differences in vowel quality and can therefore easily be elicited in a survey by providing audio samples for the participants to choose from.

In summary, administering brief online surveys to the MetaFilter community proved to be an extremely advantageous way to collect large amounts of data quickly and efficiently. The ‘testing’ environment was familiar and natural to MeFites — the surveys were administered on MetaFilter, at the convenience of those who wished to participate. Interest in the surveys and thoroughness in completing them were both high, resulting in a robust data set, output as a spreadsheet and ready for data normalization, filtering and analysis.

4.2.6 Participation and Usage Meta-Data

Another advantage of the online medium in CMC studies is the potential availability of public profile and social network data. These data can be combined with survey results or qualitative data to provide a more rounded picture of both the overall social network and the social positions of individual users. Meta-data often include statistics about participants’ involvement, such as the number of contributions they have made to the site, their number of contacts (and the number of people who consider them a contact), and their length of time on the site (based on their user join date), to list a few metrics.

The MetaFilter InfoDump is one such publicly available collection of social network data. Available for free download and located at: <http://stuff.metafilter.com/infodump/>, the InfoDump is a regularly updated collection of site data culled from the MetaFilter database. The InfoDump provides information to be combined with the survey results and includes (but is not limited to) statistics about the number of posts, comments, favorites and contacts

of every user on the site, sortable by timestamps and other metrics. The addition of these meta-data to the survey data allowed for the direct elicitation of pronunciation preferences and rationales for those choices to be contextualized in how participants engaged with the community, and how that engagement was viewed by other participants.

4.2.7 Corpora — Word Frequency Tables

The MetaFilter Corpus (Millard, 2011) is a project that was launched in January, 2011, consisting of downloadable lists of word-frequency tables based on all MeFites' posts and comments from the four main MetaFilter subsites (MeFi, AskMe, MeTa, Music). The lists are divided into daily, monthly, yearly or all-time formats, for the entirety of MetaFilter's 14-year history. Like the [InfoDump](#), this data set is also publicly available for free download (located at <http://stuff.metafilter.com/corpus/freq/>).

All word frequency tables consist of an information header and columns for the raw count, parts-per-million (PPM), and the word. Data rows are sorted in descending order, from most frequent to least frequent. An example of the first ten data rows of a typical word frequency table is shown in [Figure 12](#).

Figure 12. Image of Word Frequency Table Format

count	PPM	word
2311067	44820.6837715875	the
1421986	27577.904419744	to
1324562	25688.4696713083	a
1166008	22613.4836606386	and
1089998	21139.350641787	of
960953	18636.6602666034	i
781047	15147.5749503355	that
733580	14227.0030255121	in
724148	14044.0794281721	you
705390	13680.2879906295	is

4.2.8 Qualitative Data from Community Discourse – Posts and Comments

The fourth and final component of the collected data was in the form of public posts and comments made by MeFites on MetaFilter. Six MetaTalk threads were selected; the topic of each was the M-Set or linguistic norms and conventions on MetaFilter. These public discussion threads comprised over 1,250 comments made to the MetaFilter community about the M-Set or the enregisterment of forms, over an 11-year span.

4.2.8.1 *Rationale for Qualitative Data Collection*

While some sociolinguistic studies have utilized CMC sources such as Twitter (Russ, 2012) or community forums to support the enregisterment of variables within speech (Johnstone, 2010b), few studies have focused solely on tracking the enregisterment of spoken forms originating from or used primarily within the online medium itself (Squires, 2010). This is an understudied area of enregisterment, which is becoming increasingly useful to sociolinguistic enquiry, as many new words are entering the everyday speech of individuals through CMC (Crystal, 2011).

Within the MetaFilter community, the M-Set has semantic and communicative value — as evidenced by the content of the discussion surrounding them. The stances that were expressed by participants in discussions about the M-Set provided a way to interpret the linguistic variation that occurs in the MetaFilter community and insight into the motivations for those differences (Llamas, 2007, p. 581). As with all stance-taking, these metalinguistic discourses were situated within the larger context of the community, socially constructed and reflective of broader ideologies about language in general and group identity (Barton and Lee, 2013, pp. 108–109, p. 123).

Metalinguistic discourse about the M-Set is visible and trackable through MetaFilter's archive of community discussions. The coding of features and stances in these comments (including favorite counts) allowed the qualitative data to be transformed into quantitative results. The frequency of particular stances over time — and MeFites' orientations to those stances through their responses and favoriting behaviors — were indicative of the spread of awareness about emerging standards, common perceptions, and the continued enregisterment of forms.

4.3 Sampling Design

All participants included in this research were members of the MetaFilter community (MeFites) at the time of each data collection. The surveys were open to any MeFite who was logged into MetaFilter and claimed that they were at least 18 years of age. Prior to the first survey in 2010, it was estimated that at least 2% of the active MetaFilter population would participate, and therefore comprise a suitable sample size for analysis. The response rates far exceeded this estimate and the resulting sample size was more than sufficient to allow for results to be generalized over the entire active MetaFilter population.

The number of MetaFilter respondents for the 2010 survey was 2,521, representing 5% of the registered userbase and 15% of the active userbase for that year.²⁶ These results included surveys from the four full-time MetaFilter site moderators (mathowie, jessamyn, cortex and

26. See 3.2.3 The MetaFilter Userbase, p. 61, for an overview of MetaFilter's userbase statistics.

pb), and the temporary nighttime moderator (vacapinta).

The number of MetaFilter respondents for the 2012 survey was 1,957, representing 3% of the registered userbase and 12% of the active userbase in 2012. These surveys again included the official site moderators (increased to six since 2010: mathowie, jessamyn, cortex, pb, restless_nomad and taz).

The number of MeFites who participated in both surveys was 769, comprising 39% of the 2012 survey data. These participants were labeled the 'Panel Data' and were treated as a distinct, analyzable subset of the overall Survey Data; the Survey Data was treated as trend data, as it mostly comprised different participants over time (even though some participants took both the 2010 and the 2012 survey). The Panel Data participants represented 1% of the registered userbase and 5% of the active userbase in 2012. It should be noted that the group of participants who took both surveys is not a sufficient sample size to be generalized to the entire MetaFilter population (i.e., the registered userbase). Additionally, all MeFites who took both surveys were active users in both 2010 and 2012, which is qualitatively different from the majority of the registered userbase — 73% of the registered userbase in 2012 were not active users (i.e., did not post or comment on MetaFilter in the year prior; it is possible and likely that a majority of these MeFites were 'lurkers' or other non-engaged peripheral participants, but it is impossible to quantify how many of these registered users read the site without posting or commenting).

However, the social patterns in the Panel Data participants could be generalized to the active userbase, as the sample size was more than sufficient (5% of the active population) and the participants were representative of the population of active users in terms of their contribution frequency. The subset of users who took both surveys and selected pronunciation choices in each (N=753) allowed for M-Set preferences and attitudes (stances) participants have about them to be tracked over time.

4.4 Measures

The first step to understanding the process of enregistration for the M-Set on MetaFilter is to assess what the actual distribution of variants is and determine which social factors are correlated with those variants. Conversations with MeFites at meetups, MetaTalk discussion threads, and exploratory research of the data collected from the surveys revealed several relevant social factors to include in the statistical analyses that tested for correlations between variables. These social factors were grouped into four categories: demographic factors, social engagement factors, metalinguistic awareness factors, and M-Set stances. Each of these categories will be discussed in detail, after describing the assessment of the M-Set.

4.4.1 Measuring the Distribution of the M-Set

Assessing the distribution and frequency of the M-Set variables (i.e., ‘MeFi’ and ‘MeFite’) were the primary measures in this research. Based on survey data, the following descriptive statistics were able to be assessed:

- The distribution of each variant of ‘MeFi’ and ‘MeFite’, in 2010 and in 2012.
- The amount and direction of change in pronunciation preference for ‘MeFi’ and ‘MeFite’ variants from 2010 to 2012; all survey participants (trend report).
- The amount and direction of change in pronunciation preference for ‘MeFi’ and ‘MeFite’ variants from 2010 to 2012; only participants who took both surveys (panel report).

From these basic reports, further work could be done to determine whether there were any correlations with social factors that might significantly affect the distributions of pronunciation preferences.

4.4.2 Demographic Measures

The demographic factors included in analysis were linguistic background, geography (i.e., current country of residence), age, and gender. The linguistic background measures included data from surveys enquiring about the participant’s native language and proficiency levels (i.e., fluent, advanced, intermediate, or beginner) of languages studied or learned other than the participant’s native language. Geographic measures involved participant’s self-reported country of residence at the time of the survey. Age and gender were also self-reported by participants.

4.4.3 Measures of Social Engagement

Social engagement factors concerned the frequency and type of participation MeFites’ engaged in on MetaFilter and doing MetaFilter-related activities. Measures of social engagement considered for statistical analyses included the following:

- The year the participant joined MetaFilter (as an indicator of the length of time one has been involved with the site).
- The participant’s self-reported frequency of visitation to each of the three main subsites (MetaFilter, AskMetaFilter, and MetaTalk).
- The participant’s self-reported frequency of listening to the MetaFilter Podcast.
- The participant’s self-reported frequency of attendance at MetaFilter meetups.

4.4.4 Measures of Metalinguistic Awareness

Metalinguistic awareness was directly measured through two means:

- A Likert scale of how strongly the participant felt they do (or would) use their chosen variant of the M-Set exclusively (as opposed to using other variants).
- How much thought the participant had given to the pronunciation of the M-Set prior to the survey.²⁷

4.4.5 M-Set Stance Measures

M-Set stances included measures relating to attitudes about and preferences for the M-Set. Measures considered for statistical analyses were drawn from written rationales in surveys — paragraph-style responses from participants sharing their thoughts and opinions about the M-Set variables. These qualitative data responses were filtered and coded. Analysis revealed approximately 20 categories worthwhile to include as measures possibly correlating with pronunciation variants. These included, but were not limited to, the following measures:

- Whether or not the participant stated that their pronunciation choice was influenced by another person (e.g., a friend, family member, moderator on the podcast, a MeFite at a meetup).
- Whether or not the participant expressed a linguistic basis for their pronunciation choice (e.g., grammatical rules).
- The use of orthographically or phonologically similar forms as a justification for their pronunciation choice (e.g., “*Me’ looks like the word ‘me’, not like an abbreviation for ‘meta’*” or “*I pronounce it like ‘HiFi’ or ‘WiFi’*”).
- The visual appearance of the M-Set (e.g., “*The capitalization of the “m” and “f” makes me read it as “mee-fie” in my head for some reason.*”).
- Expressed avoidance of the form (e.g., “*I rarely use the word MeFi because it sounds awkward to me no matter how you pronounce it*”).
- Prescriptivist stances (e.g., “*Because it’s so obviously the “correct” way to pronounce it that I’ve never felt the need for a “reason?”*”).
- Positive associations with a particular variant (e.g., “*I think it sounds friendly and nontechnical.*”).
- Expressions of indifference toward the M-Set (e.g., “*don’t care*”).
- Expressions of a lack of explanation for their pronunciation choice (e.g., “*I don’t know. That’s just how it sounds in my head.*”).

Several categories displayed overlap, or provided additional data for other measures (e.g., some of these categories are reflective of metalinguistic awareness (or lack thereof)). These categories were included, merged or subdivided in the results where it made sense to do so. For example, the category designating the use of orthographically or phonologically similar

27. Previous research revealed that many MeFites were aware of the debate on MetaFilter and/or were aware of pronunciation ambiguity in the M-Set and had given the matter some thought. However, until the survey(s), it was not known what percentage of MeFites had thought about this topic.

forms was divided into several subcategories for words that MeFites frequently chose to explain their pronunciation choices (e.g., ‘me’, ‘meh’, ‘meta’, ‘HiFi’, ‘WiFi’, ‘filter’).

4.4.6 Other Measures

Word frequency tables provided data concerning the frequency of written forms of the M-Set on MetaFilter by all MeFites. This allowed the enregisterment of the M-Set to be tracked over time, from the first appearance of ‘MeFi’ in 2000, to its status today as part of the most frequent (e.g., in the top 1,000) words on MetaFilter.

Lastly, some quantitative measures were employed in the analysis of the MetaTalk threads. Comments in each thread were coded for the following features:

- Whether the comment contained a stance about the M-Set (or a particular variant within the M-Set).
- Whether the comment contained a preference for or against a particular variant — including which variant(s) the comment referenced and how the variant(s) was (were) evaluated (i.e., positively, neutrally, or negatively).
- Whether the comment contained a prescriptivist stance about language.
- The number of favorites the comment had received.
- The username of the commenter.

4.4.7 Summary of Measures

The measures used in this research provided a comprehensive overview of the distribution of the M-Set, the attitudes about the M-Set and several other relevant features related to linguistic awareness, identity, and participation in the MetaFilter community. These measures were assessed in private domains (surveys and other collected data) and public domains (MetaTalk threads). Lastly, measures taken at different points in time allowed for observations relating to language change and the spread of linguistic innovation.

4.5 Data Collection Procedures

The four types of data involved in this research required different procedures for collection; these are outlined in the following pages.

The moderation staff of MetaFilter were contacted prior to each data collection and the study designs were reviewed and approved prior to their implementation.

A database was created to contain all data that were collected over the course of the thesis research. This was hosted by the University of York and accessed through a secure VPN client. The University uses the Oracle SQL 11g framework and the database was allotted 20GB of storage space on their servers.

4.5.1 The MetaFilter Surveys

The first survey on MetaFilter was conducted as part of a Masters Thesis at San Francisco State University in 2010. The 2012 survey was similar in procedure and format, with various improvements made as a result of feedback from participants in the 2010 survey. The research designs and data collection were approved by university ethical boards; IRB (San Francisco State University in 2010) and HSSEC (University of York in 2012).

Both surveys were advertised on MetaFilter by a small banner of text at the top of each MetaFilter subsite's front page. The banners were visible to all logged in MetaFilter participants once the survey opened and were removed five days later, when the survey closed. These banners consisted of a short, clickable prompt that read "*Have a few minutes? Please participate in a MetaFilter Survey!*" The words 'MetaFilter Survey' were a hyperlink, leading to the Information & Implied Consent page of the survey.

Additionally, many MeFites were made aware of the survey by visiting MetaTalk, where a discussion post about the survey was made once the survey went live. This discussion post stayed open for 30 days, as is customary for all posts on MetaTalk.

The MetaTalk discussion post encouraged participants to take the survey before reading the remainder of the post or participating in it. This was to prevent participants from submitting responses that might have been influenced by discussion about the M-Set in the post. The MetaTalk post content for both surveys is included in **Appendix B**.

The surveys began with the Information & Implied Consent page (see **Appendix C**). Before continuing onto the survey, MeFites were required to click three check-boxes at the bottom of the Information & Implied Consent, stating that they had read the consent information and agreed to the terms, that they were over 18 years of age and participating through their primary MetaFilter account, and that they were the person associated with the account they were currently logged into and using to take the survey.

The survey consisted of one web page, with a summary of the information from the Information and Implied Consent page in a paragraph at the top. This information was followed by brief instructions for taking the survey, including notification that all information was confidential and would not be shared with anyone other than the researcher. See **Appendix D** for the 2010 survey page screenshots.

All questions in both surveys were optional. A small percentage (less than 1%) of participants submitted entirely blank or nearly-blank surveys. Their results were included in the data set and cells were coded as "answer left blank" where appropriate.

The first survey was made available for 5 days to all logged-in MetaFilter members from March 24–28th, 2010. The 2010 survey consisted of 18 questions in total: 16 were multiple

choice (some with extended answer fill-in options) and two were free-form fill-in questions. Questions ranged from site participation behaviors, thoughts about the pronunciation of the M-Set, and demographic background information.

The second survey was available from August 22–26th, 2012. This survey consisted of 20 questions in total, as well as a check-box option to generate a word-frequency table of the participant’s posting history, mailed to the participant and/or the researcher. The 2012 survey is provided in **Appendix D**.

Differences between the 2010 and the 2012 survey formats will be discussed in the next section, after a review of the survey content common to both.

The first survey question asked the participant about the frequency with which they visited each of the MetaFilter subsites. Following questions gathered other self-reported behaviors, such as how often the participant listened to the podcast or tracks on MetaFilter Music, ways in which they interacted with other MeFites off-site (e.g., on SNSes such as Facebook or Twitter, through chat channels, or other mediums and social contexts), and whether they had attended MetaFilter Meetups prior to the survey. This provided a better sense of participants’ engagement with various aspects of the MetaFilter community, including the frequency and types of interactions they may have outside of their MetaFilter social network. This information was also helpful to gauge the probability of participants’ exposure to spoken variants of the M-Set. Conversely, it was assumed that MeFites who did not listen to podcasts or music tracks and did not attend meetups or interact with MeFites outside of MetaFilter, had little to no experience of hearing spoken forms of the M-Set.

The next several questions gathered information about the pronunciation choices, rationales, and thought given to the M-Set. In the 2010 survey, MeFites were provided with six clickable audio samples for variant pronunciations of the M-Set. The audio samples were presented in a random order for each survey participant.

The wording of this question was as follows:

“Think about the word ‘MeFi’. Say it out loud if you need to. Which audio recording below best matches the way you currently say ‘MeFi’? (Click on each option to hear digital audio samples. Next to each word is an approximated transcription, written in brackets, using the International Phonetic Alphabet (IPA) corresponding to each audio sample. If you do not know IPA, do not worry; just choose by using the audio samples associated with each option and completely ignore the transcriptions. The order of these audio samples is random.)”

Figure 13. Screenshot of ‘MeFi’ Pronunciation Survey Question



Audio samples were created using AT&T Natural Voices® Text-to-Speech Demo (located at <http://www2.research.att.com/~ttsweb/tts/index.php>) and modified in Praat where necessary to conform to English stress, pitch and vowel targets that sounded natural to hearers. These audio samples were evaluated by several MeFites and linguists (including two phonologists) for naturalness and categorical discreteness. The resulting audio samples were acoustically identical minimal pairs, differing in only their first and/or second vowel.

Survey participants were asked which pronunciation they (would) prefer to use in speech for ‘MeFi’ and ‘MeFite’. They were also asked to rate how strongly they felt the pronunciations they chose for the M-Set are ones they (would) use exclusively. Participants were encouraged to share their reasons for having chosen their preferred pronunciation, including a prompt to give an explanation for any changes in pronunciation they may have made. This was a paragraph-style answer box, with no limits on length of responses.

All questions regarding pronunciation preferences for the M-Set were separated for ‘MeFi’ and ‘MeFite’, so that participants could give individual responses for each. This was especially important for participants who had differing pronunciations, rationales or opinions for each variable.

Next, participants were asked to select from three options as to how much thought they had given to the pronunciation of ‘MeFi’ or ‘MeFite’ prior to the survey. This allowed for a broad measure of participants’ metalinguistic awareness of the long-standing pronunciation debate about the M-Set, which could later be correlated with their level of involvement and length of time on the site.

The next questions in the surveys involved the collection of demographic data, starting with age and gender. In the 2010 survey, the gender question included four options: Male, Female, Transgender, and Other (with a text box for further explanation). While more inclusive than many studies that are limited to two gender options, it was discovered during the 2010 survey that this method was problematic for other reasons. For example, many trans* people²⁸ do not self-identify as the label ‘transgender’ and/or would have preferred the option of clicking more than one choice in the question. This question was changed to a free-form text answer box in the 2012 survey to allow participants to self-identify and give as much rationale as they wished to disclose. The gender question included a prompt that read, “*This is free-form, go nuts!*”. This informal wording replicated the prompts found in user profiles, thereby recreating a familiar style of instruction for the question; this informal prompt was used for several other free-form questions in the 2012 survey as well.

The next questions enquired about whether the participant was a native English speaker,

28. ‘trans*’ (with the asterisk) is one currently accepted neutral and inclusive way to refer to gender identities possibly including but not limited to: transgender, transsexual, transvestite, genderqueer, etc.

as well as any other languages the participant may have learned, spoken or studied. This information often corroborated responses given in earlier questions about rationales for pronunciation choices.

Following this, questions concerning the participant's dialect and current country of residence were provided; these provided a measure for that participant's potential exposure to other dialects or languages. For example, an American English speaker who was currently living outside the US may therefore have been familiar with alternative sound patterns that were consistent and standard for that speaker's current speech community.

The final demographic enquiry concerned race/ethnicity. In addition to not being captured correctly in the database, the question's structure was problematic — like the gender question, multiple responses were not allowed to be selected and the response format was multiple choice rather than a free-form fill-in. Responses for this question in the 2010 survey were not included in the data analysis.

The final question of the surveys allowed participants to provide any additional comments about the survey, as well as to opt-in to any further studies about their participation on MetaFilter. This question provided valuable feedback about the study design, suggestions for further areas to explore, and as encouragement and critique pertaining to the research.

4.5.1.1 *Changes Made in the 2012 MetaFilter Survey*

The two surveys were designed to be as similar as possible, for comparability in data analysis. However, after considering community feedback on the 2010 survey, as well as the data analysis of the 2010 survey, some improvements and additions were made. Only the modifications and additions are explained below.

The initial questions regarding site participation behaviors and off-site behaviors were expanded and restructured. A new subsite, IRL, was added to MetaFilter in June 2010, and therefore needed to be included in the question concerning how often the participant visits each subsite. The list of choices for visitation frequency for all subsites was slightly modified, based on participant feedback. The question regarding the ways in which participants interacted with other MeFites off-site was also expanded and restructured.

Two additional pronunciation variants were added as options to the 2012 survey (3c = [mɛfɪ], [mɛfɪt] and 4b = [maɪfaɪ], [maɪfart], including their corresponding sound files; see 3.5.4.1 **Differences in Phonotactic Distribution**, p. 80 for more information). Two new pronunciation questions were also added, asking participants to show how they would describe their pronunciations of 'MeFi' and 'MeFite' in writing. These questions had a limited character length (20) and therefore forced participants to use short, rhyming, or analogous wordforms in their descriptions, rather than long-form paragraph explanations. In many

cases, especially where descriptions unambiguously referred to a particular pronunciation variant, this question also allowed for comparison of the sound file selected to the description provided. Some surveys were corrected as a result. This was especially helpful when/where participants were unable to listen to the sound clips and informed the researcher that they had difficulties and/or left the question blank, guessed, or selected variants at random. All instances were clarified and corrected in the data.

Several question formats were changed from multiple choice formats in the 2010 survey to free-form fill-in answer formats in the 2012 survey. This applied to data collected concerning gender identity, ethnicity, dialect, and the addition of a short-answer fill-in question about nationality. These modifications allowed participants to express themselves with as little or as much detail as they saw fit, rather than be forced into selecting from a list of options, which may not have been entirely accurate or wholly representative of how the participants perceived themselves.

Some question formats were changed from short-answer fill-in options in the 2010 survey to multiple choice formats in the 2012 survey. These questions concerned place of residency (US state and/or non-US country), native language and other language experience, where the participant could fill-in up to three other languages they had experience with and select from a list of options about their level of expertise in each.²⁹ This greatly aided data normalization and analysis. For example, the current country of residence was a short-answer fill-in for the 2010 survey, which required extensive work to normalize the data (e.g., mostly owing to variation in spelling, capitalization, and punctuation), as well as not allowing for some important distinctions.³⁰ In the 2012 survey, the residence question was a drop-down list containing all of the recognized countries at the time of the survey, in alphabetical order. The output of this 2012 survey question was therefore already normalized and ready for data analysis.

A question about the participant's current post code was also added to the 2012 survey. This could allow for the creation of more precise geographic maps and data visualizations of pronunciation choices and other social factors beyond broad categories such as country or US state. However, owing to constraints on time and capabilities, post code data was not used in any analyses.

4.5.1.2 *Final Collection Procedures for the Survey Data*

The site moderators coded and hosted both surveys. After the close of each survey,

29. This survey question limited participants to list up to three languages studied. While providing additional lines would have facilitated a more thorough account of a participant's linguistic experience, the goal of keeping the survey concise was prioritized.

30. For example, some participants listed 'United Kingdom' as their current country of residence, not allowing the data to be further separated into the dialectally-diverse countries and territories that comprise the UK.

a private link was made available to the researcher for downloading all survey responses in a comma-separated values (.csv) file. These files were downloaded and the files were reformatted and cleaned up using TextWrangler (free text editor software for Mac) and OpenRefine (free, open source data manipulation tool). The files were then added to the Oracle SQL database.

4.5.2 InfoDump Collection Procedures

InfoDump files were collected twice, each within one month of the surveys' closing dates, by visiting <http://stuff.metafilter.com/infodump/> and downloading the file labeled 'infodump-all.zip'. The files were unzipped, formatted and imported into the database.

4.5.3 Word Frequency Table Collection Procedures

Word-frequency tables were collected from the MetaFilter Corpus by visiting <http://stuff.metafilter.com/corpus/freq/> and downloading all of the available files sorted by year and subsite. These files were decompressed, formatted and imported into the database.

4.5.4 MetaTalk Thread Collection Procedures

Three of the six selected MetaTalk threads were posted by MeFites in years prior to the pronunciation surveys. These posts made direct enquiries to the MetaFilter community about their pronunciation of the M-Set. The content of the posts can be found in **Appendix A**.

I made the remaining three posts on MetaTalk during the course of the thesis research. Two of these three posts accompanied the 2010 and 2012 surveys, as a place for the community to discuss the research and their participation in it. The 2010 pronunciation survey MetaTalk thread, posted in March, 2010, generated 472 comments; the 2012 pronunciation survey MetaTalk thread, posted in August 2012, generated 306 comments. The content of these posts is provided in **Appendix B**.

The last MetaTalk post related to the research enquired about norms, in-jokes and memes on MetaFilter, to directly elicit qualitative data about the MetaFilter register, and enregisterments that were specific to or salient in the community. The main text of that post can be found in **Appendix E**, along with a list of selected results from that enquiry.

These six MetaTalk threads were visited and viewed as HTML source code. The full text from these pages was copied and pasted as text files in TextWrangler. The HTML was stripped out and the files were reformatted into spreadsheet columns. The data were further cleaned up using OpenRefine. The data were manually coded and then added to the database.

4.6 Data Analysis Procedures

The aims of data analysis were to find out which social factors were significant predictors of variant choice and to observe and report on trends in usage and opinion for the M-Set variables over time, by key subgroups within the MetaFilter community. This demonstrates the spread of a particular linguistic innovation and the mechanisms by which those changes occurred. The procedures undertaken to accomplish these tasks are outlined in the following pages.

4.6.1 Data Correction and Adjustment

Various validation procedures were necessarily undertaken to make the data comparable and properly prepared for analysis. First, the database was extensively queried and reorganized to sort all data related to the M-Set. New tables were created, combining the data from the two surveys, InfoDump, MetaFilter Corpus, and coded MetaTalk discussion threads for each participant. The resulting data required further manipulation and intensive preparation involving error correction, normalization and additional coding.

4.6.1.1 *Error Correction*

Adjustments were made to responses in the surveys at the request of participants who contacted the researcher with corrections or clarifications. These changes or omissions were usually owing to technical difficulties or mistakes participants made (or thought they made) in filling out the surveys.

In addition to responding to emails from participants, the entirety of the MetaTalk thread and all paragraph responses in the surveys were reviewed for comments concerning possible corrections to survey data. Participants were contacted when possible and necessary, to confirm responses and changes to survey data.

Lastly, the 2012 survey included two questions requesting participants to explain their pronunciation choices in writing (in 20 characters or less). This allowed me to verify the ‘eye-dialect’ spellings matched the pronunciation variants chosen and the M-Set rationales given. Data were adjusted where responses were unambiguously errors in selecting the audio sample.

4.6.1.2 *Data Normalization*

Responses to short-answer fill-in questions from the surveys needed extensive review and normalization. These included questions regarding ways of interacting with other MeFites, current country of residence (2010 survey only), gender identity, ethnicity, nationality, and dialect. The last four categories in this list applied to the 2012 survey only.

For example, the gender identity category contained 377 unique responses (after variation

in capitalization, punctuation and superfluous spaces, etc., were removed). Responses in this category ranged from expected answers such as ‘male’ or ‘female’ to obscure or euphemistic descriptions of male or female gender categories. A substantial portion of the variation was owing to the vast range of gender identities that fall outside a male/female binary construct, including but not limited to, trans* identities, intersex, and genderqueer descriptions. Many participants also volunteered additional information pertaining to sexual orientation, gender expression or other features relating to gender identity. For example, it was not uncommon to find responses such as “*Biologically female cisgender lesbian*” or “*Male, but fond of dresses*”.

For all short-answer fill-in survey questions, a duplicate category was created for normalized responses. In the case of gender identity, the 377 unique responses were narrowed down to three gender identity categories: male, female, and QUILTBAG.³¹

Data were also normalized for comparability between surveys. For example, the options available to designate the frequency of visitation to each of the MetaFilter subsites differed between the two surveys. This required that responses be reformulated into fewer, broader categories that allowed the responses to be compared over time. Similarly, the different measures used to capture a participant’s current country of residence between the two surveys meant that the countries of the United Kingdom needed to be relabeled as such, so that they were comparable; the original values were maintained in a separate column for later qualitative analysis.

4.6.1.3 Coding M-Set Rationales from Surveys

Of the 2,521 surveys in 2010, there were 2,005 participants (80%) who supplied a written rationale for their chosen pronunciation of MeFi. In the 2012 survey, 1,472 (75%) of the 1,957 participants supplied a rationale for their pronunciation. These figures were slightly smaller for MeFite; 1,806 surveys (72%) in 2010 and 1,302 surveys (67%) in 2012 contained rationales for the pronunciation of MeFite.

In total, there were 6,585 written rationales for the pronunciation of the M-Set. These rationales were initially reviewed to develop a sense of the salient categories to code for. The rationales were manually coded and reviewed for errors. Coding resulted in approximately twenty new categories to include in quantitative analyses (see 4.4.5 M-Set Stance Measures, p. 99).

31. As per the Queer Dictionary, <http://queerdictionary.tumblr.com/post/3899608042/quiltbag>: “QUILTBAG is an acronym. It stands for Queer/Questioning, Undecided, Intersex, Lesbian, Transgender/ Transsexual, Bisexual, Allied/Asexual, Gay/Genderqueer. It is meant to be a more inclusive term than GLBT/ LGBT and to be more pronounceable (and memorable) than some of the other variations or extensions on the GLBT/LGBT abbreviation.”

4.6.2 Quantitative and Qualitative Analyses

All factors able to be quantified were analyzed independently, using descriptive statistics involving chi-square tests to determine whether the factor was correlated with ‘MeFi’ pronunciation. The results from each outcome were supported by qualitative analyses from survey rationales and public comments on MetaFilter.

Qualitative analyses further revealed important factors in pronunciation choice. This new information could then be looked at in the broader context of discourse within MetaFilter. Discussions about the M-Set were reread and comments representing pivotal ideas in the discourse were flagged for further qualitative analysis.

Particular stances were also tracked in the selected MetaTalk threads. This provided additional support for the statistical outcomes in the survey data, as well as helped explain how attitudes about the M-Set had evolved over time. Common themes pointed to indexical associations that participants have about the variants, which demonstrate ways in which these variables are being enregistered within the community.

4.7 Summary of Methodology

The four sources of data (surveys, the InfoDump, the MetaFilter Corpus and MetaTalk threads) complemented each other to provide a rounded picture of the actual distribution of variants, the perceived distribution of variants, and the evolution and awareness of their enregisterments over time. The 4,478 total surveys collected (2010 and 2012 survey combined) provided over 6,500 individual responses about the pronunciation of the M-Set. Data analysis revealed whether the distribution of choices and stances aligned with the distribution of publicly-stated choices and stances by MeFites, and if not, how those distributions differed. The InfoDump and corpus files showed the extent to which other factors influenced the choices participants made, and how that affected the enregisterment of forms over time, as negotiated through a text-based medium that is a persistent and meaningful record of a shared history and identity.

Chapter 5: Data Results

5.1 Introduction to Data Results

This chapter presents results from the data collection and analysis procedures that were outlined in **Chapter 4: Methodology, p. 89**. The measures used to analyze data were grouped into categories and will be presented here, starting with the distribution of the M-Set, followed by metalinguistic awareness factors, and concluding with demographic factors. Results pertaining to social engagement measures and M-Set stances will be covered in **Chapter 6: Enregisterment, p. 153**. All measures included in this thesis were analyzed across three subsets of data, where it was applicable to do so. See **Table 6** for an overview of the data subsets.

Table 6. Data Subsets

Data Set Name	Population	Sample Size	Data Source	% of Active Population
Survey Data	All MetaFilter survey participants.	2,521 in 2010 1,957 in 2012	MetaFilter Surveys	15% in 2010 12% in 2012
Panel Data	Those who participated in both the 2010 and the 2012 survey.	769 across both years	MetaFilter Surveys	5% in 2010 5% in 2012
Rationales	Participants who provided an explanation for their M-Set pronunciation in their survey.	1,974 in 2010 1,472 in 2012	MetaFilter Surveys	12% in 2010 9% in 2012

The Survey Data are comprised of all responses from MetaFilter participants who took either the 2010 or the 2012 survey (including those who took both surveys). Of the 2,521 survey participants in 2010, 2,471 (98%) were active MetaFilter members in the preceding twelve months of the survey (measured as a full year ending on the first day of the 2010 survey, March 24, 2010), representing 15% of the total active MetaFilter population. Of the 1,957 survey participants in 2012, 1,905 (97%) were active MetaFilter members in the non-calendar year prior to the survey (measured as a full year ending on the first day of the 2012 survey, August 22, 2012) and represented 12% of the total active MetaFilter population.

The Panel Data is a subset of the Survey Data. These were drawn from the responses of participants who provided surveys in both 2010 and 2012. Surveys submitted by the same participants in different years allowed factors and attitudes about the M-Set to be observed over time. There were 769 participants who took both surveys, representing 34% of the entire survey data and 5% of the total active MetaFilter userbase.

Of the 2,521 surveys in 2010, 1,974 (78%) included rationales explaining their reasoning for their ‘MeFi’ pronunciation choice. Of the 1,957 surveys in 2012, 1,472 (75%)

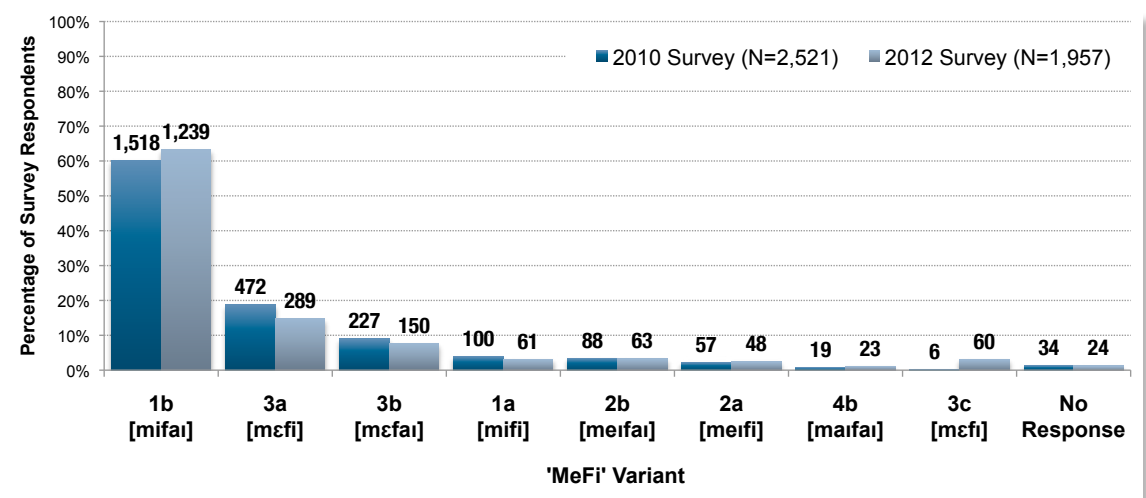
included rationales. These participants represented 12% of the total active MetaFilter userbase in 2010 and 9% in 2012.

5.2 The Distribution of the M-Set

5.2.1 The Pronunciation of ‘MeFi’ — Survey Data

The Survey Data revealed the distribution of the preferred pronunciations of ‘MeFi’ by MetaFilter participants. All eight variants, including their variant codes, separated by survey year, are shown in Figure 14.

Figure 14. Pronunciation of ‘MeFi’ by Variant and Survey Year



The most preferred variant overall, for both survey years, was 1b – [mifaɪ]. This variant accounted for 60% of the data in 2010 and 63% in 2012.

The second and third most preferred pronunciations were 3a – [mɛfi] and 3b – [mɛfaɪ], both of which share the first vowel [ɛ] but differ in their second vowels. Preference for 3a – [mɛfi] was roughly twice as frequent than that of the preference for 3b – [mɛfaɪ] (19% as compared to 9% in 2010, and 15% as compared to 8% in 2012).

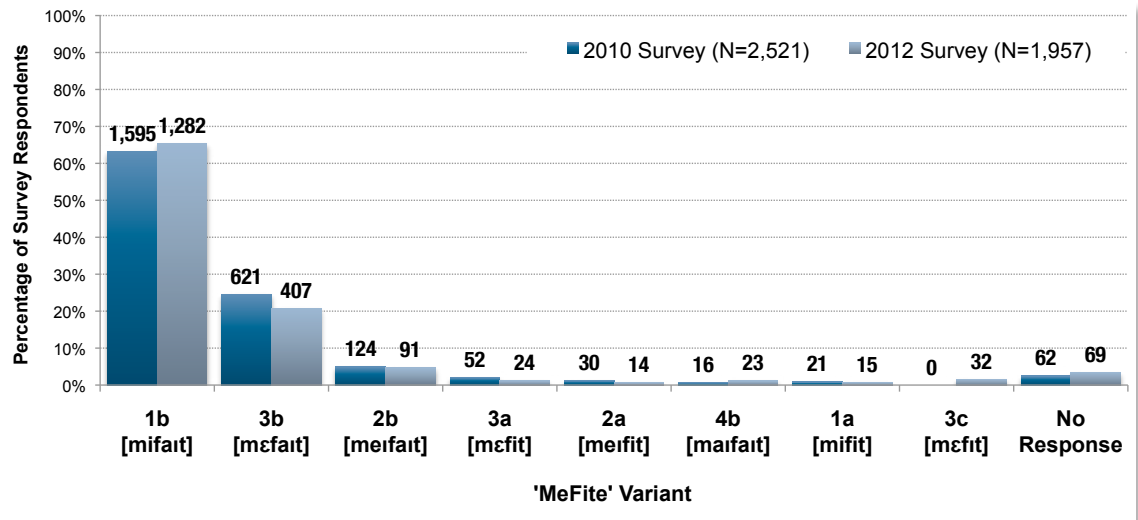
These three variants encompassed 88% of the Survey Data in 2010 and 86% of the Survey Data in 2012. Through posts on MetaFilter (especially in MetaTalk) and comments by moderators on podcasts or by MeFites at meetups, there was a common perception of these variants as the most commonly preferred ones. However, until these surveys, the actual distribution of the most preferred variants could not be verified.

As explained in 4.5.1.1 *Changes Made in the 2012 MetaFilter Survey*, p. 104, the 3c – [mɛfi(t)] and 4b – [maɪfaɪ(t)] variants were not options in the 2010 survey; results for those variants in 2010 data were added manually after the survey closed, based on discussions with those participants about their choices. Therefore, the 2012 totals for these variants were much higher than in 2010.

5.2.2 The Pronunciation of ‘MeFite’ — Survey Data

The Survey Data revealed that the pronunciation of ‘MeFite’ had a similar distribution as that of ‘MeFi’, as shown in Figure 15.

Figure 15. Pronunciation of ‘MeFite’ by Variant and Survey Year



The most preferred variant for ‘MeFite’ was 1b – [mifart]. This variant made up 63% of the data responses in 2010 and 66% in 2012.

The [-ɑrt] ending was strongly preferred, representing 93% of the responses in 2010 and 92% in 2012. In survey rationales, many MeFites stated that their preference for an [ɑrt] ending had to do with semantic associations of “*belonging*” and “*being a denizen of*”, which the ‘-ite’ suffix inspired for them. A variety of analogous forms ending in ‘-ite’ were used to illustrate this; the most common referred to being an inhabitant of a city (e.g., ‘Denverite’, ‘Londonite’) or member of an ethnic or social group (e.g., ‘Israelite’, ‘Luddite’, ‘socialite’). Also frequently mentioned was analogy with ‘fight’, which many felt aptly described aspects of MetaFilter culture as well as associations with MeFight Club, a popular MetaFilter spin-off site for collaborative online gaming. Additionally, many MeFites expressed negative associations with what they perceived as the “*feet*” suffix and therefore sounding “*gross*” or “*inappropriate*”. Those who preferred the [fit] ending often stated that they did so out of a desire for consistency with their preferred pronunciation of ‘MeF[i]’.

5.2.3 The Pronunciation of the M-Set — Panel Data

The distributions of the M-Set variants for participants who took both surveys did not significantly differ from the overall 2010 and 2012 Survey Data distributions for the M-Set. As the Panel Data represented 34% of the Survey Data and 5% of the active MetaFilter population, some results of the Panel Data may be generalized over the rest of the Survey Data as well as the MetaFilter population. This is important, as the Panel Data results are analyzed

from the perspective of individuals' choices as they change over time and are therefore an invaluable element in understanding this process of enregistration. The distributions of 'MeFi' and 'MeFite' in the Panel Data are shown in Figure 16 and Figure 17.

Figure 16. Pronunciation of 'MeFi' by Variant and Survey Year — Panel Data, N=769

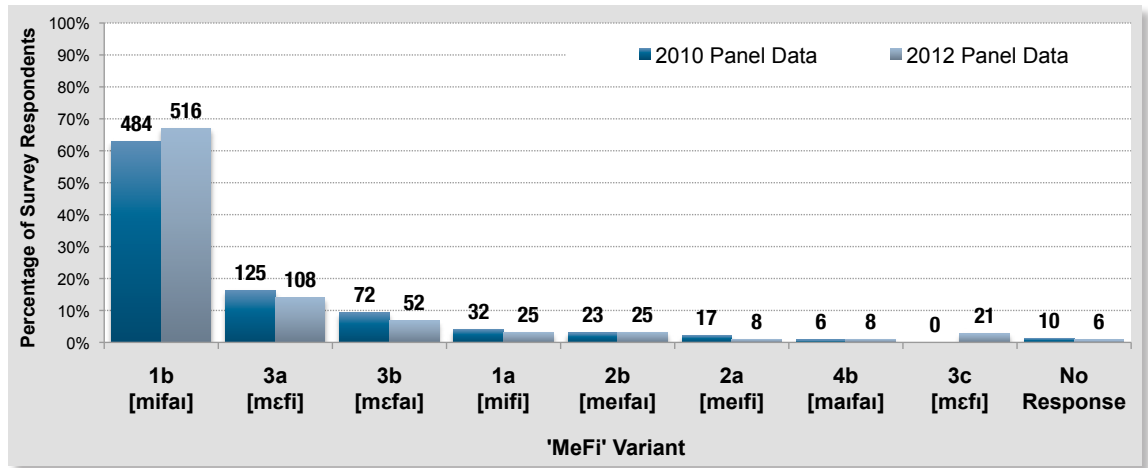
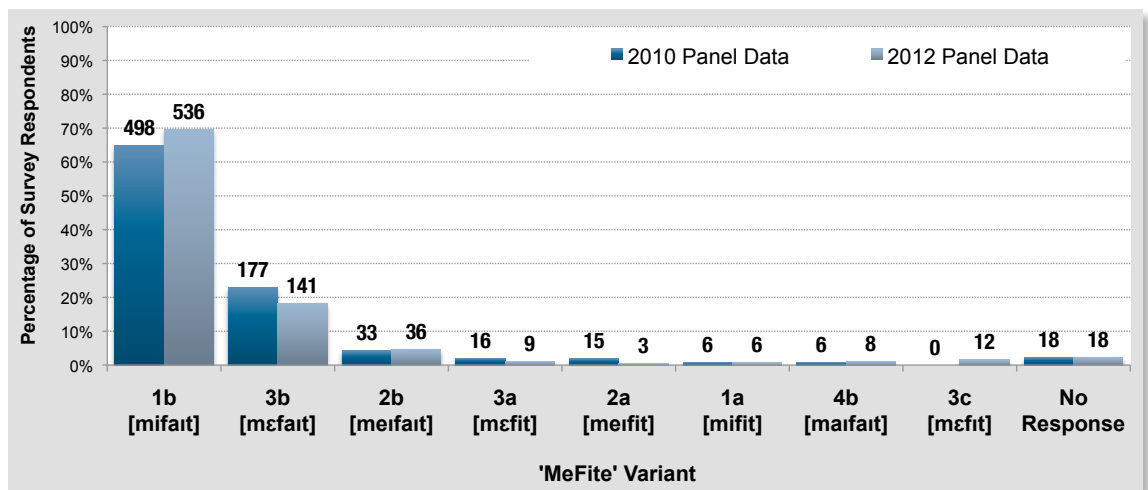


Figure 17. Pronunciation of 'MeFite' by Variant and Survey Year — Panel Data, N=769



5.2.4 Changes in M-Set Distributions Over Time

The Survey Data examined across both years revealed overall trends in increased or decreased preference for M-Set variants. The Panel Data provided an even closer view of these changes over time, whereby it was possible to see fine differences in the degree to which the individual variants were preferred over others in 2012 as compared to 2010.

5.2.4.1 Change over time for 'MeFi' — Survey Data

Table 7 shows the overall trends for 'MeFi' variants, illustrating the percentage of increase or decrease in each variant in the 2012 survey as compared to the 2010 outcomes.

Table 7. Change in Variant Choice for ‘MeFi’ Over Time

‘MeFi’ Variant	% of 2010 Survey Data	% of 2012 Survey Data	% of Change Between Surveys
1a - [mifi]	3.97%	3.12%	↓ 0.85%
1b - [mifaɪ]	60.21%	63.31%	↑ 3.10%
2a - [meɪfi]	2.26%	2.45%	↑ 0.19%
2b - [meɪfaɪ]	3.49%	3.22%	↓ 0.27%
3a - [mɛfi]	18.72%	14.77%	↓ 3.96%
3b - [mɛfaɪ]	9.00%	7.66%	↓ 1.34%
3c - [mɛfi]	0.24%	3.07%	↑ 2.83%
4b - [maɪfaɪ]	0.75%	1.18%	↑ 0.42%
Answer left blank	1.35%	1.23%	↓ 0.12%

There was a 3.1% overall percentage point increase in the 1b - [mifaɪ] variant and a 3.96% overall decrease for the 3a - [mɛfi] variant in the 2012 Survey Data. These distributional differences in ‘MeFi’ pronunciation between the 2010 and 2012 Survey Data were statistically significant: $\chi^2 = 15.782$, $df = 5$, $p = 0.0074$.³²

It should also be noted that the 2.83% percentage point increase in preference for the 3c - [mɛfi] and 0.42% increase in the 4b - [maɪfaɪ] variants were the result of them being added as options in the 2012 survey.

5.2.4.2 Change over time for ‘MeFite’ — Survey Data

Table 8 shows increases and decreases in preference for ‘MeFite’ variants in the 2012 surveys as compared to the 2010 results.

Table 8. Change in Variant Choice for ‘MeFite’ Over Time

‘MeFite’ Variant	% of 2010 Survey Data	% of 2012 Survey Data	% of Change Between Surveys
1a - [mifit]	0.83%	0.77%	↓ 0.07%
1b - [mifart]	63.27%	65.51%	↑ 2.24%
2a - [meɪfit]	1.19%	0.72%	↓ 0.47%
2b - [meɪfart]	4.92%	4.65%	↓ 0.27%
3a - [mɛfit]	2.06%	1.23%	↓ 0.84%
3b - [mɛfart]	24.63%	20.80%	↓ 3.84%
3c - [mɛfit]	0.00%	1.64%	↑ 1.64%
4b - [maɪfart]	0.63%	1.18%	↑ 0.54%
Answer left blank	2.46%	3.53%	↑ 1.07%

32. All chi-square tests analyzing change over time in these data excluded categories 3c and 4b; they were not options in the 2010 survey and would therefore skew test results. The 2010 data shown for these participants’ pronunciations were added manually, based on individual discussions with those participants.

Data Results

The change in the distribution of ‘MeFite’ pronunciation between the two survey years was also statistically significant: $\chi^2 = 14.066$, $df = 5$, $p = 0.0151$, with a 2.24% percentage point increase in preference for 1b – [mifɑɪ], further demonstrating an overall trend over time toward the 1b variants.

The 3b – [mɛfɑɪ] variant was chosen by 24.63% of MeFites in 2010 and 20.8% in 2012, resulting in a 3.85% overall percentage point decrease in preference over time. All other variants which were included as options in the 2010 survey also showed a decrease in preference in 2012, suggesting that a small but significant number of these MeFites may now prefer the 1b pronunciation.

5.2.4.3 Change over time for the M-Set — Panel Data

The Panel Data allow the results to be organized according to the participants who changed their pronunciation choice between 2010 and 2012 versus those who remained with their original choices. First, the overall trends will be shown, for comparison with the Survey Data. Then the numbers and percentages of participants who switched variants in 2012 will be shown.

Table 9. Shift in Variant Choice for ‘MeFi’ Over Time — Panel Data

‘MeFi’ Variant	% of 2010 Panel Data	% of 2012 Panel Data	% of Change Between Surveys
1a – [mifi]	4.16%	3.25%	↓ 0.91%
1b – [mifɑɪ]	62.94%	67.10%	↑ 4.16%
2a – [mɛrfi]	2.21%	1.04%	↓ 1.17%
2b – [mɛrfɑɪ]	2.99%	3.25%	↑ 0.26%
3a – [mɛfi]	16.25%	14.04%	↓ 2.21%
3b – [mɛfɑɪ]	9.36%	6.76%	↓ 2.60%
3c – [mɛfɪ]	0.00%	2.73%	↑ 2.73%
4b – [maɪfɑɪ]	0.78%	1.04%	↑ 0.26%
Answer left blank	1.30%	0.78%	↓ 0.52%

For the Panel Data participants, the change in preference for ‘MeFi’ between survey years was not statistically significant ($p = 0.09$), despite an overall 4.16% percentage point increase in preference for 1b – [mifɑɪ]. This is contrasted with the change in distribution of ‘MeFi’ in the Survey Data, an effect which was found to be significant even though the overall increase in preference for 1b – [mifɑɪ] was smaller, at 3.1% percentage points. These differing outcomes can be explained by the much larger sample size of the Survey Data than the Panel Data, as well as by the decreases in preference for the other variants for both data sets, which were more evenly distributed in the Panel Data than they were in the Survey Data. For example, there was a 3.96%

Data Results

decrease in 3a – [mɛfi] and 1.34% decrease in 3b – [mɛfaɪ] in the Survey Data, but only 2.21% and 2.6% decreases for those variants, respectively, in the Panel Data.

Another important point to note is that while the overall trends in the Panel Data did not yield significant results, it does not necessarily follow that the size of the change over time for ‘MeFi’ did not. This is because the actual amount of variant switching was much greater in 2012 than is represented in Table 9, as much of the switching may have served to cancel out changes when the numbers are represented as overall percentages of increases and decreases.

Examining the Panel Data more closely, it was found that 165 of 769 Panel Data participants had changed their ‘MeFi’ pronunciation choice by 2012. Removing 16 participants who left the ‘MeFi’ pronunciation question blank in either 2010 or 2012, participants who changed pronunciation represented 22% of the remaining Panel Data. Table 10 illustrates these changes in pronunciation of ‘MeFi’ by individual variant³³, including the most preferred variant participants switched to from each variant category.

Table 10. ‘MeFi’ Variant Switching in Panel Data Participants

‘MeFi’ Variant	Total Count in 2010	Count and (%) who preferred this same variant in 2012	Count and (%) who switched to other variants in 2012	Count and (%) of most preferred other variant in 2012
1b – [mɪfaɪ]	484	446 (92.15%)	38 (7.85%)	3a – 12 (31.58%)
3a – [mɛfi]	125	78 (62.4%)	47 (37.60%)	1b – 19 (40.43%)
3b – [mɛfaɪ]	72	35 (48.61%)	37 (51.39%)	1b – 16 (43.24%)
1a – [mifi]	32	16 (50%)	16 (50%)	1b – 8 (50%)
2b – [mɛɪfaɪ]	23	5 (21.74%)	18 (78.26%)	1b – 14 (77.78%)
2a – [mɛɪfi]	17	3 (17.65%)	14 (82.35%)	1b – 7 (50%)
Answer left blank	10	0 (0%)	10 (100%)	1b – 6 (60%)
4b – [maɪfaɪ]	6	5 (83.33%)	1 (16.67%)	Blank – 1 (100%)
3c – [mɛfi]	0	0 (0%)	0 (0%)	N/A

The participants who chose the most preferred variants in 2010 also showed the least amount of switching to other variants in 2012. Of the participants who opted for the most preferred variant (i.e., 1b – [mɪfaɪ]) in 2010, only 7.85% switched to another variant in their 2012 survey (i.e., over 92% of the 1b group made the same choice in 2012 as they did in 2010). The variant that these participants switched to the most was 3a – [mɛfi], making up 31.58% of the share of the other variants that the formerly 1b participants could choose from. The 3a – [mɛfi] variant was also the second most popular variant choice overall, so it is not surprising that this was the most frequently chosen variant for 2010 1b – [mɪfaɪ] participants

33. Variants in this table are sorted in order of most to least preferred.

who changed their pronunciation choice. The 1b - [mifart] variant was the most preferred for participants to switch to in all other categories (excluding 4b, where one participant left the pronunciation choice question blank in 2012).

Table 11. Change in Variant Choice for ‘MeFite’ Over Time — Panel Data

‘MeFite’ Variant	% of 2010 Panel Data	% of 2012 Panel Data	% of Change Between Surveys
1a - [mifit]	0.78%	0.78%	0.00%
1b - [mifart]	64.76%	69.70%	↑ 4.94%
2a - [merfit]	1.95%	0.39%	↓ 1.56%
2b - [merfart]	4.29%	4.68%	↑ 0.39%
3a - [mɛfit]	2.08%	1.17%	↓ 0.91%
3b - [mɛfart]	23.02%	18.34%	↓ 4.68%
3c - [mɛfɪt]	0.00%	1.56%	↑ 1.56%
4b - [maɪfart]	0.78%	1.04%	↑ 0.26%
Answer left blank	2.34%	2.34%	0.00%

Change in preference for ‘MeFite’ between survey years achieved statistical significance, $\chi^2 = 15.431$, $df = 5$, $p = 0.0086$. This was most notably demonstrated by a 4.94% increase in percentage points in preference for 1b - [mifart] and a 4.68% decrease in preference for 3b - [mɛfart]. These were the two most popular variants for ‘MeFite’ and therefore represented the most data in the sample.

Examining individual variants, patterns similar to those seen for ‘MeFi’ are evident (with a few minor exceptions) for panel participants who changed their pronunciation of ‘MeFite’ in 2012 from their 2010 choice. Of the 733 Panel Data participants who selected a pronunciation variant for ‘MeFite’ in both survey years, 160 (22%) of them selected a different variant in 2012.

Table 12. ‘MeFite’ Variant Switching in Panel Data Participants

‘MeFite’ Variant	Count in 2010	Count and (%) who preferred same variant in 2012	Count and (%) who switched to other variants in 2012	Count and (%) of most preferred other variant in 2012
1b - [mifart]	498	446 (89.56%)	52 (10.44%)	3b - 19 (36.54%)
3b - [mɛfart]	177	110 (62.15%)	67 (37.85%)	1b - 47 (70.15%)
2b - [merfart]	33	8 (24.24%)	25 (75.76%)	1b - 17 (68%)
Answer left blank	18	0 (0%)	18 (100%)	1b - 13 (72%)
3a - [mɛfit]	16	4 (25%)	12 (75%)	3c - 4 (25%)
2a - [merfit]	15	0 (0%)	15 (100%)	1b - 8 (53.33%)
1a - [mifit]	6	1 (16.67%)	5 (83.33%)	1b - 4 (80%)
4b - [maɪfart]	6	4 (66.67%)	2 (33.33%)	Blank - 2 (100%)
3c - [mɛfɪt]	0	0 (0%)	0 (0%)	N/A

Again, the participants who opted for the most frequently preferred variants of ‘MeFite’ in 2010 showed the least amount of switching in 2012.³⁴ The most commonly switched-to variant was 1b - [mifart], with exceptions in both the 3a - [mɛfit] and 4b - [marfart] categories (3a switchers switched to 3c more than they switched to 1b, and two participants in the 4a group left their 2012 survey blank for pronunciation of ‘MeFite’).

Very few participants gave explicit rationales for why they had switched to other variants in 2012. Additionally, the rationales given were often inconclusive, ambiguous or contradictory. More general reasons for switching can be gleaned from analysis of common rationales for variant choices, discussed in 6.5 M-Set Stances, p. 187.

5.3 Metalinguistic Awareness Factors

Participants’ awareness of linguistic variation or language use in general can influence their linguistic behavior. Data for two measures of metalinguistic awareness were directly elicited in the surveys. The first measure concerned how strongly participants felt that the pronunciations they had chosen for the M-Set variants were ones that they would use to the exclusion of others. The second measure asked participants how much thought they had given to the pronunciation of M-Set terms prior to that year’s survey.

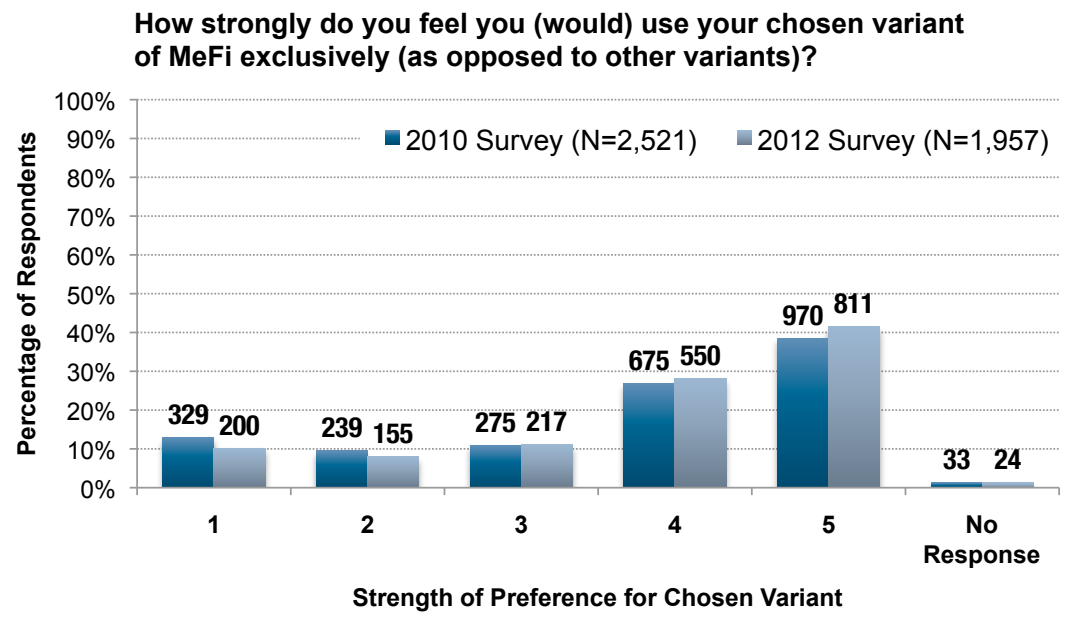
From these two measures, it could be shown how much MeFites might vary their pronunciation for the M-Set (e.g., how strongly they felt they might use other variants or not) and how much conscious thought they had had about that variation (e.g., as an indicator of awareness of the existence of variation and/or the debate about M-Set pronunciation on MetaFilter). Other measures of metalinguistic awareness from survey rationales will also be presented in this section.

5.3.1 Exclusivity of Use of Preferred M-Set Variants

Many MeFites’ pronunciation choices for the M-Set are unwavering, especially for those who have previously engaged in debates about the M-Set. However, other MeFites feel less strongly that their chosen variant is the one that they would use predominantly. Therefore, a measure of metalinguistic awareness was implemented, to gauge how strongly participants felt that they would use their chosen variant of ‘MeFi’ exclusively. This question aimed to target the amount of variance MeFites said they had in their pronunciation choice. The Likert measure ranged from 1–5, where category 1 was labeled with the category number and the word ‘Indifferent’, and category 5 was labeled with the category number and the words ‘Very strongly’.

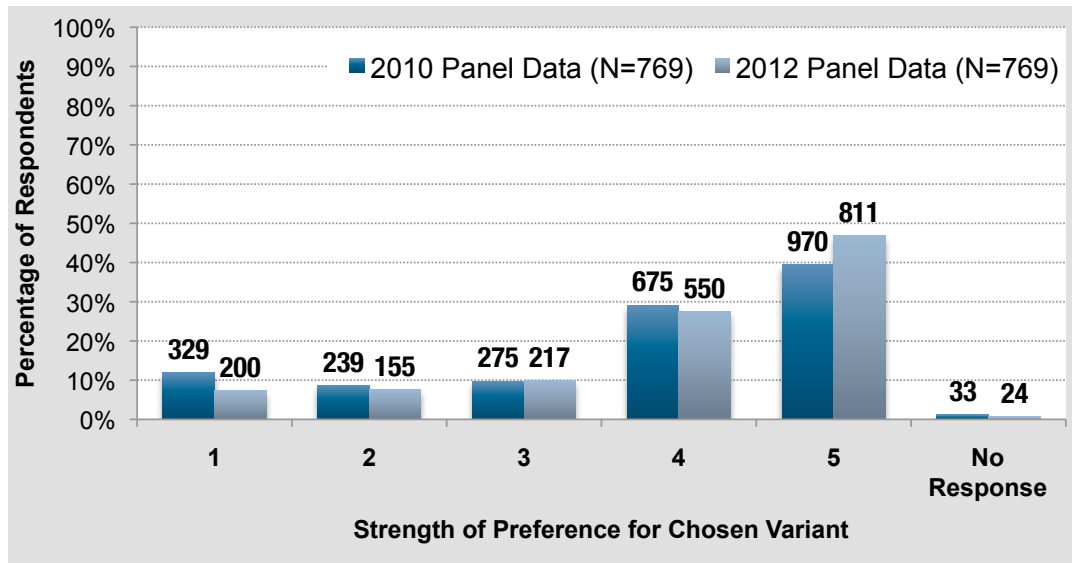
34. Some percentages may be based on data where the counts of the variants were low (i.e., very few participants selected these variants in 2010).

Measure of Exclusive Use of 'MeFi' Variant — 2010 and 2012 Survey Data



$$\chi^2 = 13.757, df = 5, p\text{-value} = 0.0172$$

Figure 18. Measure of Exclusive Use of 'MeFi' Variant — 2010 and 2012 Panel Data

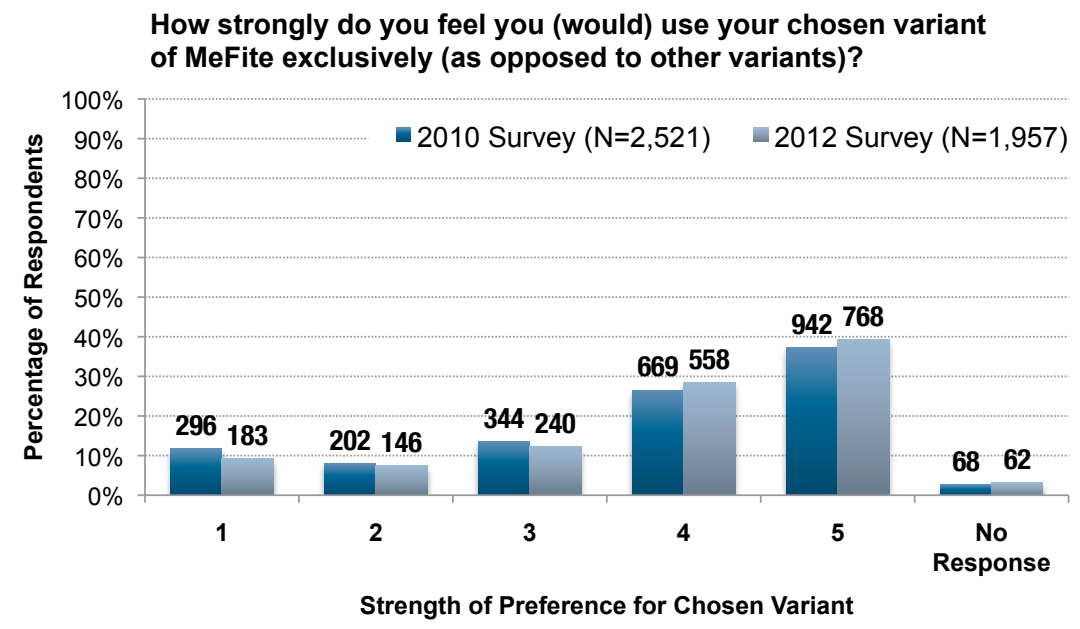


$$\chi^2 = 15.387, df = 5, p\text{-value} = 0.0088$$

The differences between the 2010 and 2012 samples in exclusive preference for chosen 'MeFi' variant were found to be statistically significant in both the Survey and Panel Data sets. A greater percentage of participants in 2012 chose higher exclusivity levels (categories 4 and 5) than in 2010; a lesser percentage of participants in 2012 chose lower exclusivity levels (categories 1 and 2) than in 2010. Overall, MeFites in the 2012 survey reported that they felt more strongly about the exclusive use of their chosen variant than they did in 2010. Additionally, the 2012 Panel Data participants significantly differed from the 2012 Survey data participants, with those who

took both surveys (i.e., the Panel Data) demonstrating the highest levels of exclusive use of their preferred variant out of any other data set or survey year: $\chi^2 = 9.765$, $df = 4$, $p = 0.0446$.

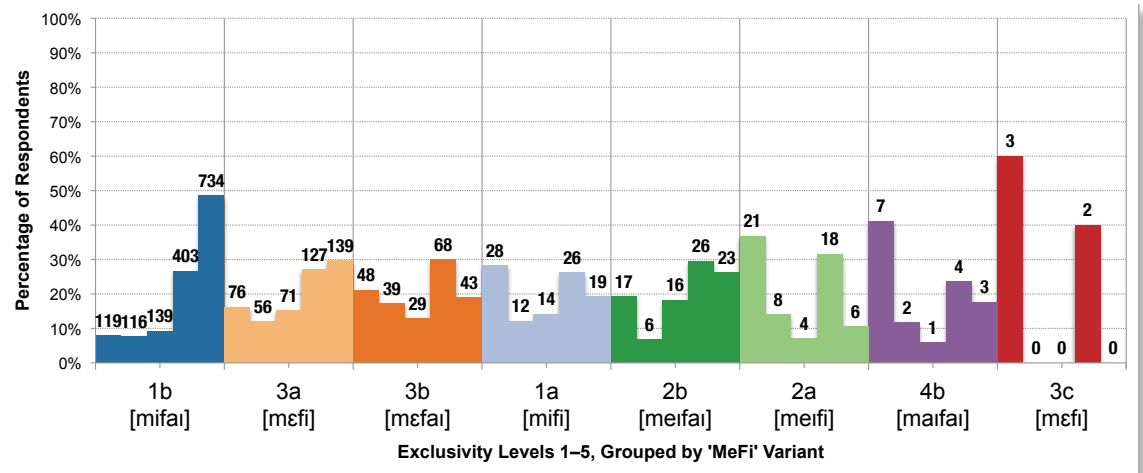
Figure 19. Measure of Exclusive Use of ‘MeFite’ Variant — Survey Data



Similar results were found for the ‘MeFite’ variable. There was a 3% percentage point decrease in the ‘Indifferent’ category in 2012. This also corresponded with increases in categories toward the ‘Very Strongly’ end of the scale. These differences in exclusive use between the two survey years were found to be statistically significant (but slightly less so than for the ‘MeFi’ variable): $\chi^2 = 10.499$, $df = 4$, $p = 0.0328$.

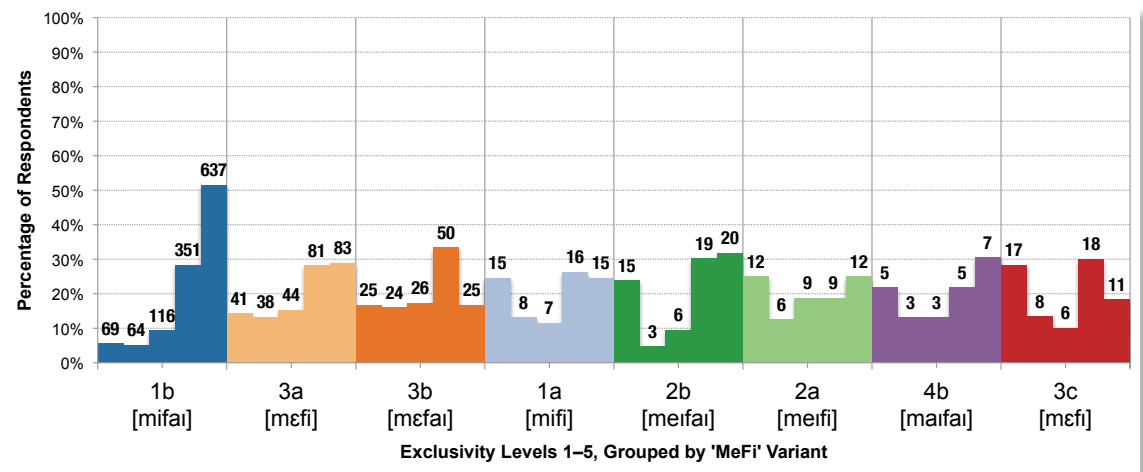
Lastly, the differences between exclusive use of ‘MeFi’ variants and exclusive use of ‘MeFite’ variants in the 2010 survey were statistically significant. A greater percentage of respondents felt more strongly about exclusive use of their chosen ‘MeFi’ pronunciation than they did about their ‘MeFite’ pronunciations: $\chi^2 = 12.728$, $df = 4$, $p = 0.0127$. This suggests that there was more metalinguistic awareness about the pronunciation of ‘MeFi’ than of ‘MeFite’ in 2010.

Figure 20. 'MeFi' Pronunciation by Exclusivity of Use — 2010 Survey Data, N=2,473



$\chi^2 = 255.482, df = 28, p\text{-value} < 0.0001$

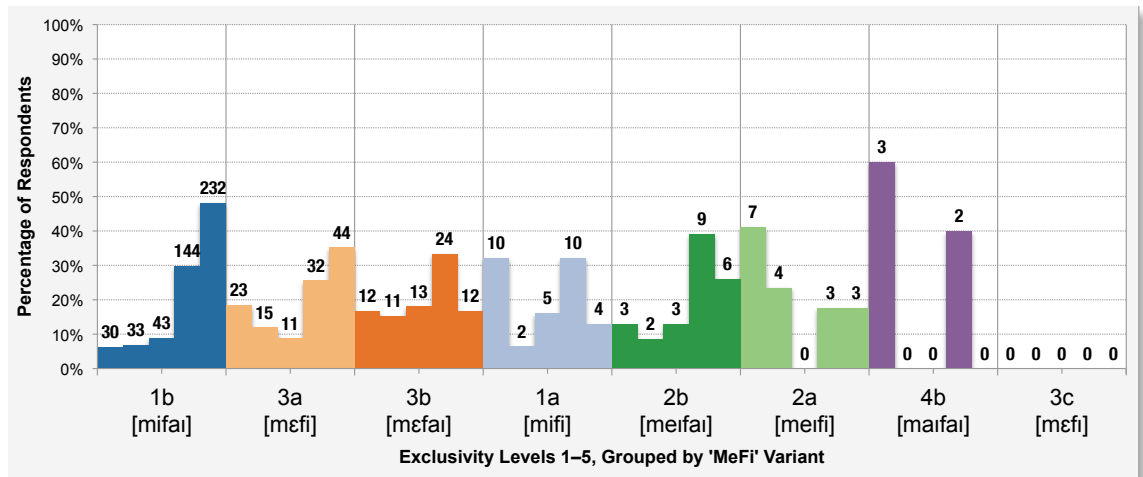
Figure 21. 'MeFi' Pronunciation by Exclusivity of Use — 2012 Survey Data, N=1,929



$\chi^2 = 256.577, df = 28, p\text{-value} < 0.0001$

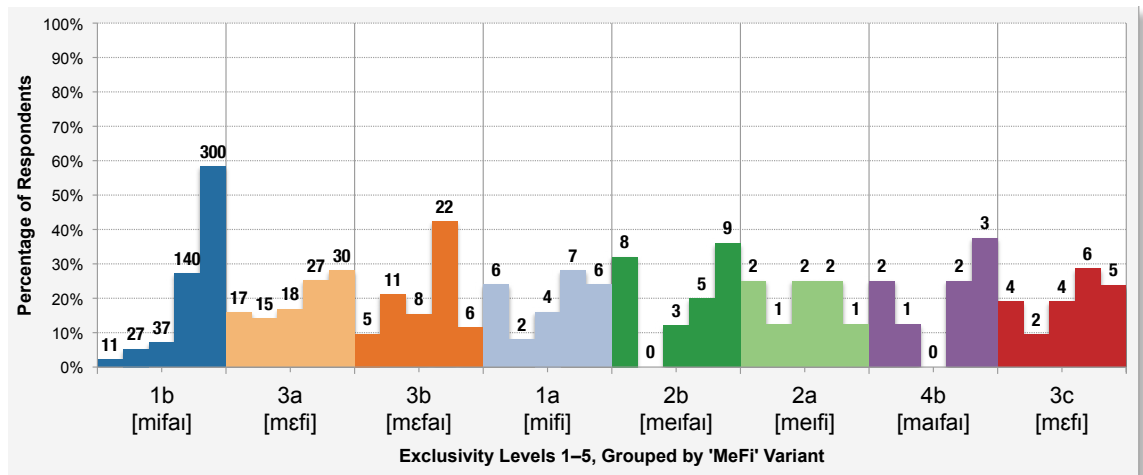
Note: For each variant, the columns were ordered from category 1 – 5, with category 1 = “Indifferent” (leftmost) and category 5 = “Very strongly” (rightmost). The survey question prompt was: “How strongly do you feel that the pronunciation of ‘MeFi’ that you have chosen is the one you (would) use exclusively (as opposed to the other options)?”

Figure 22. 'MeFi' Pronunciation by Exclusivity of Use — 2010 Panel Data, N=755



$$\chi^2 = 72.59, df = 16, p\text{-value} < 0.0001$$

Figure 23. 'MeFi' Pronunciation by Exclusivity of Use — 2012 Panel Data, N=761



$$\chi^2 = 106.1441, df = 8, p\text{-value} < 0.0001$$

Note: For each variant, the columns were ordered from category 1 – 5, with category 1 = “Indifferent” (leftmost) and category 5 = “Very strongly” (rightmost). The survey question prompt was: “How strongly do you feel that the pronunciation of ‘MeFi’ that you have chosen is the one you (would) use exclusively (as opposed to the other options)?”

The most preferred variant, 1b – [mifaɪ], showed a clear trend across both survey years — most of the participants who chose this variant felt very strongly that they would use it exclusively. Of the participants who chose 3a – [mɛfi] or 3b – [mɛfaɪ], more felt less strongly about exclusive use of their variant than did the 1b – [mifaɪ] group; that is, greater numbers from the 3a or 3b groups felt indifferent about exclusive use than in the 1b group. For the least preferred variants, equal or greater numbers of participants felt indifferent about those choices than those who felt strongly or very strongly.

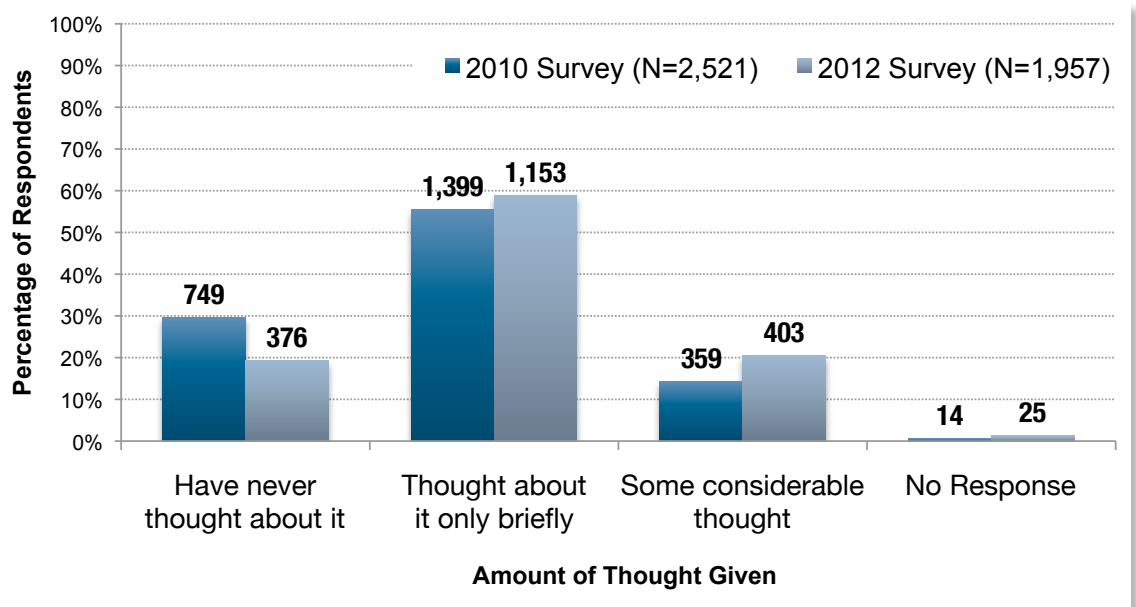
Similar overall patterns of exclusivity were found for variants of ‘MeFite’ in 2010 and 2012,

as well as for the M-Set in the Panel Data. These results help substantiate earlier findings, in which significant shifts toward the 1b – [mifaɪ] and [mifaɪt] variants were observed over time. In summary, a considerable number of MeFites switched to 1b variants, and a considerable number of MeFites also stated that they felt very strongly that they would use those variants exclusively. The majority of participants who preferred less popular variants claimed to feel indifferent and said they might use other variants (possibly including the more popular ones).

5.3.2 Amount of Thought Given to the Pronunciation of the M-Set

The survey question concerning the amount of thought given to the pronunciation of the M-Set asked whether thought had been given to either variant (or both variants) prior to that year’s survey. This was a categorical measure, comprising three options: “No thought given”, “Only brief thought given”, or “Considerable thought given”.

Figure 24. Amount of Thought Given to the Pronunciation of the M-Set



$$\chi^2 = 76.73, df = 2, p\text{-value} < 0.0001$$

The majority of MeFites from both survey years had “*thought about the pronunciation of the M-Set only briefly prior to the survey*” (55% in 2010 and 59% in 2012). The second largest category were those who had “*never thought about the pronunciation of the M-Set prior to the survey*” (30% in 2010 and 19% in 2012). Lastly, 14% in 2010 and 21% in 2012 had given the matter “*some considerable thought*”.

Since 34% of the 2012 survey participants had also taken the 2010 survey, it was expected that there would be a decrease in the “*never thought about it*” category and an increase in the other two categories from 2010 to 2012. This was indeed the case. In 2010, 30% of respondents

had never thought about the pronunciation of the M-Set prior to the survey, but only 19% stated they had never thought about it prior to the 2012 survey, representing a decrease of 11 percentage points over time. There was a 4% percentage point increase in “*brief thought*” between 2010 and 2012 and a 7% increase in “*considerable thought*” over time (14% in 2010 and 21% in 2012). These differences were found to be statistically significant: $\chi^2 = 76.73$, $df = 2$, $p < 0.0001$. Lastly, the 2012 Panel Data participants significantly differed from the 2012 Survey Data participants, with Panel Data participants having given more thought to the pronunciation of the M-Set: $\chi^2 = 77.012$, $df = 2$, $p < 0.0001$. This result is not surprising considering that the Panel Data participants had taken the 2010 survey, whereas the majority of the 2012 Survey Data participants (which includes the Panel Data participants) had not.

5.3.2.1 ‘MeFi’ Pronunciation and Amount of Thought Given

Having devoted some considerable thought to the pronunciation of the M-Set prior to the survey was hypothesized to lead to a higher preference for the 1b – [mifaɪ(t)] variants among the respondents, as this might have been reflective of awareness of the debate and of others’ stances, through prior discussions read or heard about through MetaFilter and related activities. However, there were no significant correlations between any category of thought given to the pronunciation of the M-Set and ‘MeFi’ pronunciations in either the Survey or Panel Data sets. This is not to say that awareness of the debate, others’ stances, or involvement with other MetaFilter-related activities does not play a role in the pronunciation of the M-Set. It seems likely that the amount of thought given to the M-Set does play a role in pronunciation outcomes, but that it shows no bias toward any particular variant. Other measures of metalinguistic awareness would suggest that this is the case, as it is not necessarily the amount of thought given that influences participants towards certain outcomes, but rather the type of metalinguistic thought that is attended to that achieves this effect.

5.3.3 Other Measures of Metalinguistic Awareness

Many participants gave responses in their survey rationales that demonstrated little or no conscious thought about how they made their pronunciation choices. That is, they expressed a lack of metalinguistic awareness about the M-Set or the debate over pronunciation. In these instances, participants often gave statements in their survey rationales such as, “*That’s just how it is in my head!*” or “*It’s what my brain decided.*”

This rationale was not given by participants from any one pronunciation group significantly more than any other. This is not a surprising result, given that the measure concerns a *lack* of conscious thought given to pronunciation. However, what is interesting to note here is the sheer number of respondents that gave this type of response. In the 2010

survey, 10% of respondents who gave rationales stated some variant of a statement saying that they had arrived at their pronunciation based on their brain resolving any ambiguity without interference from deliberate thought about the matter; 9% responded with this type of rationale in the 2012 survey. This result suggests that there is a sizeable portion of the community who had arrived at a pronunciation choice without an explicit reason for doing so (that they may have been consciously aware of at the time). It is likely that the actual percentages are even higher, but this measure can only account for those who had overtly stated something to this effect.

5.3.3.1 *Metalinguistic Awareness: “don’t know.”*

This category included all survey rationales in which MeFites stated that they “*didn’t know*”, “*had no idea*”, “*no clue*” and similar responses regarding their pronunciation of ‘MeFi’. The category is taken to be an indicator of a complete lack of metalinguistic awareness or thought given as to why a pronunciation choice was made.

These responses comprised 5% of 2010 survey rationales and 6% of 2012 rationales. This rationale type was not skewed toward any one pronunciation group.

There was also substantial overlap between MeFites giving this rationale and stating some form of “automatic” brain processes, e.g., “*I don’t know. That’s just how it is in my head.*” Therefore, the lack of bias toward any particular variant for this rationale category was also not surprising.

5.3.3.2 *Citing Linguistic Rules*

At the other end of the spectrum, many MeFites demonstrated extensive metalinguistic awareness by citing specific linguistic or grammatical rules they had applied to the M-Set to derive their preferred pronunciation(s).

Grammatical rules or morpho-phonological processes were mentioned in 4% of 2010 survey rationales and 3% of 2012 survey rationales. The mention of linguistic knowledge or awareness was not significantly skewed toward any one pronunciation group, but the rules that were cited and their application differed notably between groups.

The 1b – [mifɑɪ] pronunciation group frequently cited linguistic rules relating to stressed syllables and long vowels, both usually applying to the pronunciation of the first syllable of ‘MeFi’. The 3a – [mɛfi] and 3b – [mɛfɑɪ] group often cited a long vowel rule, but almost always pertaining to the last syllable in ‘MeFi’. For the first syllable, the [mɛ-] groups often referred to abbreviation processes, citing the full form ‘MetaFilter’ as their guide to the pronunciation of the truncated form (usually with an exception for the ‘Fi’ portion of the abbreviation).

This shows that while metalinguistic awareness was helpful, speakers were selective

about how they applied that knowledge. Additionally, many survey respondents demonstrated awareness of inconsistencies or selective application of rules, adding further linguistic justifications for those exceptions.

This is not to say that participants were being unjustly selective, but rather that they were reconciling both inconsistencies in grapheme-phoneme correspondences (GPC), as well as rules that they have learned to deal with such inconsistencies. Participants went as far as they felt was sufficient or necessary to resolve any potential pronunciation issues they recognized. For a review of some of the common GPC rules and inconsistencies in English dialects, please refer to **3.5.3 Grapheme-Phoneme Correspondence (GPC) Rules, p. 76.**

Other common linguistic reasons mentioned concerned the perception of the first syllable of ‘MeFi’ as a closed syllable (i.e., having a checked vowel) or as an open syllable (i.e., having an unchecked vowel). Those who perceived the first syllable to be open tended to opt for the long vowel pronunciations. Conversely, those who perceived the first syllable to be closed (i.e., mVf.V) tended to prefer the lax vowel [ɛ] in ‘MeFi’. For the 1b – [mifaɪ] group, this often aligned with their visual perception of the form as to be made of two parts, separated by the capital ‘F’ in ‘MeFi’ (to be covered later this section, **5.3.3.5 ‘MeFi’ as Made of “Two Parts”, p. 127**). For the 3 – [mɛ-] pronunciation groups, this aligned with their perception of ‘MeFi’ as its own word, but one that is phonetically similar to its parent form ‘MetaFilter’. For a more complete description of syllabification issues, as they pertain to the ‘M-Set’, please refer to **3.5.6 Syllabification, p. 86.**

5.3.3.3 *The General Appearance of ‘MeFi’*

A common rationale that participants gave concerned how the form ‘MeFi’ appeared to them. The most common expressions of this were variations of the phrase “*That’s just how it looks to me.*” Other variations included descriptions of the appearance of the letters, mentions of capitalization or general appearance of the form.

In 2010, 291 (15%) of survey rationales stated a feature of the letters, capitalization or overall shape of the word as an influence on their pronunciation of ‘MeFi’. In 2012, 195 (13%) of rationales mentioned a feature of how ‘MeFi’ appeared to them as an influence in making their choice. This result was overwhelmingly skewed towards MeFites who preferred the 1b – [mifaɪ] pronunciation variant (82% (2010) and 84% (2012) of respondents with rationales referring to how the form looked to them in the 2010 and 2012 surveys, respectively). The 3 – [mɛ-] variants represented 14% (2010) and 11% (2012) of this overall distribution. The remaining variants accounted for less than 5% of the remaining citations.

5.3.3.4 *CamelCase*

A subset of the rationales concerning appearance included MeFites who specified the presence of CamelCase as a specific influence in their pronunciation. Mentions of CamelCase (or any other expression referring to capitalization) accounted for 4% of all the survey rationales in 2010 and 3% in 2012. Citing CamelCase as a factor influencing pronunciation choice was even more heavily skewed toward the 1b – [mifaɪ] variant than general mentions of how ‘MeFi’ looked — 92% (2010) and 90% (2012) of those who mentioned CamelCase chose this pronunciation.

It is clear from this result that there was a strong correlation between those who attended to the visual form of ‘MeFi’ as a cue to pronunciation and those who felt that 1b – [mifaɪ] matched that cue. However, for those preferring other pronunciation variants, cues along different dimensions took precedence.

5.3.3.5 *‘MeFi’ as Made of “Two Parts”*

Similar to the general appearance and CamelCase categories, many participants stated that they perceived ‘MeFi’ as to be made of two separate parts or words. The “*two parts*” rationale category was not a subset of the general appearance category (although there was substantial overlap) because many MeFites mentioned the separation of the two syllables without explicitly stating that it was a feature of the visual representation of the form. These instances were coded as belonging to the “*two parts*” category, but not in the appearance category, unless an additional rationale stated this as such.

For both survey years, 10% of surveys with rationales included a statement referring to ‘MeFi’ as being made of two parts. Similar to the results for CamelCase and appearance, the “*two parts*” category was overwhelmingly skewed toward the 1b – [mifaɪ] pronunciation group. MeFites who preferred this variant and stated that they perceived ‘MeFi’ as being made of two parts accounted for 89% of “*two parts*” rationales in the 2010 survey and 95% of “*two parts*” rationales in the 2012 survey. The 6% increase in percentage points from 2010 to 2012 for MeFites who preferred the 1b – [mifaɪ] pronunciation and gave this rationale was not significant. However, that this rationale was mentioned in almost 10% of all survey rationales predominantly by the 1b – [mifaɪ] group showed that, along with the other appearance-based rationales, this reason was one of the more common justifications given for a 1b – [mifaɪ] pronunciation.

5.3.4 **Analogies in Survey Rationales**

Also frequently employed for justifying pronunciation choices was the use of analogies, where MeFites borrowed from orthographically, semantically or otherwise similar forms

they had associated with the M-Set to arrive at a preferred pronunciation. This category included associations MeFites have with the words and sounds that make up variants of the M-Set. For example, many recognized the word ‘me’ in [mifɑɪ] and felt positive associations accompanying the semantic link between the word ‘me’ and its appearance in ‘MeFi’ (e.g., that ‘MeFi’ is a community about them and their interests).

In this subsection, the most frequent lexical-semantic associations will be reviewed. The six lexical items mentioned the most in survey rationales were ‘me’, ‘meh’, ‘meta’ (on its own, or as in ‘MetaFilter’), ‘my’, ‘HiFi’ and ‘WiFi’. Some of these words were biased toward a certain pronunciation variant — e.g., ‘me’ was categorically associated with the 1b – [mifɑɪ] variant. ‘HiFi’ and ‘WiFi’ were mostly mentioned to justify preferences for the final syllable of ‘MeFi’, i.e., ‘Fi’. These latter two forms were also used to explain pronunciation for some first-syllable choices as well (e.g., 4b – [maɪfɑɪ] rhyming with ‘HiFi’ or ‘WiFi’).

5.3.4.1 *Like ‘me’*

The pronoun ‘me’ appears in ‘MeFi’ and has a natural semantic connection to ‘MeFi’ for many MeFites who feel that the site is about them and their interests. For the 1 – [mi-] pronunciation groups, this association overrode the desire to maintain phonetic consistency between pronunciations of ‘MeFi’ and ‘MetaFilter’; this was frequently explicitly stated in rationales.

This justification for ‘MeFi’ pronunciation was categorically preferred by the 1 – [mi-] pronunciation groups and was mentioned in 22% of 2010 surveys with rationales and 29% of 2012 surveys with rationales. The 7% increase in citing this reason in the 2012 survey (up to 29% from 22% in 2010) was highly significant: $\chi^2 = 25.383$, $df = 1$, $p < 0.0001$. That is, there was a significantly greater number of MeFites in 2012 than in 2010 who preferred a [mi-] pronunciation and stated a semantic or lexical connection between ‘me’ and ‘MeFi’. This is interesting to note, given that there were no significant increases in citing appearance, CamelCase or seeing ‘MeFi’ made up of two parts for the [mi-] pronunciation groups between survey years. From this, it can be inferred that the significant increase observed here was due to a more positive evaluation of the semantic or lexical connection than previously and was not necessarily indicative of an increased awareness or positive association with the appearance of ‘me’ in MeFi.

When other pronunciation groups mentioned ‘me’ in survey rationales, they disassociated themselves from it by stating that a pronunciation like the word ‘me’ did not make sense with the full form, ‘MetaFilter’. Even though these mentions were very infrequent (only 15 instances of dispreferring the word or pronoun ‘me’ were reported across both surveys), they show how disassociation from ‘me’ for the other pronunciation groups was not

necessarily based on the semantics of ‘me’, but rather pronunciation continuity with the full form ‘MetaFilter’. This is contrasted with the [mi-] groups, where the ‘me’ association was based on lexical semantics, to varying degrees (i.e., some mentioned positive or negative³⁵ associations with the semantics of ‘me’ explicitly, while others simply recognized the pronoun ‘me’ embedded in the form ‘MeFi’).

5.3.4.2 Like ‘meh’

The word ‘meh’³⁶, popularized by the television show *The Simpsons* and meaning “expressing a lack of interest or enthusiasm” (OED Online, 2013), was frequently cited in rationales. ‘Meh’ was positively referenced by those who preferred a 3 – [mɛ-] and 2 – [meɪ-] pronunciation and negatively referenced by those who preferred a 1 – [mi-] pronunciation.

Survey rationales using the word ‘meh’ to explain or justify that participant’s pronunciation choice were categorically used by the [mɛ-] and [meɪ-] pronunciation groups and accounted for 2% (2010) and 3% (2012) of all surveys with rationales. Similar to references to ‘me’, the justification of ‘meh’ for these non-[mi-] groups was largely based on phonological grounds (e.g., phonetic consistency with ‘MetaFilter’) and not primarily on connections between the semantics of ‘meh’ and possible attributes of MetaFilter.

Mentions of ‘meh’ by the [mi-] group mainly referred to negative semantic associations, justifying their dispreference for those phonetic forms of the M-Set. The most commonly cited negative associations included ‘meh’ as indexing “weakness” or a “lack of interest”.

5.3.4.3 Like ‘meta’

Although the predominant pronunciation of the word ‘meta’ is [mɛrə] or [mɛtə], other pronunciations exist and can even be more acceptable than [mɛrə] or [mɛtə] in certain contexts or dialects of English.³⁷ This leads to some ambiguity in the pronunciation of ‘meta’ in ‘MetaFilter’, but it is mostly accepted and understood within the community as [mɛrə] or [mɛtə]. Most survey rationales that cited the word ‘meta’ as an influence in their pronunciation of ‘MeFi’ explicitly referred to ‘meta’ from ‘MetaFilter’. However, some rationales referred to other uses and pronunciations of ‘meta’.

The word ‘meta’, as in ‘MetaFilter’ or otherwise, was almost categorically referenced by the 3 – [mɛ-] or 2 – [meɪ-] pronunciation groups and comprised 13% of 2010 surveys

35. Some participants who referred to the word ‘me’ in their survey rationales did not have positive attitudes toward the semantic associations of ‘me’ (seeing it as selfish-sounding or egotistical). Yet others took pride in these negative evaluations, as covertly prestigious stances.

36. Usually pronounced [mɛ], but sometimes also [meɪ].

37. The river Meta, located in eastern Colombia and western Venezuela, can be pronounced [mitə]; speakers of some English dialects would pronounce ‘meta’ as [mɛrtə] (see 3.5.3.1 **The Pronunciation of <e> in English**, p. 76 for further explanation).

with rationales and 14% of 2012 surveys with rationales. In almost all instances where ‘meta’ was positively referred to by the 1 – [mi-] pronunciation groups it was mentioned that a pronunciation of ‘MeFi’ like ‘meta’ in ‘MetaFilter’ would make more sense and/or be consistent. Many of these participants also stated that they were not exclusive about their 1 – [mi-] pronunciation choice and might or already did use other forms.

5.3.4.4 *Like ‘my’*

More than half of the 39 participants across both survey years who chose the 4b – [maɪfɑɪ] pronunciation made reference to the word ‘my’. Several of them further associated the semantics of ‘my’ as part of their choice, as in a favorable possession of the community (or belonging to it).

5.3.4.5 *Like ‘HiFi’ or ‘WiFi’*

The word ‘HiFi’ was mentioned 378 times (19%) in 2010 survey rationales and 307 times (21%) in 2012 survey rationales. This was quite surprising, as this was approximately twice the number of mentions of ‘WiFi’ (i.e., ‘WiFi’ was mentioned in 9% (2010) and 13% (2012) of all surveys with rationales) and not a particularly relevant or frequent term in general English today as compared to ‘WiFi’.³⁸ This was also a very commonly-cited rationale overall, as compared to other rationales.

‘HiFi’ was predominantly mentioned by the 1b – [mɪfɑɪ] pronunciation group in survey rationales. For these participants, the most common explanation for this mention included analogizing from ‘HiFi’, but replacing the first syllable with the pronoun ‘me’.

Like the 1b – [mɪfɑɪ] pronunciation group, the 3b – [mɛfɑɪ] pronunciation group analogized with ‘HiFi’ for the final syllable only, but these participants often stated that they preferred to replace the first syllable of ‘MeFi’ to match [mɛ-] in ‘MetaFilter’. This is interesting, given that this pronunciation group also explicitly did not mention that they perceived ‘MeFi’ as to be composed of two separate parts, even though their justification for their pronunciation choice presupposed such a distinction.

The three mentions of ‘HiFi’ from the 4b – [maɪfɑɪ] pronunciation group (across both survey years) stated rhyming with ‘HiFi’ for both syllables, even though the spelling differs (which was only explicitly mentioned once).

The remaining mentions of ‘HiFi’ from other pronunciation groups either disassociated from the form, stating that they did not pronounce ‘MeFi’ like ‘HiFi’, or only emulated the final syllable (and didn’t provide further explanation).

38. ‘WiFi’ and allographic variants were two to ten times more frequent than ‘HiFi’ and variants in the Corpus of Contemporary American English (COCA).

Analogies with ‘WiFi’ followed a very similar pattern to those for ‘HiFi’. Mentions of ‘WiFi’ in survey rationales usually accompanied mentions of ‘HiFi’, with ‘HiFi’ often mentioned first. ‘LoFi’ was also mentioned, but very seldom.

As stated previously, justifications of ‘MeFi’ as analogous to ‘WiFi’ patterned very similarly to those for ‘HiFi’. However, ‘WiFi’ was mentioned much less frequently than ‘HiFi’ overall. The reason for this disparity is unclear, but it is possible that the semantics of ‘HiFi’ may have a stronger connection to ‘MeFi’ than ‘WiFi’, in that ‘HiFi’ relates to the high quality of the signal or communicative message, whereas ‘WiFi’ is relegated to the method in which the message is delivered (i.e., wirelessly). Additionally, according to the Oxford English Dictionary (OED Online, 2014), ‘HiFi’ was first attested in 1950, whereas ‘WiFi’ was attested in 1999, but only became prevalent in common vernacular within the last decade.

5.3.4.6 *Like ‘MeFite’*

Across both surveys, 4% of participants who gave survey rationales stated that their pronunciation of ‘MeFi’ was derived from their pronunciation of ‘MeFite’. ‘MeFite’ as a driver of ‘MeFi’ pronunciation was not significantly biased toward any pronunciation variants of the M-Set.

Generally, ‘MeFi’ is the main focus of the debate and the more commonly-used and frequently-heard form. However, it is not surprising that ‘MeFite’ might influence or drive pronunciation choice for ‘MeFi’ for a certain number of participants, as ‘MeFite’ is less ambiguous with respect to its second syllable. Additionally, other aspects of the MetaFilter community, such as spin-off sites like MeFight Club, might influence some participants to pay more attention to ‘MeFite’ than ‘MeFi’.

5.3.4.7 *Other Analogies in Survey Rationales*

Other analogous words mentioned by the various pronunciation groups included but were not limited to: ‘Fee Fi Fo Fum’ (i.e., the first line from an English fairytale), ‘met’, ‘meet’, ‘Semper Fi’ (i.e., the shortened form of ‘Semper Fidelis’, meaning “Always Faithful”, and the motto of the US Marine Corps), and ‘semi’. These words and phrases were infrequent and were often mentioned in conjunction with the other forms discussed above. Forms used by MeFites as analogies also demonstrated a wide range of orthographic and allographic variation, in addition to many of them demonstrating their own pronunciation variations (e.g., ‘semi’ as [semɑɪ] or [semi], depending on context and style).

5.4 Demographic Factors

Analysis of demographic measures provides valuable information about underlying

factors that might bias certain groups of individuals towards particular variants. The demographic measures analyzed in this study were language background and experience, country of residence (geography), age, and gender. It was hypothesized that linguistic history and geography would both be strongly correlated with certain pronunciation variants, as cultures differ with respect to grammatical rules and attitudes about language and its use. Age was also hypothesized to be correlated with M-Set pronunciation, as the connections between age and sociolinguistic variation and change have been well established in many previous studies (Labov, 2001; Boberg, 2004; Bailey, 2008). However, it was possible that the outcomes of this study might differ from traditional outcomes with respect to age, as the speech community under investigation here interacts primarily through text-based communication and so the mechanisms for the spread of innovation are fundamentally altered. Gender was hypothesized not to be significantly correlated with M-Set pronunciation, given the long history of gender discussions and the strong cultural values of gender equality on MetaFilter. Additionally, preliminary research of discussions on MetaFilter about the M-Set revealed that evaluations of the variants relating to gender were rare and therefore participants were neither propagating those evaluations nor aligning their own gender identities with gendered perceptions of the variants.

5.4.1 Language Background and Experience

Measures of language background and experience included comments from survey rationales mentioning linguistic history or experience and answers to two survey questions pertaining to the participants' native languages and any other languages that they had spoken, studied or were otherwise familiar with.

5.4.1.1 *Linguistic Variety Mentioned in Survey Rationales*

The influence of a linguistic variety was commonly mentioned in rationales. Participants cited their native languages, languages they had learned or studied, or regional influences such as dialect(s) they speak or are familiar with.

Making up 5% of 2010 and 4% of 2012 survey rationales, MeFites often stated their experience with other languages or dialects as an influence on their pronunciation choice (see **Example 1, A-C**). This was heavily skewed toward the **3a** – [mɛfi] pronunciation group, comprising 41% (2010) and 44% (2012) of this rationale category. This was compared to the **1b** – [mifaɪ] pronunciation group, in which linguistic varieties were mentioned 36% (2010) and 15% (2012) of the time.

Example 1. Rationales Citing Linguistic Variety as an Influence on Pronunciation Choice

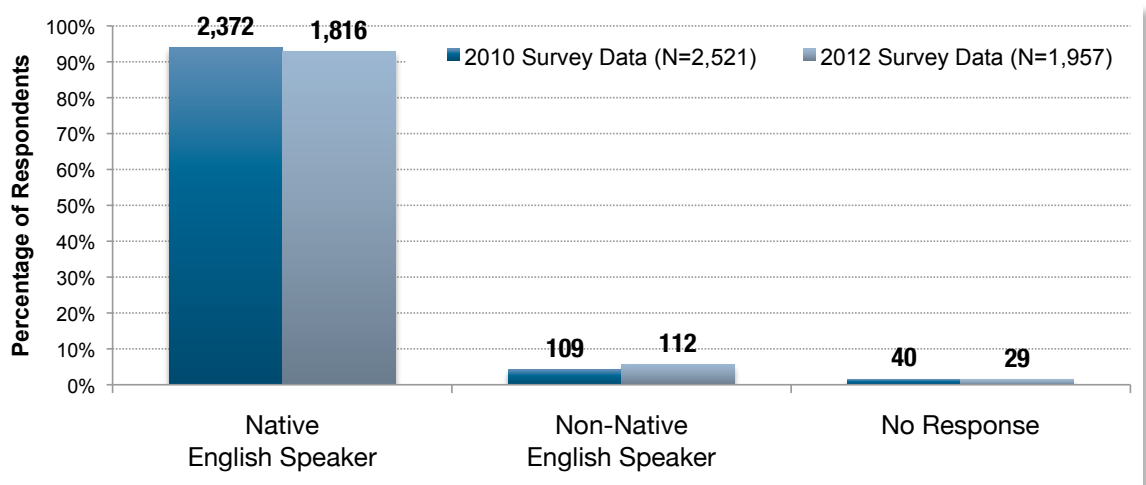
- A. As a native English speaker, I see the words Me and Fi, as in Hi-Fi. Thus the pronunciation.
- B. As a native Spanish speaker, I have always “pronounced” the website name in my mind as “meh-tah-fee-l-tehr”, so in my mind I also pronounce the word MeFi as “meh-fee”. I believe that had I been a native English speaker I would probably pronounce it “may-fee”, which is the way I might be inclined to pronounce it if I ever get together with mefiters.
- C. English is my second language. The pronunciation I’ve chosen is probably due to my Norwegian accent.

These qualitative data examples showcase several ways in which participants’ linguistic background might influence pronunciation outcomes. Participants’ levels of awareness about linguistic varieties as an influence on their pronunciation choices varied. Some MeFites asserted their language background as a possible influence in the form of a guess or suggestion, while others definitively made a case for their native language or dialect as a factor in their choice with examples and additional supporting information.

5.4.1.2 English as a Native Language

As MetaFilter is a site accessible to people from around the globe, it was expected that a percentage of these participants would be not be native speakers of English. Until this survey, however, it was unknown how much of the userbase could be accurately estimated to be non-native speakers.

Figure 25. English as a Native Language — 2010 and 2012 Survey Data



$\chi^2 = 4.567, df = 1, p = 0.0326$

Data Results

Of the 2,521 surveys collected in 2010, 2,372 (94%) respondents stated that they were native speakers of English, whereas 1,816 (93%) of the 1,957 survey respondents in the 2012 survey stated they were native speakers of English. This increase from 2010 to 2012 in non-native English-speaking participants was significant ($\chi^2 = 4.567$, $df = 1$, $p = 0.0326$). The Panel Data participants did not significantly differ from the Survey Data with respect to being a native speaker of English, for either survey year.

Native English and Non-Native English speakers did not significantly differ in their pronunciation choices for ‘MeFi’ in either survey year. This suggests that, while the population of participants changed over time with respect to English-speaking status, this was not a significant factor influencing the choice of M-Set variant. This was also despite the fact that participants who preferred the 3a – [mɛfi] variant attributed non-native English speaking status as a rationale for their pronunciation choice more than any other group did. This can be interpreted to mean that while the 3a pronunciation group was more likely to claim non-native English speaking as a factor in their decision-making, the participants’ native language did not actually significantly influence decision-making in this case; equivalent numbers of non-native English speakers preferred other variants (and did not state that their linguistic history was a factor to the same degree that those who preferred 3a variants did).

5.4.1.3 Other Language Experience

In 2010, 898 (36%) respondents stated that they had experience with studying, learning or speaking a language other than English. In 2012, the format of the language experience questions was improved and restructured, resulting in 1,395 (71%) of the survey respondents sharing their experiences with languages other than English. The most frequently reported non-English languages studied or learned are listed in Table 13.

Table 13. Self-Reported Language Experience from the 2012 Survey

Rank	Fluent Level:	Advanced Level:	Intermediate Level:	Beginner Level:
1	French (49)	French (105)	French (307)	French (54)
2	Spanish (38)	Spanish (71)	Spanish (225)	Spanish (48)
3	German (15)	German (36)	German (105)	German (33)
4	Japanese (8)	Japanese (14)	Latin (46)	Russian (23)
5	Italian (7)	Italian (12)	Japanese (30)	Italian (19)
6	Chinese – All varieties (6)	Chinese – All varieties (10)	Chinese – All varieties (23)	Japanese (19)
7	Dutch (5)	Latin (10)	Italian (23)	Greek (17)
8	Portuguese (5)	Portuguese (7)	Greek (21)	Chinese – Any variety (16)
9	Russian (5)	Swedish (7)	Russian (17)	Latin (14)
10	Danish (5)	Dutch (6)	Dutch (8)	Hebrew (10)

Note: Many respondents listed more than one language; therefore, the total responses far exceed the total number of survey respondents.

The 2012 survey showed that MeFites have the most experience with French, closely followed by Spanish and German. Over 150 unique languages were reported across all proficiency categories. Varieties reported included rare and endangered languages, computer programming languages, various constructed languages ('ConLangs', e.g., Klingon), and more abstract or possibly non-serious interpretations of 'language', such as 'Love', 'Internet', and 'MetaFilter'.

As previously covered in **4.5.1.1 Changes Made in the 2012 MetaFilter Survey, p. 104**, the language experience question differed between the two surveys, making statistical comparisons over time between these results unfeasible. However, some results pertaining to participants' country of residence and language experience are worth noting.

In 2010, there was a significant difference in self-reported language experience by country, with Canadians stating that they speak another language proficiently³⁹ at least 15% more than residents of the US, the UK, and Australia: $\chi^2 = 22.693$, $df = 3$, $p < 0.0001$. This is likely owing to the influence of French in Canada, which is a native language of over 22% of the population and spoken by over 30% of the population, according to the Canadian census (Corbeil, 2012). French is also recognized as one of Canada's two official languages (Boberg, 2010; Corbeil, 2012).

These differing orientations to cultural values involving linguistic diversity and language policy between the US and Canada are likely to influence the linguistic choices of individuals. While Canada has less overall English dialect diversity than the US, French (as well as British) linguistic and cultural influences are much more present in Canada (Boberg, 2008a).

In 2012, with the language experience survey question changed and respondents reporting up to three languages they had any experience with (and rating them at four levels of proficiency), results showed that Canada still had a slightly higher response to language experience (at least 4% more respondents reported other language experience than respondents from the other countries), but this difference was not significant.

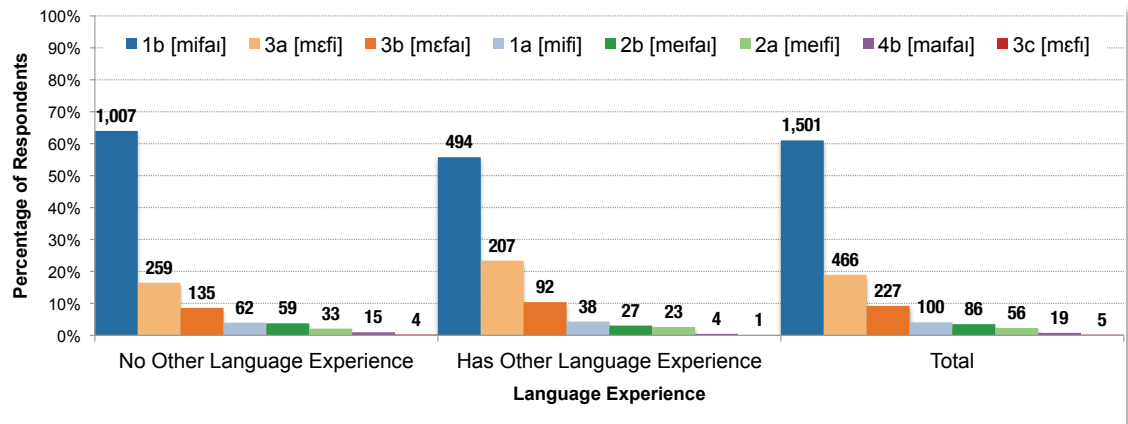
5.4.1.4 *'MeFi' Pronunciation and Non-English Language Experience*

This metric assessed whether there were any correlations between 'MeFi' pronunciation and any experience with languages other than English.

39. Self-reported language 'proficiency' is fairly subjective and led to the reframing and restructuring of the language experience question in the 2012 survey.

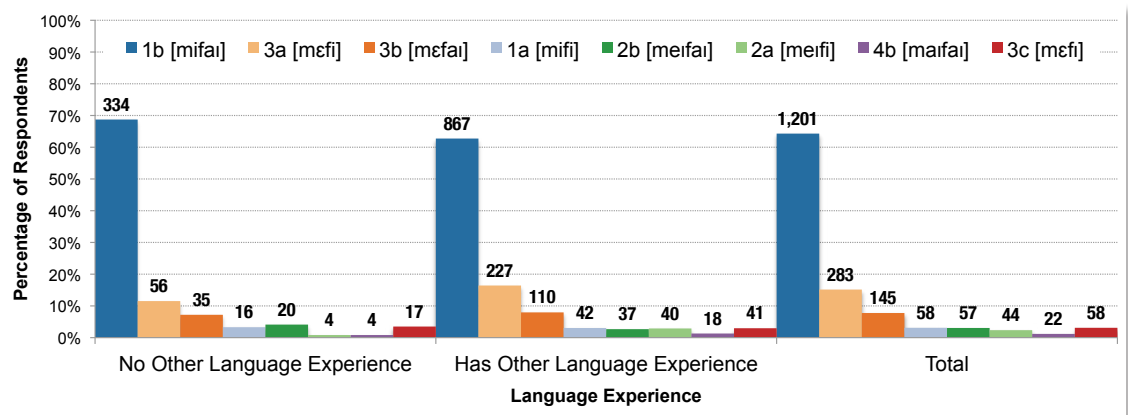
Data Results

Figure 26. 'MeFi' Pronunciation by Language Experience – 2010 Survey Data, N=2,460



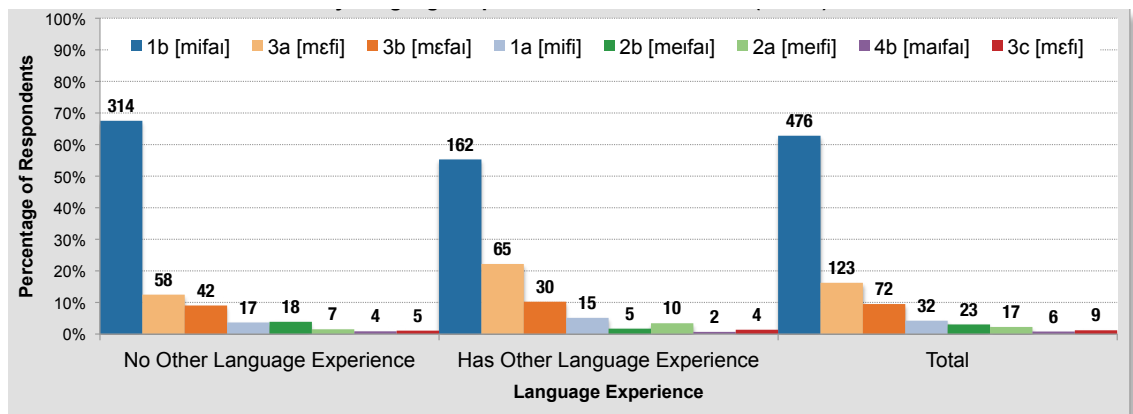
$\chi^2 = 26.559, df = 7, p\text{-value} = 0.0004$

Figure 27. 'MeFi' Pronunciation by Language Experience – 2012 Survey Data, N=1,868



$\chi^2 = 18.605, df = 7, p\text{-value} = 0.0117$

Figure 28. 'MeFi' Pronunciation by Language Experience – 2010 Panel Data, N=758



$\chi^2 = 21.81, df = 7, p\text{-value} = 0.0027$

While native language was not correlated with pronunciation choice in any way (5.4.1.2 English as a Native Language, p. 133), other (non-English) language experience was significantly correlated in both survey years and data sets, excluding the 2012 Panel Data (not

shown). The 2012 measure of language experience was more accurate than the 2010 measure, in that it allowed participants to specify three languages and four proficiency levels that they had experience with. This is compared to the 2010 measure, in which participants stated whether or not they had experience with other (non-English) languages, but it was left to the participant to determine what the definition of experience was. This may lend some insight to the disparity between the 2010 and the 2012 Panel Data results, with additional explanation given in 7.5 **Research Hypotheses and Inconsistent Outcomes**, p. 211.

5.4.2 Geography

The measures for geography were based on participants' self-reported country of residence at the time of the survey. While additional geographic information would have provided a more complete picture of participants' geographic background, this information was not included in the data collection. This was done to keep the surveys brief, as well as to keep possibly intrusive questions to a minimum.

The 2,521 survey respondents in 2010 represented at least 49 self-reported countries of residence; 34 survey participants did not state a current country of residence. The 15 most represented countries in the 2010 survey were as follows (number of survey participants from each country in parentheses): the United States (1,862), Canada (231), the United Kingdom (162), Australia (55), Japan (16), New Zealand (16), Germany (15), France (13), Netherlands (12), Ireland (9), Mexico (9), Sweden (9), Italy (6), Belgium (5) and China (5).

The 1,957 survey respondents in 2012 represented at least 47 self-reported countries of residence.⁴⁰ The 15 most represented countries in the 2012 survey were as follows: the United States (1,434), Canada (166), England (119), Australia (54), Netherlands (19), Germany (14), Scotland (12), Ireland (10), New Zealand (10), Japan (8), Belgium (6), Denmark (5), Finland (5), France (5) and Switzerland (5).⁴¹

US State data was accurately captured in the 2012 survey. The 15 most represented US States were as follows: California (216), New York (148), Massachusetts (105), Oregon (74), Illinois (71), Washington (67), Texas (61), Pennsylvania (54), Virginia (54), North Carolina (44), Georgia (31), Maryland (31), Michigan (31), Washington DC (30) and Minnesota (30).

40. Thirty survey participants did not state a current country of residence. Of these, twelve selected 'Not a US resident' in the previous question but declined to state their country of residence. Therefore, it is possible that additional countries are represented in the data, but these were not shared with the researcher.

41. The country of residence question was a free-form fill-in in 2010; in 2012 the question was modified to a drop-down menu, with the countries that make up the UK listed separately.

Figure 29. 2010 World Map of Survey Respondents' Country of Residence

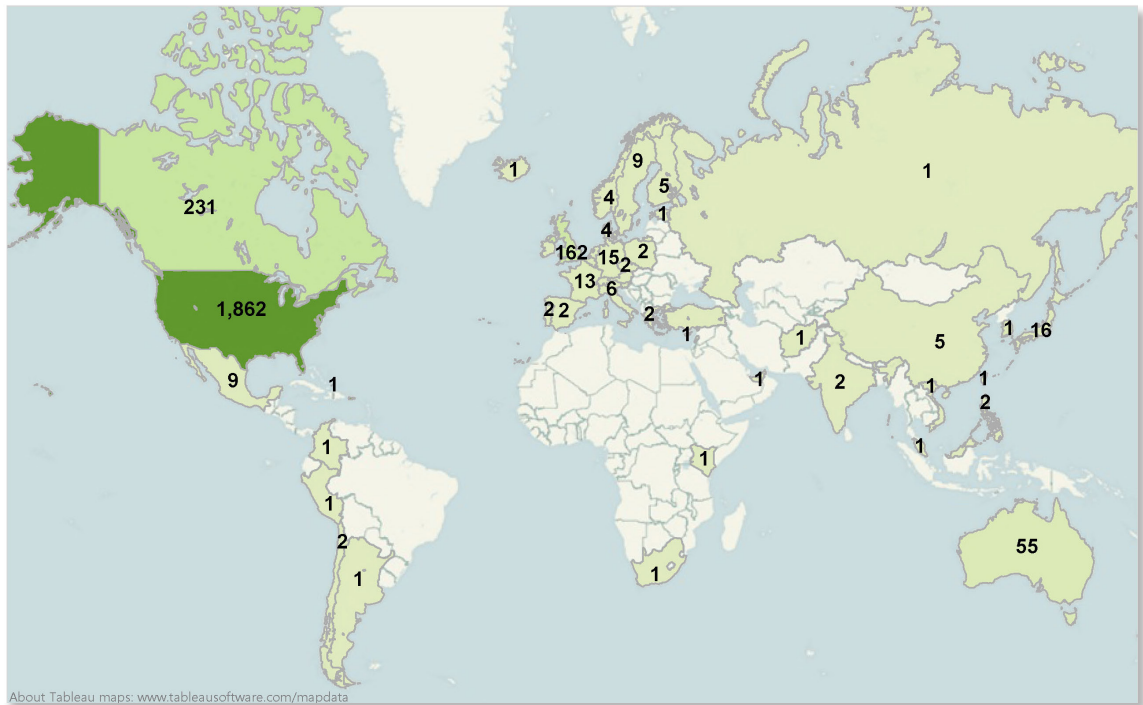
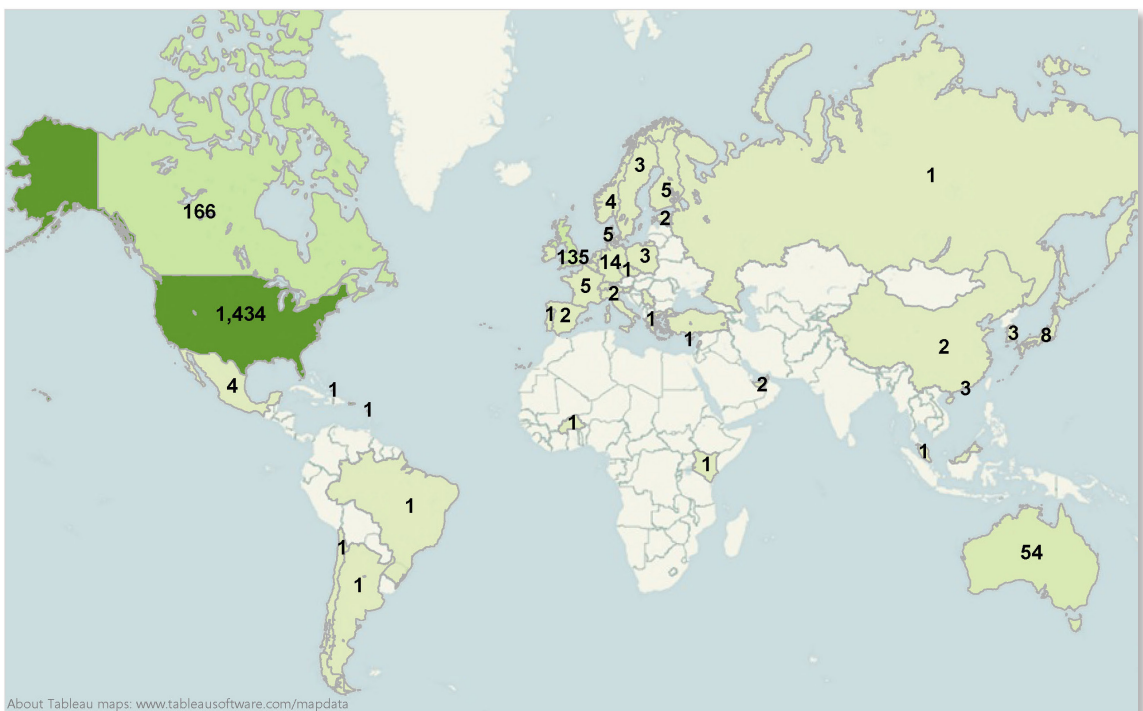


Figure 30. 2012 World Map of Survey Respondents' Country of Residence



5.4.2.1 'MeFi' Pronunciation and Geography

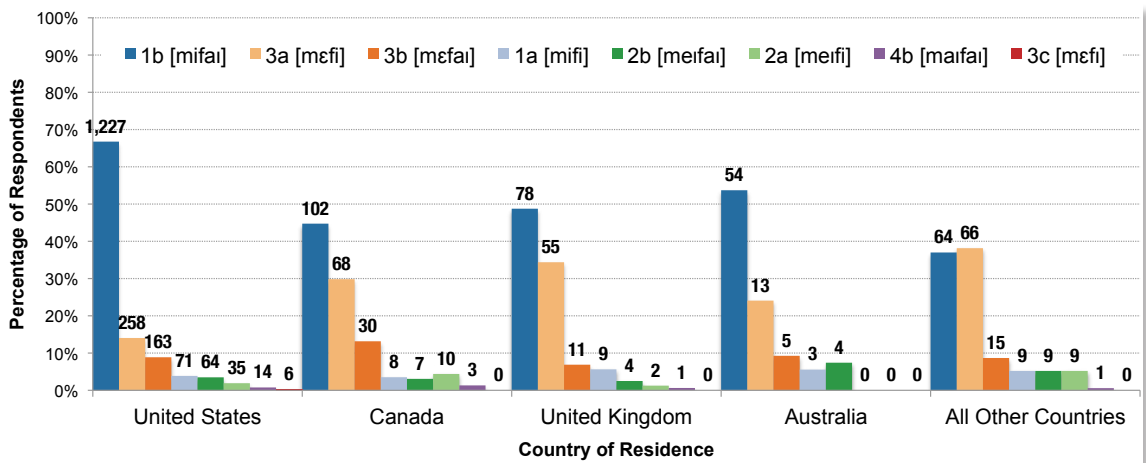
Geographic data were sorted into five categories based on self-reported country of residence: United States, Canada, United Kingdom, Australia and All Other Countries. Ambiguous and blank responses were labelled as 'NA' and not included in analysis. There were no significant differences in the distributions of participants' countries of residence in

Data Results

survey years for both the Survey and Panel Data sets.

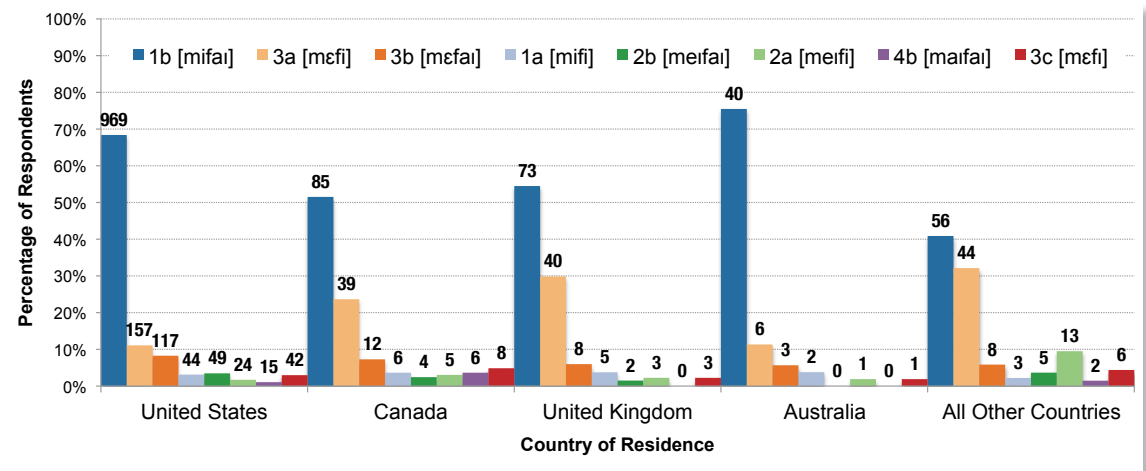
The most MetaFilter-populated country was the United States, representing 75% of the Survey Data in 2010 and in 2012 and shown in **Figure 31** and **Figure 32**. While this distribution is heavily skewed, it is generally reflective of the actual geographic distribution of the MetaFilter userbase. Another potential issue to note is that each country was treated as a uniform entity, ignoring any intra-regional variation, as not enough data were collected for this level of detail. This is particularly problematic in the UK data, where there is considerable dialect variation across several national territories, each having unique cultures and language ideologies (Kortmann and Upton, 2008; Llamas, 2010). However, these broad geographic divisions provided a good starting point for analyzing variation in the pronunciation of the M-Set.

Figure 31. 'MeFi' Pronunciation by Country of Residence – 2010 Survey Data, N=2,453



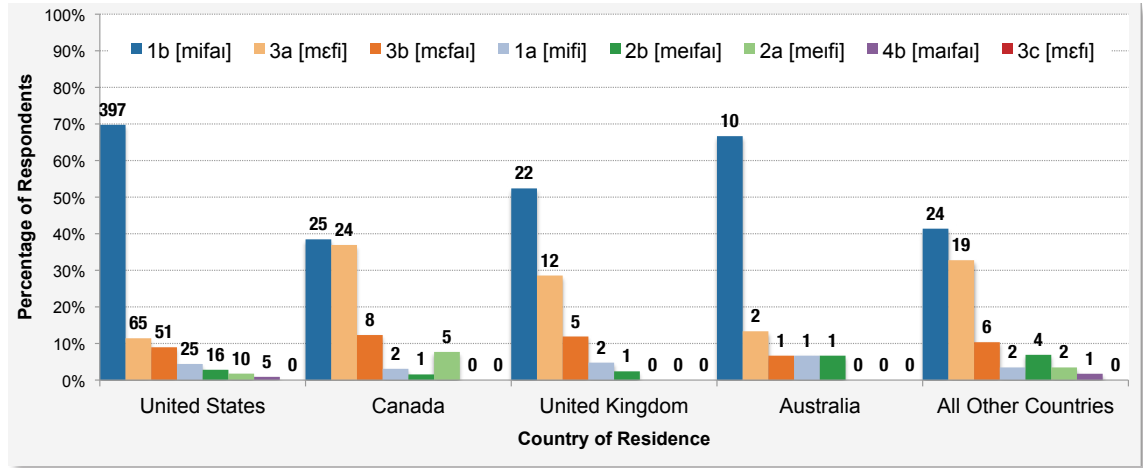
$$\chi^2 = 141.1183, df = 8, p\text{-value} < 0.0001$$

Figure 32. 'MeFi' Pronunciation by Country of Residence – 2012 Survey Data, N=1,906



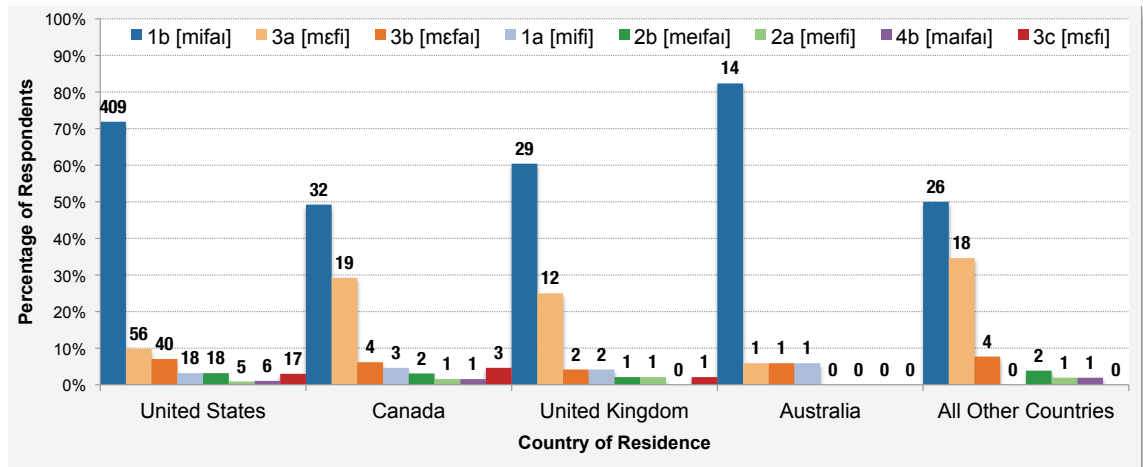
$$\chi^2 = 95.8713, df = 8, p\text{-value} < 0.0001$$

Figure 33. 'MeFi' Pronunciation by Country of Residence – 2010 Panel Data, N=749



$$\chi^2 = 56.045, df = 8, p\text{-value} < 0.0001$$

Figure 34. 'MeFi' Pronunciation by Country of Residence – 2012 Panel Data, N=751



$$\chi^2 = 46.5422, df = 8, p\text{-value} < 0.0001$$

Country of residence was highly correlated with 'MeFi' pronunciation, with Canadian and UK MeFites preferring a significantly greater percentage of 3a – [mɛfi] and 3b – [mɛfɑɪ] variants than US residents did. None of the distributions of 'MeFi' pronunciations for the US, Canada, the UK, or Australia significantly changed between the survey years in the Survey Data.

Of the four countries included in the Survey Data, the US and Canada showed the biggest distributional difference, with 22% less preference for the 1b - [mifai] variant by Canadian MeFites in the 2010 Survey Data and 17% less in 2012. These outcomes may be surprising, given the geographical proximity and the cultural influence the US has on Canada. General Canadian English and General American English are also very similar with respect to the vowels involved in this debate (Trudgill and Hannah 2008, p. 53). Therefore, it is unlikely that the results can be based purely on phonological grounds.

Canada's sociocultural history may play a greater role than the phonology of English in these pronunciation outcomes. The large differences in 'MeFi' pronunciation distributions between the US and Canada may be at least partially explained by differences in cultural histories and sociolinguistic attitudes between the two countries (Boberg, 2000). Canadian MeFites, while being heavily influenced by American culture in general and being in greater social proximity to US-centric topics on MetaFilter, are more linguistically influenced by historical and cultural affiliation with the British and the French than Americans are. Canada's unique linguistic position has been aptly elucidated by Schneider (2008, p. 24) in the following quote: "*Canadian English in general is said to have been characterized by a tension between its British roots (reinforced by loyalists who opted for living in Canada after America's independence) and the continuous linguistic and cultural pressure [...] exerted by its big southern neighbor.*"

Additionally, even though Canada today is dominated by English speakers, the country was originally established by French colonies rather than English ones⁴² (Boberg, 2008a, p. 145), thereby adding another layer of complexity to Canadians' linguistic history. In the century prior to the Treaty of Paris in 1763, there were virtually no English speakers in Canada; in the years afterward English settlement took over, ultimately resulting in the official bilingual status of Canada today. Apart from Quebec and its neighboring regions, English is generally spoken. Exceptions to this exist for the larger cities, where there is much cultural and linguistic diversity — for example, in Toronto, approximately only 59% of the population are English-speaking, 1% French-speaking, and the remainder being speakers of other languages of the world; similar percentages are claimed for Vancouver (Boberg, 2008a, p. 145).

Therefore, the history of influence of Canadian French (and Canadian MeFites' knowledge of French as a first or second language), may have a greater influence on pronunciation outcomes than originally hypothesized, even if bilingualism or experience with other languages was not directly correlated with pronunciation choice. That is, non-English language experience (mainly, French) may be linked to pronunciation choices as a facet of Canadians' national identity, rather than as a standalone factor accounting for pronunciation outcomes in general (and regardless of participants' country of residence).

Differences in sociolinguistic attitudes between American English and Canadian English may also play a role in pronunciation outcomes for 'MeFi'. These attitudes sometimes arise from arguments about whether Canadian English was more or less influenced by American English versus British English, which are very complicated to disentangle, for various

42. With the exception of Newfoundland, which was established in 1583 by English settlers and is the oldest English-speaking colony in North America; Newfoundland remains linguistically distinct from the rest of Canada to this day (Boberg, 2008a, p. 145).

reasons. Scargill (1957, pp. 611-612), as cited in Boberg (2008a, p. 147) explains several of these complexities, as well as some caveats about assumptions on the origins of Canadian English. The most relevant of these rationales to this study is the warning against ascribing key features of Canadian English to American influence, as they just as easily could have arisen from Northern or Western Britons, who comprised the majority of 19th century British immigrants and were magnitudes greater in number than the American settlers. The influences are not mutually exclusive, however, and it is suggested by Boberg (2008a, p. 147) that Canadian English today may have retained some features of colonial American English which are no longer present in American English today (presumably due to different influences on the type and rate of linguistic change over time in America versus Canada). Regardless, Canadian English today “*varies between standard British and American forms on a long list of variables concerning phonemic incidence, morphosyntax, lexicon, and general usage*” (Boberg, 2008a, p. 148). This extends to Canadian spelling, which employs features of both American and British English. More recently and among younger Canadian speakers especially, there has been a trend toward American conventions across all levels of linguistic structure, including spelling (Boberg, 2008a, p. 149). As a result, in some ways, variation in American English is increasing, but in others it is on the decrease — with the main correlates being demographics and education (Kretzschmar, 2008, p. 42). However, some features are uniquely Canadian⁴³, the most pertinent to this study being Canadian Raising, as described in **3.5.4 Phonetic Realizations of the M-Set, p. 79**. These features may remain distinct, and new ones may arise as well.

The overwhelming influence of America on Canadian speech and culture does not signal the beginning of the end for Canadian distinctness. As Boberg (2008a, p. 158) states: “*...the sound of Canadian English will be closely bound up with Canadians’ sense of their national identity for many generations to come.*” In his research on geolinguistic diffusion across the US-Canadian border, Boberg (2000, p. 23) also states, “*In general, it seems safe to say that Canadians do not want to sound like Americans, so that when a variant is marked [+American] rather than, say, [+young] or [+trendy] it will not be readily transferred.*” Therefore, the pronunciation of ‘MeFi’ may provide opportunities for various types of sociolinguistic and cultural attitudes to be asserted, whereby linguistic choices which may be perceived as more American-sounding could be less preferred than those variants that Canadian participants perceive to be more reflective of a Canadian identity. Regardless, this need for individuation — be it along community, political, linguistic, or national lines — will always exist and may find

43. It should be noted that it is the indexical relationship between Canadians and the PRICE/MOUTH raising that is unique identifier, and not the exclusivity of this feature to Canadians, in which it is not unique; Canadian Raising has been observed in several other non-Canadian regions, including but not limited to, Martha’s Vineyard, Massachussets, and Philadelphia (Boberg, 2008a, p. 153).

itself indirectly expressed in subtle but meaningful ways, such as in the selection of a phonetic pattern for a group name.

Turning to matters involving the UK and the US, in the 2010 Survey Data, 49% of UK MeFites preferred the 1b - [mifɑɪ] variant compared to 67% of US MeFites; 54% of UK MeFites preferred the 1b - [mifɑɪ] variant as compared to 68% in 2012. Similar to Canada, the distribution of pronunciation preferences between the UK and the US were highly significant across both survey years and in both the Survey and Panel Data sets.

The UK has much more dialect variation in general than the other countries (Llamas, 2010, pp. 229-230), as well as identifiable pronunciation rules that differ from American speech for the vowels involved in the M-Set, which could result in more variation in pronunciation preferences. For example, in many dialects of England, the phoneme /i/, generally realized as [i] or [i:] in other varieties of English, can also be realized as [ei ~ ti ~ əi] (Beal, 2008, p. 130; Clark, 2008, p. 160; Upton, 2008, pp. 271-272). In Belfast English, /i/ may be commonly realized as an even lower variant, as [ɛ:] or [ɛ] (Hickey, 2008, p. 93). For additional examples of phonetic differences between the UK and the US, please refer to **3.5.4 Phonetic Realizations of the M-Set, p. 79**.

The examples given in the previous paragraph assumed that the underlying preference for the first vowel of the M-Set variables was /i/, with varying surface realizations based on dialectal differences. However, differences in phonotactic distributions between the UK and US dialects could result in different underlying representations of the M-Set between these two geographic areas as well. For example, the rule restricting the phonotactically allowable vowels in unchecked, stressed positions in words for American English speakers (Ladefoged, 2006), may result in a preference toward /i/ as the V1 in 'MeFi' for speakers in the US, whereas the possibility of /i/ or the equally viable /ɛ/ phoneme in the UK. These possibilities are explained in greater detail in **3.5.3 Grapheme-Phoneme Correspondence (GPC) Rules, p. 76** and **3.5.6 Syllabification, p. 86**.

From all of these examples, it is possible to see how the differences in 'MeFi' distributions between the UK and the US may at least be partially accounted for on phonological grounds, at various levels of phonetic processing and production. There may be ideological or other identity-related factors at play as well, but at this point of the investigation, they are much harder to target. Employing a methodology specifically designed for the level of detail required to investigate this further would be worthwhile in future research, and could go toward explaining both the Canadian and UK differences as compared to the US.

The 2010 Survey Data showed that the Australian distribution of pronunciation choices did not significantly differ from the US distribution, but was also no more diverse than

the Canadian or UK distributions (i.e., Australia did not significantly differ from either Canada or the UK). In 2012, Australia showed significant distributional differences in ‘MeFi’ pronunciation from both Canada ($\chi^2 = 6.998$, $df = 1$, $p = 0.0082$) and the UK ($\chi^2 = 7.224$, $df = 1$, $p = 0.0072$). Australians may be therefore demonstrating a change over time with respect to ‘MeFi’ pronunciation, but with such small numbers of participants in either survey years, the outcome is not entirely reliable and also does not represent an overall change over time in the Survey Data.

The explanation for the lack of significant results in 2010, and the differences between Australia and both Canada and the UK in 2012 involve several conflicting influential factors that need to be examined separately. Australia is geographically distant from the countries studied here, but has cultural commonalities with all of them, with its historical roots in British culture, heavy influence from American culture and several other similarities with Canadian culture (demography, historical independence, government and healthcare systems, etc.). Even though Australia’s (and New Zealand’s) geographic isolation has allowed the distinctiveness of those varieties of English to flourish, the amount of regional variation within Australia (and New Zealand) is minimal as compared to other geographic regions (e.g., the British Isles, North America) (Burrige and Kortmann, 2008, p. 24). However, globalization has contributed to greater linguistic diversity in Australia, as accounted for by unprecedented numbers of tourists, refugees and migrants (Burrige and Kortmann, 2008, p. 25). *“It would be surprising therefore, given the global presence of the United States and the inevitable loosening of ties between Britain and its former Antipodean colonies, if there were not some sort of linguistic steamrolling going on”* (Burrige and Kortmann, 2008, p. 26). This linguistic attitude would suggest that the pronunciation of the M-Set might be influenced by these participants’ (possibly implicit) perception of what American MeFites might prefer for the M-Set. However, Australian and New Zealand English speakers are generally negative towards the “Americanization” of their speech variety (Burrige and Kortmann, 2008, p. 26). Taking this into consideration, if *“some sort of linguistic steamrolling”* is occurring in these instances, it is likely happening below these participants’ levels of linguistic awareness, or perhaps at deeper levels of linguistic structure (e.g., shifts in overall sound patterns, as opposed to more noticeable lexical borrowings).

There are several possible and conflicting factors here, but given the small sample size of Australian MeFites, exploring them further is not be feasible nor would their results be conclusive at this time. Again, these results open up possibilities for further study, at finer-grained levels of detail.

Representing the remaining 7% of the geographic data, all other countries were

combined into one category. This yielded a highly significant difference from US and Australian distributions for ‘MeFi’ pronunciations.

The majority of the countries included in the ‘All Other Countries’ category were non-English speaking. Even though native language was not directly correlated with ‘MeFi’ pronunciation, it is very possible that this was a contributing factor toward pronunciation outcomes when divided along geographic lines. That is, there were a significant proportion of non-native English speakers residing in the US, Canada, UK, and Australia whose pronunciation distributions mirrored that of the native English speakers. However, when looking at non-English-speaking countries — where the majority of residents do not speak English as a native language — the ‘MeFi’ distribution differed greatly. This fact, combined with qualitative data from survey rationales where other language varieties were mentioned in justifying preferences, may help explain results showing an overwhelming bias toward 3a - [mɛfi] and 3b - [mɛfaɪ] pronunciations from participants residing in these other countries.

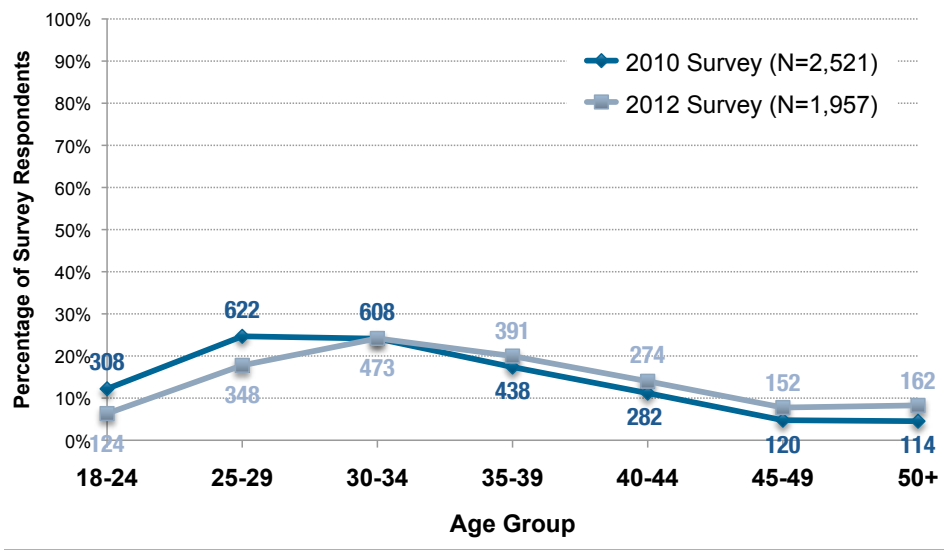
Striking differences between distributions of ‘MeFi’ preferences and participants’ countries of residence illustrate how geography — as a proxy for language background and cultural influence — introduces a notable demographic bias to this pronunciation debate. Demographic factors such as country of residence play an important role in the negotiation of the M-Set in that these factors influence outcomes, regardless of whether participants interacting can easily convey these details to other participants (or perceive this information about other participants) during their interaction online.

5.4.3 Age

The average age of MeFites in the 2010 survey was 33, with a range from 18–79.⁴⁴ The median age was 32. The average age of MeFites in the 2012 survey was 36, with a range from 18–81; the median age was 35. A slight shift in the average age of MeFites in 2012 was expected, as 34% of the 2012 survey participants had taken the 2010 survey and aged two years since that time. However, the average age increased by three years, suggesting a demographic change toward a slightly older userbase.

44. Participants were required to check a box on the online survey consent form stating that they were at least 18 years of age. It is possible that some participants stated that they were 18 or older when in fact they were not, but it is not likely that this occurred to a significant degree or in numbers that would affect the data.

Figure 35. Age Distribution of Survey Participants

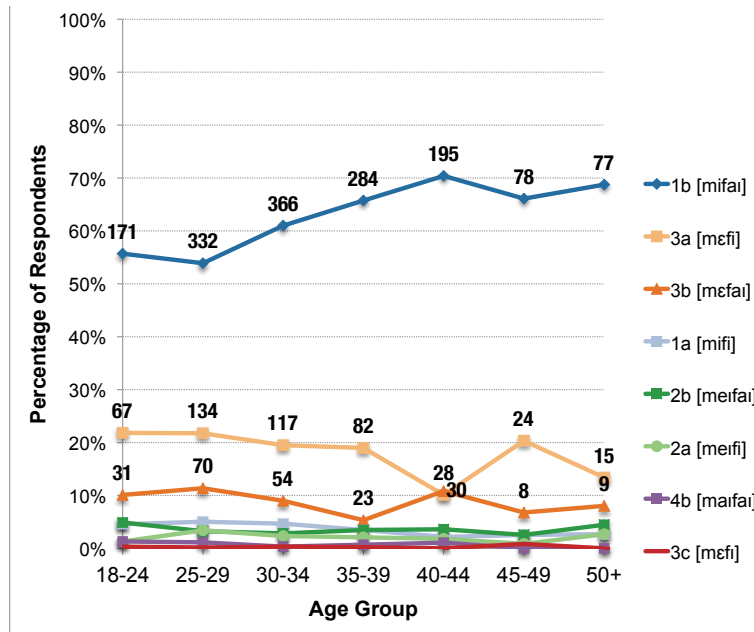


Note: 29 participants (1%) in the 2010 Survey and 33 participants (2%) in the 2012 Survey did not respond to the age question.

A significant difference was found between the Panel Data and the Survey Data participants in 2010. The Panel Data participants were slightly older than the overall 2010 Survey Data population ($\chi^2 = 14.596$, $df = 6$, $p = 0.0237$), with an average age of 34 in 2010. However, in 2012, where the participants in the 2012 Survey Data were significantly older than the 2010 Survey Data participants, there were no significant differences between the ages of the 2012 Panel Data participants and the overall 2012 Survey Data population.

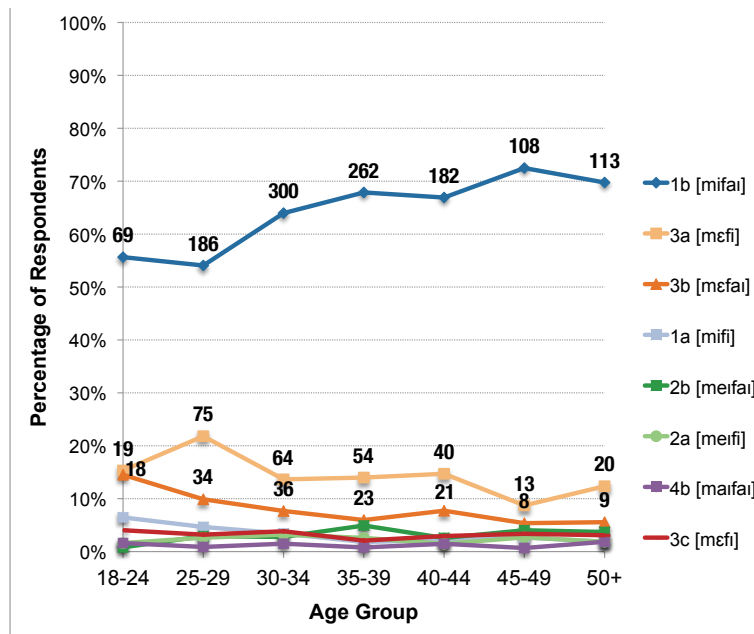
5.4.3.1 'MeFi' Pronunciation and Age

Figure 36. 'MeFi' Pronunciation by Age Group — 2010 Survey Data, N=2,462



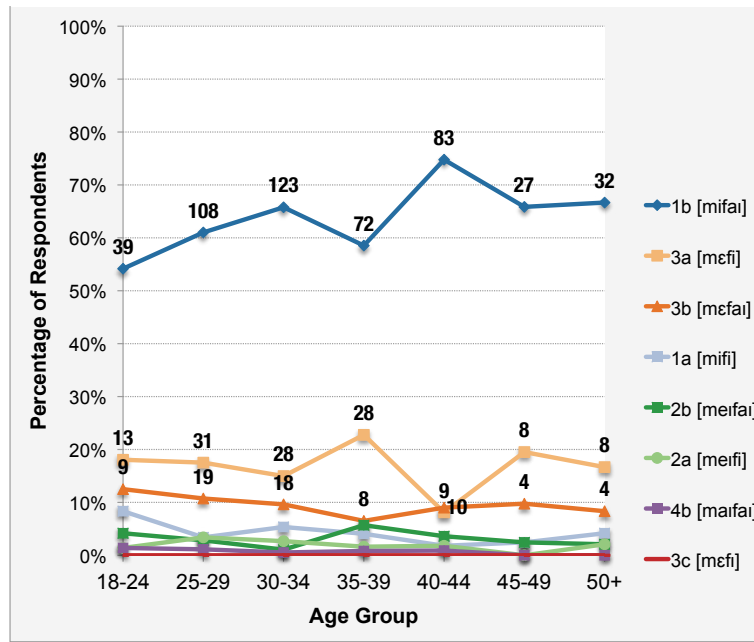
$\chi^2 = 42.2611, df = 12, p < 0.0001$

Figure 37. 'MeFi' Pronunciation by Age Group — 2012 Survey Data, N=1,906



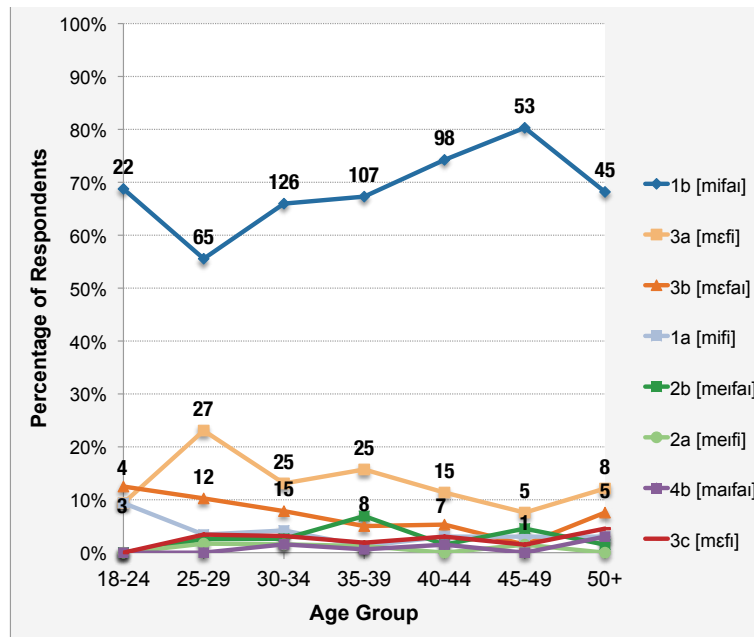
$\chi^2 = 39.3902, df = 12, p < 0.0001$

Figure 38. 'MeFi' Pronunciation by Age Group — 2010 Panel Data, N=759



$\chi^2 = 14.2676, df = 12, p = 0.2839$

Figure 39. 'MeFi' Pronunciation by Age Group — 2012 Panel Data, N=763



$\chi^2 = 22.9696, df = 12, p = 0.0279$

In testing for age-related correlations, participants who left either the age question or 'MeFi' pronunciation question blank were removed from analysis. This could have proved problematic for the age results in the Panel Data, where some participants may have omitted answers to survey questions in one year, but not the other. This may explain why, in 2012 Panel Data, the younger age groups were underrepresented as compared to their 2010 group sizes. This may also at least partially account for a significant outcome in the 2012 Panel Data, but

not the 2010 Panel Data. Additional analysis of the 2012 Panel Data revealed that when the 25–29 age group was removed, the results were no longer significant, suggesting that these participants' choices in 2012 were attributable to the skewed outcome.

In the Survey Data, age was significantly positively correlated with increased preference for the 1b - [mifɑɪ] variant. This was also found in the 2012 Panel Data, with an exception for the youngest age group, 18–24, whose preference for [mifɑɪ] rivaled that of the oldest age groups. While the hypothesis that age was significantly correlated with pronunciation variants was confirmed by the findings of this study, it is unclear why the results for were are so varied and inconsistent.

However, the somewhat simplistic treatment of age in this survey methodology may at least partly account for these inconsistent findings. Age was collected solely based upon the self-reports by participants of how many years they had lived (i.e., numerical ratio data). Furthermore, participants were categorized into relatively equal age groups for statistical analysis, which converted the numerical data into categorical data. This broad collection and treatment of age data did not take into account any qualitative aspects, such as life stage or experiential information of participants — these are factors which are more generally linked with various age-graded changes in linguistic behavior and could possibly be correlated with M-Set pronunciation.

While previous literature on age-led sound change often points to younger speakers as the leaders of linguistic innovation (Milroy and Milroy, 1985; Chambers, 2002; Guy, 2011), it is important to consider that within the culture of MetaFilter — the userbase of which overwhelmingly consists of participants in their mid-30's and is largely centered on the issues and interests of that generation — a youth-based explanation of trend patterns may not be apt. That said, it may well be that several age-related trends are co-occurring, resulting in a significant bias toward the 1b - [mifɑɪ] variant in older age groups, as observed in the Survey Data findings.

It is difficult to fully account for the correlations between 'MeFi' pronunciation and age without further investigation (which is unfortunately beyond what could be achieved in this case study). Age was also not mentioned in survey rationales or comments about the M-Set on MetaFilter; participants made no overt connections between M-Set pronunciations and age. This makes it challenging to hypothesize about how age might influence pronunciation, but the lack of social ascriptions or indexical relationships involving age are interesting to note for another reason — that is, similar to geography, age and pronunciation are related, but speakers have limited (if any) means with which to discern this information about others and therefore make these connections themselves.

5.4.4 Gender

The distribution according to respondent gender significantly differed between the 2010 Survey Data and 2010 Panel Data: $\chi^2 = 17.666$, $df = 2$, $p = 0.0001$. The Panel Data participants were more evenly balanced for male/female (55%/42%) than the overall 2010 Survey Data participants, who were skewed towards male (63%/35%).

Overall, the 2012 Survey Data showed a significant shift toward a more gender-equal population than the 2010 population ($\chi^2 = 113.956$, $df = 2$, $p < 0.0001$), and did not significantly differ from the 2012 Panel Data.

Table 14. Gender Distribution of MeFites in the 2010 and 2012 Surveys

Gender Category	2010 Count	2010 Percent	2012 Count	2012 Percent
Male	1,598	63%	998	51%
Female	873	35%	834	43%
QUILTBAG*	17	1%	86	4%
Declined to state	33	1%	39	2%
Total	2,521	100%	1,957	100%

*See 4.6.1.2 Data Normalization, p. 107 for an explanation of this term.

These results show that while MetaFilter still had a male majority in 2012, there has been a substantial shift towards a more balanced gender representation on MetaFilter since 2010. MetaFilter members have always shown interest in discussing gender issues. Participants often demonstrate, share and promote their knowledge and experience in areas of gender awareness and gender politics. The level of interest in these topics has increased in recent years, further explaining the more substantial shift in demographics as compared to earlier studies of MetaFilter by other researchers, in which the gender distribution had remained skewed but was relatively stable over time⁴⁵ (Lawton, 2005; Warnick, 2010; Sessions, 2010).

As new members join the community and may not fully realize the extensive history of discussion on these topics, a continual revisiting of previous discussion topics occurs, and especially so if these new or unfamiliar participants demonstrate behaviors that violate site norms concerning respect and gender equality (e.g., engaging in harassment, sexism, etc.). Revisiting these issues educates new members and reinforces the site culture and norms while also promoting a more inclusive environment for everyone to participate in. This in turn encourages a more balanced gender representation on the site.

45. Warnick (2010) reported 48% Male, 47% Female, 10% declined to state in his 2009 survey data. Sessions (2010) and Lawton (2005) both reported results from a survey conducted in 2004 by MetaFilter user 'fvw'. Referring to the same data, Sessions reported fvws findings differently, with MetaFilter being 68% male (Lawton reported 63% male). The survey data are no longer available for verifying the actual results.

The other factors leading to the gender difference between survey years were changes in methodological approach for the survey question. In the 2010 survey, the gender question was a radial button style with four options: Male, Female, Transgender, Other (with a fill-in explanation option). While it was appreciated by many that this question had more than two gender options (i.e., Male or Female), the four options that were available were problematic and limiting for other reasons. In the 2012 survey, the question was modified to be a free-form fill-in, allowing participants to describe their gender identity however they wished. This change was overwhelmingly preferred by the survey participants and allowed for a much more thorough understanding of how gender is perceived and constructed by participants of the MetaFilter community.

5.5 Summary of Data Results

Starting with the distribution of the M-Set, it was shown that **1b** variants ([mifɑɪ] and [mifɑɪt]) were the most preferred, and preference for them significantly increased in the 2012 data. Additionally, MeFites who changed their pronunciation in 2012 (from their 2010 choice) switched to **1b** variants more than they did to any other.

Measures of metalinguistic awareness from survey questions and rationales revealed that the amount of thought given to the M-Set was not necessarily correlated with ‘MeFi’ pronunciation, but various types of attention paid to linguistic forms and features were. For example, feeling very strongly about the exclusive use of preferred variants and attuning to the visual features of ‘MeFi’ were significantly correlated with higher preference for **1b** - [mifɑɪ(t)] variants. Overall, the range of metalinguistic awareness among participants was great, with many unable to explain their choice or be completely indifferent about them, and many others using extensive reasoning and high metalinguistic awareness to support their preferred variant(s).

Data results also showed that M-Set choices were correlated with various dimensions of demography. Pronunciation was seen to be slightly biased toward **1b** - [mifɑɪ(t)] variants, by all demographic measures (e.g. native language, non-English language experience, country of residence, age, gender, etc.), but the amount of bias differed significantly within most demographic factors (e.g., by country of residence and age, but not by language history factors or gender).

Even though demographic features of participants are not visible during online interaction, these results showed that demographic categories play an influential role in that interaction. The demographic categories participants belong to influence their sociolinguistic behavior, which in turn shapes the co-created identity of MeFites and MetaFilter.

While these demographic factors are important in explaining some of the differences

in the distribution of 'MeFi' pronunciation, they cannot fully account for what is happening within the user community as participants engage with each other and contribute to the practice over time. Various social engagement factors were examined and these results are presented and analyzed in **Chapter 6**, as they directly relate to the process of enregisterment.

Chapter 6: Enregisterment

6.1 Introduction to Enregisterment

In this chapter, additional data results pertaining to social engagement on MetaFilter will be presented and incorporated into an analysis of the enregisterment of the M-Set. These data show how enregisterment does not simply reflect the distribution of variants according to demographic factors, but rather is a result of a specific sociohistorical context which is shaped by the ongoing practice of its participants.

This chapter begins with a comparison of the enregisterment process in the MetaFilter community to the results of previous studies of enregisterment, highlighting the importance of the social structure of the practice in influencing how participants communicate and what types of social information they have access to. Following this, Agha's (2003) concept of the message chain will be adopted and extended to illustrate how various types of social engagement foster enregisterment. The expanded message chain concept will be applied to the presentation of data from surveys concerning how participants' social engagement with MetaFilter relates to M-Set pronunciation. Lastly, comment data from surveys and MetaTalk posts will reveal some of the common stances participants take about the M-Set, leading to an understanding of how the forms became enregistered with some of the particular indexical values that they have acquired.

6.2 The Enregisterment Process in CMC

In earlier studies involving FtF communities, the enregisterment of variables was partially accounted for by participants hearing linguistic forms and being able to indexically link them to the macrosocial demographic categories of the groups of individuals that they encountered using those forms. For the enregisterment of the M-Set, there is little opportunity for participants to hear forms and to associate those instances with the speakers and the social categories those speakers identify with.

However, not being able to hear variants or see participants does not necessarily mean that enregisterment is halted, or even slowed down. What occurs instead is that the indexicalities for the variants are primarily formed through other types of associations made by participants. The main channel through which MeFites acquire information about possible indexical associations for variants (other than by developing their own estimations and evaluations) is through direct and indirect messaging regarding the variables. The data results presented in this chapter show how the amount of information participants are exposed to is positively correlated with their social engagement in the community, which

results in participants' increased meta-linguistic awareness about the M-Set, language use, and participants associated with these topics. This is in line with the social constraints of CMC environments in general, where many social factors are obscured and paralinguistic and pragmatic features are often overtly expressed in the linguistic communication itself, e.g., in text as emoticons or explicit statements about states, feelings, responses, etc. (Rheingold, 1993; Warnick 2010; Greiffenstern, 2010). The outcomes are indexical relationships based on associations which are immediately visible, accessible and verifiable to participants. Therefore, in this online context, associations have developed having to do with the M-Set's lexical-semantic similarities with other forms, or other features of the forms themselves (e.g., CamelCase, and its indexical associations of informality and internet names). Additionally, associations have developed regarding authority figures who use particular variants (and what social categories they are perceived or known to belong to), rather than directly to demographic categories such as geography or age, which bias pronunciations but are not visible to or immediately verifiable by participants.

Early studies of enregisterment (Agha, 2003, 2005; Johnstone, Andrus, Danielson, 2006; Beal, 2009) did not elaborate on the possibility of observers not being able to hear features or see those who used them; these studies did not need to make this distinction. Even in the case of RP, for which enregisterment was largely aided by the dissemination of guidebooks about speech to wider circulations of people over time, existing notions of class (e.g., by instruction through schooling) aided the enregisterment of the variety and made the associations with demographic categories explicit (Agha, 2003; see **2.5.5.1 Enregisterment of RP – Associations with Socioeconomic Class, p. 33** for a more complete explanation of this example).

There is limited published research so far that foregrounds this distinction — the possibility that situations of enregisterment can occur whereby the social factors that may be correlated with variation are not visible or otherwise verifiable to speakers. To date, Squires' 2010 article on the enregisterment of internet language has taken the biggest step in detailing such a case, with her data providing evidence in which the sociocultural and historical context accounted for and aided the enregisterment of internet language more clearly than the correlational indices to social factors did (Squires, 2010, p. 460).

Previous studies involving the role of networks in linguistic innovation, specifically those by Milroy, L. (1980), Milroy and Milroy (1985), have also made steps toward the role of social context and community participation as equally if not more important than demographic variables in accounting for diffusion and change. While their studies focused on the spread of innovative features and not enregisterment specifically, there is considerable overlap in the nature of and process by which both concepts are defined and operate within

communities. Outcomes from all of these studies will be drawn upon throughout this chapter to provide an interpretation of the data which considers the roles of demographic and social engagement factors as both distinct and valuable influences on linguistic behavior within a specific community context.

6.3 Message Chains

Information about M-Set variants is disseminated through message chains, across various media and modalities (see 2.5.2.1 **Message Chains**, p. 27 for more on this concept). In previous studies of enregisterment, message chains were also responsible for the spread of ideas, but in those cases the social categories of message senders and receivers were aligned, meaning that information about the demographic categories (e.g., age, gender, geographic background) of the speakers was accessible (i.e., visible or hearable) alongside the content of their messages (Agha, 2003, 2005; Johnstone, Andrus, Danielson, 2006; Beal, 2009; Hall-Lew and Stephens, 2011). In the case of MetaFilter, these categories are not usually visible to message-senders. Therefore, their values do not get readily associated with message content, and the indexical relationships that are formed do not frequently serve to make those connections (even though the connections between the demographic categories and the variants do exist, and correlate with the choice of variant). In addition to the results outlined in **Chapter 5**, this is evidenced by social commentary about the variants, which rarely included associations with demographic factors (excluding geographic location, usually as a proxy for dialect). When associations did occur, they were almost always as a reference to participants' non-English background as a possible explanation for what they perceived to be outlier choices in M-Set pronunciations (and not their age or gender). These types of statements were given much more frequently in survey rationales than in public discussions online about the M-Set, further demonstrating that message chains containing messages linking demographic characteristics with pronunciations were seldom seen. Rarer still were ascriptions linking age and pronunciation, even as speculation. These types of stances were virtually nonexistent in both survey rationales and online discussions, despite the data showing significant trends in age groups and pronunciation, with older groups showing greater preference for **1b** – [mifaɪ(t)] variants (see 5.4.3 **Age**, p. 145).

It is not surprising that message chains about the M-Set rarely if at all linked such characteristics to pronunciation, given the limitations of observing demographic information in the text-based medium. Therefore, what is of greater interest for our understanding of enregisterment in this context is to describe which characteristics or ideas participants do link to pronunciation. These connections can be observed by examining patterns in the message chains that are created online and are reflected in participants' stances in survey rationales.

6.3.1 Message Chain Components

The minimal requirements of a message chain consist of a sender, a receiver, a message to be passed along the chain, and some sort of shared discursive experience that allows this contact to occur. Other features, such as the modality or medium in which the unit is transferred, the values the cultural unit contains and/or the content of the message, and the composition and number of participants comprising the audience or message-sender, are all mutable and undefined. For example, in Agha’s (2003, 2005) account of the enregisterment of RP, message chains enabling the enregisterment of the variety involved many published texts, grammar guides, etc. (as senders of the message), disseminated over time, in a one-to-many relationship (whereby a single text might have been received by a large audience of receivers).

The ‘message’ component of a message chain includes any form of expression — an idea, a stance, an evaluation or a cultural unit of value (e.g., features, symbols or objects having meaning to individuals or groups). The message does not need to be linguistic; any communicative act will suffice. However, all message chains in this research were linguistic. The modality and medium in which the message is conveyed is entirely flexible as well, encompassing any way a message can be expressed from one individual or entity to another.

All MeFites are part of a shared discursive experience, a message “*chain network*” in Agha’s (2003) terms. MeFites do not need to be aware of being a part of this message chain network, or even being aware of each other (Agha, 2003, p. 248). As long as the minimal requirements are met, the message chains can be created, they can spread through the message chain network, and they can help foster the enregisterment of variables, and do so regardless of whether participants in the message chain network are aware of this process.

Figure 40. Example of a Text-Based MetaFilter Message Chain

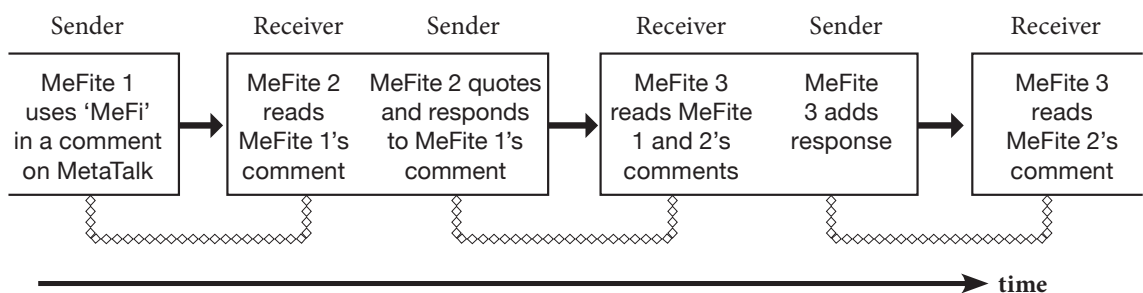
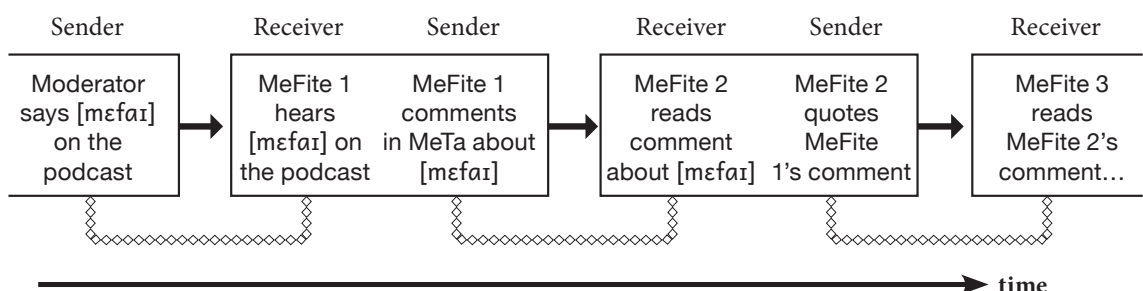


Figure 41. Example of a Multi-Modal MetaFilter Message Chain



The message chains shown in these examples differ in modality and audience reach. The text-based message chain in **Figure 40** is more accessible to a wider audience, as the text-based MetaFilter subsites act as the central hub in the constellation of practices. The information that is exchanged in these strictly text-based message chains is direct and easily traceable, even after time has passed. However, the text-based message chain is unlikely to contain information about pronunciation, unless that is the topic of the online discussion (regardless, it often has been the topic on MetaFilter).

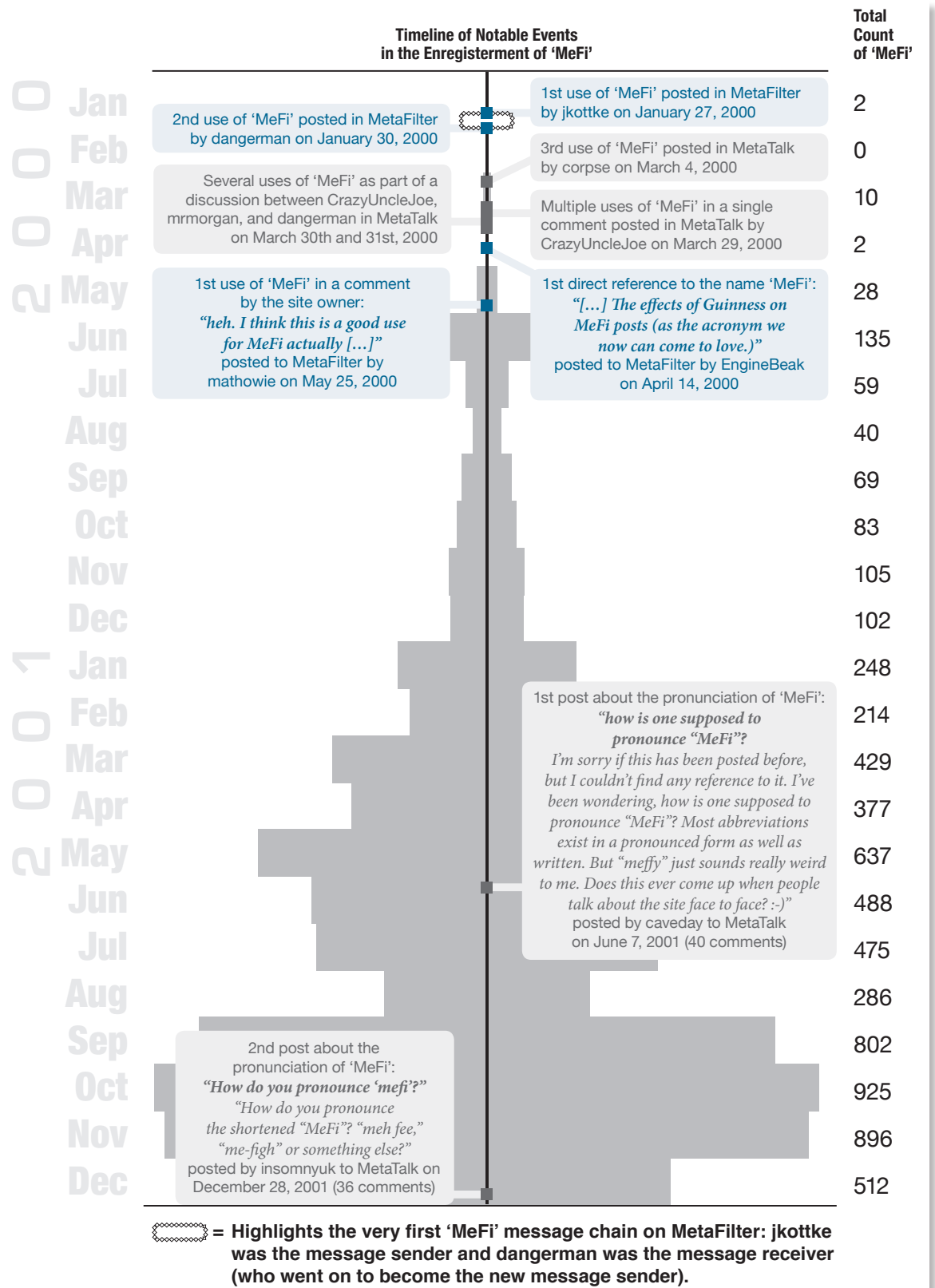
Conversely, the multi-modal message chain in **Figure 41** includes observable pronunciation(s) when the modality is spoken (e.g., podcast, meetups), but this information is not as widely disseminated, as drastically fewer participants listen to the podcast than read the site. Additionally, the participation types (e.g., core, active, occasional, peripheral, transactional) of receivers of text-based message chains is much broader than that for multi-modal message chains, as mostly core and active participants are invested in the practice in ways that allow them to be exposed to multi-modal message chains (e.g., through listening to the podcast, attending meetups, and other less popular and more in-group activities).

In both message chain examples, links in the chain may continue to build as other participants (even non-MeFites) become receivers by reading some or all of these comments. Some of them may go on to become message senders themselves. From this, the original message (i.e., the moderator pronouncing ‘MeFi’ in a particular way) reaches many others and an array of evaluations may be ascribed to the messages along the way.

6.3.2 MetaFilter Message Chains and the Frequency of ‘MeFi’

Actual examples of message chains involved in the enregisterment of the M-Set on MetaFilter are demonstrated in **Figure 42**. The very first message chain has been highlighted and subsequent notable events in the enregisterment history are explained.

Figure 42. Timeline of Notable Events in the Enregistration of 'MeFi'



In research concerning diffusion of new linguistic features, a distinction is made between *innovators* and *early adopters* (Rogers and Shoemaker, 1971; Granovetter, 1973, 1983; Milroy and Milroy, 1985). Innovators are the first to use a form or feature, but they are not necessarily

crucial to its success. Whether or not an innovation is picked up by a wider audience is largely influenced by the early adopters, who accept the innovation and model its use for other non-innovators (Milroy and Milroy, 1985, p. 367). This has been the case with ‘MeFi’ on MetaFilter. In **Example 2**, jkottke spontaneously creates the nickname ‘MeFi’, but it is not until weeks later that it is picked up again with any frequency. A particularly notable example of this is shown in **Example 3**; CrazyUncleJoe is an early adopter of the innovation, and models it four times in a comment that explicitly links the nickname to various aspects of the community.

Example 2. Innovation of ‘MeFi’

That’s Captain Cranky Pants to you! Or perhaps Captain Reality Pants?

And I like the idea of having users “earn” the right to post as a compromise between the perceived elitism of having only approved posters and the crap-o-rama free-for-all MeFi is now experiencing. Perhaps a little quiz they have to take, or only posting after a week or 10 comments, whichever comes first. How about attaching karma to posters instead of posts?

Anyway.

posted by jkottke at 9:11 PM on January 27, 2000

Example 3. Early Adoption of ‘MeFi’

MetaTalk is *supposed* to be geared more to discussion of MetaFilter and (I think) weblog-specific issues (although the current “hot topic” on MeFi would indicate that the latter isn’t really true). MetaTalk has several posts that got here because of a weird posting bug, and a few posts that discuss the process of MeFi, the T-Shirts of MeFi, and (coming soon) sexual fantasies we’ve had about the owner of MeFi. That, and they are different colors.

posted by CrazyUncleJoe at 6:43 PM on March 29, 2000

Generally, early adopters are more central to the group than innovators (Milroy and Milroy, 1985, p. 367), although that is not necessarily the case in this particular example of innovation. For a variety of reasons, it is difficult for both researchers and community participants to assess the participation level of any given member at a particular time on MetaFilter, even when the history of their contributions to the site is publicly available through their profile page. However, the survey methodology was able to address this to some degree.⁴⁶

Early adopters have the influence and reach through their strong ties within their communities, allowing the innovation to be observed by several others. Eventually the innovations are adopted by a wider audience through a series of message chains which start off slowly at first, then increase rapidly over a short period of time. Eventually, the innovation is disseminated and reaches a stable point. This entire process (from innovation to early

46. Asking participants how frequently they visited subsites, listened to podcasts, attended meetups, etc. gives some indication of how (broadly) invested in the community they are, which reflects a more engaged role as a core or active participant (as opposed to a more peripheral role with limited or narrow engagement).

adoption to stability) resembles an S-shaped curve, indicative of change over time (Milroy and Milroy, 1985, p. 367).

Figure 43. Early Stage: Frequency of 'MeFi' by Month, 2000–2001

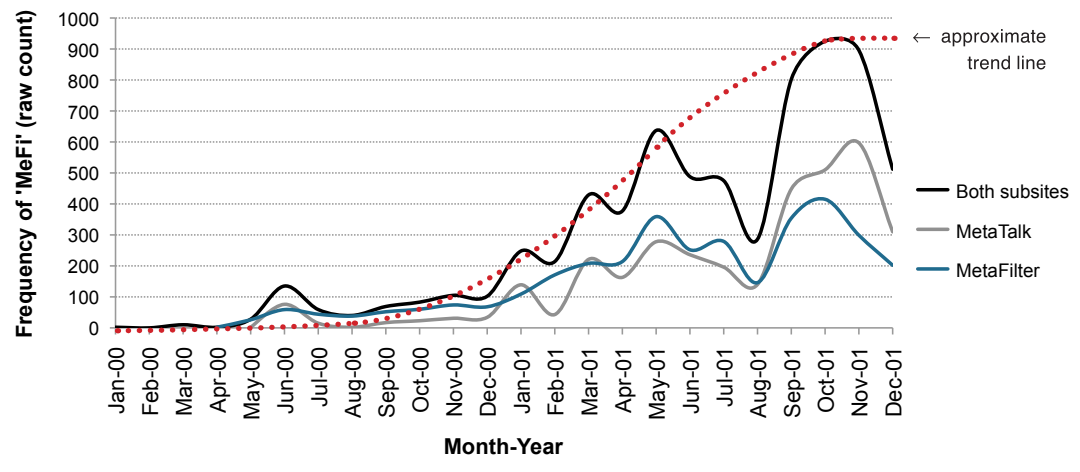


Figure 43 shows the early stage of the diffusion of ‘MeFi’ through the community, i.e., the leftmost bottom tail of the S-shaped curve. The dip in frequency shown in August 2001 reflects the temporary closure of MetaFilter while the site owner, Matt Haughey (username mathowie) went on vacation. The overall number of site comments, including instances of ‘MeFi’ were drastically reduced during this month.

In the early stage, instances of use of ‘MeFi’ were sporadic at first, but a steady, rapid increase is visible from approximately January 2001 onward, a full year from when ‘MeFi’ was first coined on the site. This process is also observable in Figure 42. **Timeline of Notable Events in the Enregistration of ‘MeFi’, p. 158**, and includes monthly totals of the frequency of ‘MeFi’.

Figure 44. S-Shaped Curve Showing Adoption of 'MeFi' by Year

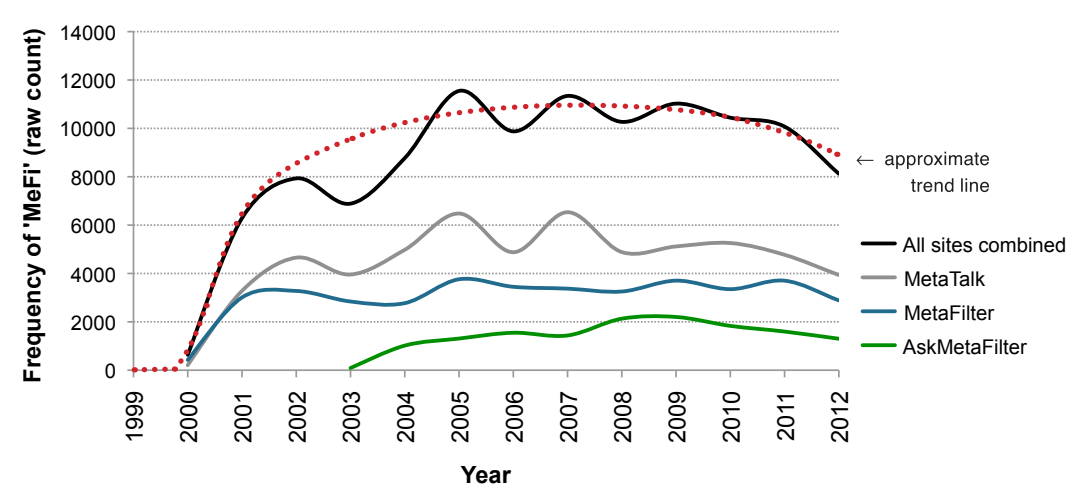


Figure 44 shows the entire timeline, up to 2012, of the innovation, adoption and stable use of ‘MeFi’. There were few notable events in MetaFilter history which may make it seem like this is not a more typical S-shaped curve at first glance. A few of the peaks and dips in ‘MeFi’ frequency can be explained by these events, and resulted in corresponding increases or decreases in overall participation. For the majority of 2003, the site was closed to new user registrations. The site also experienced several outages and the servers were subsequently moved to new locations. This led to less participation in 2003 than in the years prior and following. The peaks in ‘MeFi’ use in 2005 and 2007 can be accounted for by increased participation owing to very long and involved posts on MetaTalk about MetaFilter culture and policies occurring in late February 2005 and November 2007.

With these outliers accounted for, the overall shape of ‘MeFi’ usage over time shows a slow uptake in early 2000 (see Figure 43), a rapid increase from 2001 through 2002, and stable usage from 2004 through to 2011. It is unclear why usage in 2012 dropped to 2002 levels, but it was likely owing to an overall decline in participation from November 2012 onward, when changes in Google’s indexing algorithm caused MetaFilter’s site traffic to plummet by 40% overnight (Auerbach, 2014; Sullivan, 2014).⁴⁷ Regardless, the overall pattern of ‘MeFi’ usage across the subsites resembles the S-shaped curve that is typical of the spread of a linguistic innovation through a community. The MetaFilter subsite is more reflective of this pattern, although MetaTalk resembles the S-shape as well once the outliers (i.e., 2003, February 2005, November 2007) are considered.

It should also be noted here that the use of ‘slow’ to characterize the early stage of innovation diffusion is a relative term. It took approximately one year from the time of first use of ‘MeFi’ until the beginning of a rapid increase in usage could be observed. This would be an unusually rapid rate of uptake for innovation adoption in a typical FtF speech community, but is not unreasonable to be considered ‘slow’ in a CMC environment, where register progressions occur rapidly and memes, news, and other forms of information diffuse much more quickly than in traditional communicative spaces.

6.4 Social Engagement Factors and the Enregistration of the M-Set

Now that the overall pattern and frequency of ‘MeFi’ has been explored, it is necessary to look at the role of social engagement on the use of ‘MeFi’, its influence on the pronunciation of the M-Set, and how social engagement has contributed to the enregistration of these innovations as meaningful parts of the MetaFilter register. These factors were outlined in 4.4.3 **Measures of Social Engagement, p. 98**, and are listed with additional detail in Table 15.

47. This change contributed to loss of revenue and resulted in MetaFilter necessarily laying off half of its moderation staff as of June, 2014.

Table 15. Characteristics of Social Engagement Measures

Measure	Modality	Measurement Metric	Interactivity/Mode	Reach of Participation Levels
Subsite Visitation	Online reading, text-based interaction	Frequency of subsite visitation	Passive/Highly variable	Easily accessible to all participation levels
Podcast Listening	Accessed online, speech modality	Frequency of podcast listening	Passive/Mostly imagination and alignment	Mostly core and active participants
Meetup Attendance	Offline, FtF speech modality	Frequency of attending meetups	Interactive/Mostly engagement, but variable	Mostly core and active participants

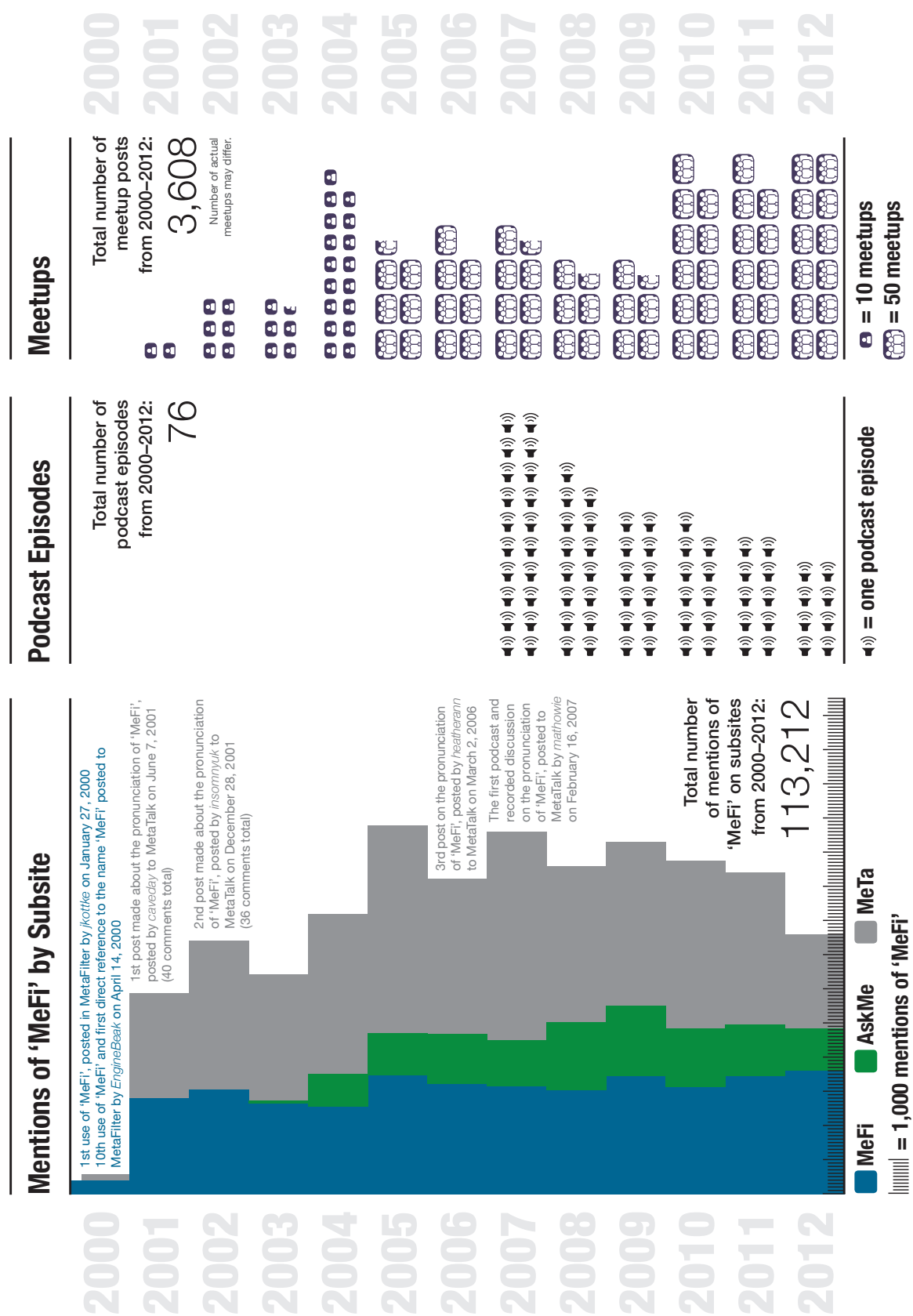
It is evident that these measures of social engagement fundamentally differ in terms of structure, modality, interactivity, and reach. Therefore, the amount and type of participation which occurs by these measures varies.

Additionally, the modes of belonging present for each measure can vary, with some activities seeing greater participation through engagement, whereas other activities being experienced more frequently through imagination or alignment modes (see 2.2.1.4 Types of Participation, p. 13 for more on modes of belonging). This results in qualitative differences in the message chains that are created and exchanged through each medium (e.g., measure), as MetaFilter members shift modes depending on the medium in which they are participating.

Differences in amount of content available over time for each measure is exemplified by the infographic displayed in Figure 45. This provides additional context for assessing the reach and popularity of each subsite, compared to the number of podcasts available and meetups planned over time.

Following, data results from the surveys pertaining to social engagement measures are analyzed within the context of MetaFilter’s unique social constellation. Message chains are also explained in terms of the mediums in which they are sent and received, contributing to a more comprehensive picture of the enregisterment of the M-Set variables. The rest of this chapter gives an account of this process, as it occurred on MetaFilter.

Figure 45. Social Engagement Factors and the Enregistration Timeline



The first column in **Figure 45**, “Mentions of ‘MeFi’ by Subsite”, displays the frequency results previously discussed in **6.3.2 MetaFilter Message Chains and the Frequency of ‘MeFi’**, p. 157. The second and third columns contextualize this data, by placing it alongside data involving podcasts and meetups. This allows the frequency of use of ‘MeFi’ on the subsites to be compared with opportunities for hearing spoken instances of use, over time and by two audio-based measures.

Figure 42. Timeline of Notable Events in the Enregistration of ‘MeFi’, p. 158 showed how the first focussed discussions about the pronunciation of the M-Set occurred on the site starting in mid-June, 2001. It was nearly six years later that regular, spoken instances of the M-Set would be accessible through the podcast. During this time, meetups did occur, but with much less frequency, and they were usually held in bigger American cities, where the greater numbers of MeFites were residing.

The numerous discussions that occurred on the site regarding the pronunciation of the M-Set prior to the podcast and/or the popularity of meetups demonstrates that awareness of pronunciation issues had been established for many years, in text, and with little opportunity for spoken feedback and evaluation. The podcast and meetups brought new attention to these issues, and allowed for innovative and expanded types of message chains to occur. In the following pages of this chapter, these social engagement factors will be investigated further and their role in enregistration will be more fully explored.

6.4.1 Year of Joining MetaFilter

The year of joining MetaFilter is not a measure of social engagement per se, but it can be reflective of the length of time one has known the community and engaged with its practices. Length of membership within a community can also be a proxy for measuring social capital, as longevity loosely translates to authority in knowing the community’s unique history.

Figure 46. Participants’ Year of Joining MetaFilter – 2010 and 2012 Surveys

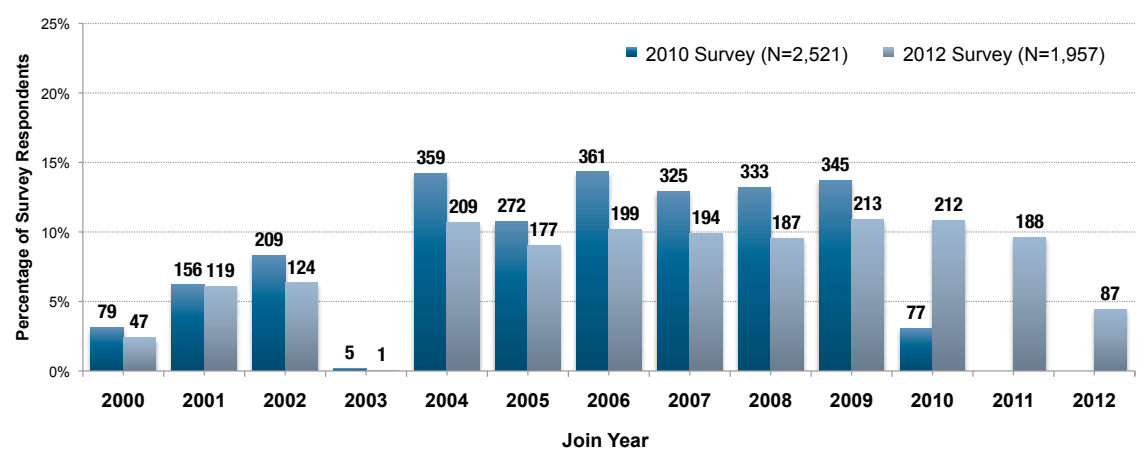


Figure 46 shows the distributions of the years in which survey participants joined MetaFilter. The site was closed to new signups in 2003⁴⁸, which led to a spike in new memberships when signups were reopened in 2004. The 2010 survey was conducted in March, allowing only three months of 2010 memberships to be included. The 2012 survey was conducted in August, capturing only eight months of 2012 signups. Accounting for these exceptions, no year was under- or over-represented in the surveys. Additionally, there were no significant differences in the years that Panel Data participants joined MetaFilter as compared to the rest of the survey population.

6.4.1.1 *‘MeFi’ Pronunciation and Year of Joining MetaFilter*

Year of joining MetaFilter and M-Set pronunciation were not significantly correlated in the Survey or Panel Data sets, for either year. This was a surprising outcome, as it was hypothesized that participants who joined or were active members in years when the debate was more frequently discussed (e.g., 2005–2007, see **Figure 45. Social Engagement Factors and the Enregistration Timeline**, p. 163) might be biased toward 1b – [mifaɪ(t)] variants. However, similar to the lack of correlation in the ‘amount of thought given’ measure⁴⁹, it may not be the year in which a participant has joined (or the length of time one has been a member of MetaFilter) that matters in pronunciation outcomes, but how the participant spent that time, e.g., whether or not the participant attended meetups, listened to podcasts, interacted with other MeFites.

6.4.2 **Frequency of Visitation to MetaFilter Subsites**

Reading posts and comments on MetaFilter subsites is the primary way MeFites connect with each other in the community. Comments made on the site reach a wider audience than do messages conveyed through participation by any other means. In this sense, the message chains on MetaFilter subsites are the most influential, in that they contain messages created by a participant which reach many and remain accessible over time.

As all content on MetaFilter is public⁵⁰, these messages can be read by anybody outside the community as well. Comments and posts can be easily shared beyond the confines of MetaFilter, either using the Twitter or Facebook sharing buttons on every post or by manually copying and pasting links in emails or on other sites. Additionally, many comments and posts reach wider audiences through Google or other search engines; historically, AskMetaFilter has often been a very frequent and high-ranking result for internet users’ queries, due to its

48. See 3.2.3 **The MetaFilter Userbase**, p. 61 for more information about the signup closure.

49. See 5.3.2 **Amount of Thought Given to the Pronunciation of the M-Set**, p. 123 for data results concerning this measure.

50. Excluding some aspects of users’ profile pages as well as the infrequent deletion of posts or comments due to violation of site rules.

extensive history of indexed archives. This has the potential of both drawing people into the community as well as disseminating information and messaging about the community.

The three most popular subsites on MetaFilter — The MetaFilter front page, AskMetaFilter, and MetaTalk — offer qualitatively different ways of interacting with the MetaFilter community. Each subsite has its own purpose and therefore the subsites vary with respect to how internally- versus externally-focused they are.

Subsite visitation frequency data are shown in the following figures for the 2010 and 2012 surveys. It should be noted that the 2012 survey included an extra category, ‘Never’, as some participants gave helpful feedback in 2010, stating that they actively avoid some subsites or simply never visit them. This extra category provided a distinction between very infrequent visitation versus none at all.

Figure 47. MetaFilter Subsite Visitation Frequency – 2010 Surveys, N=2,521

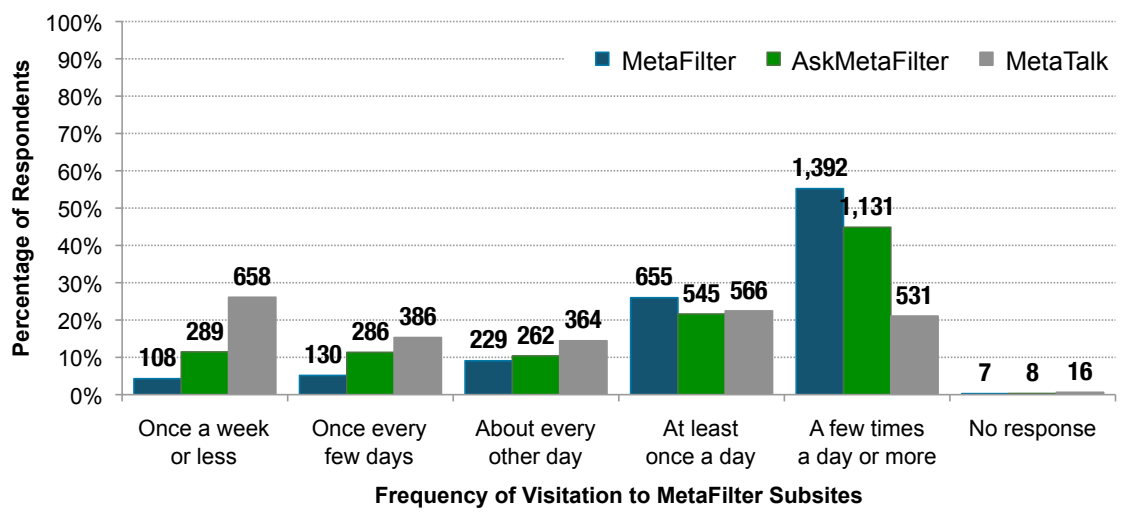
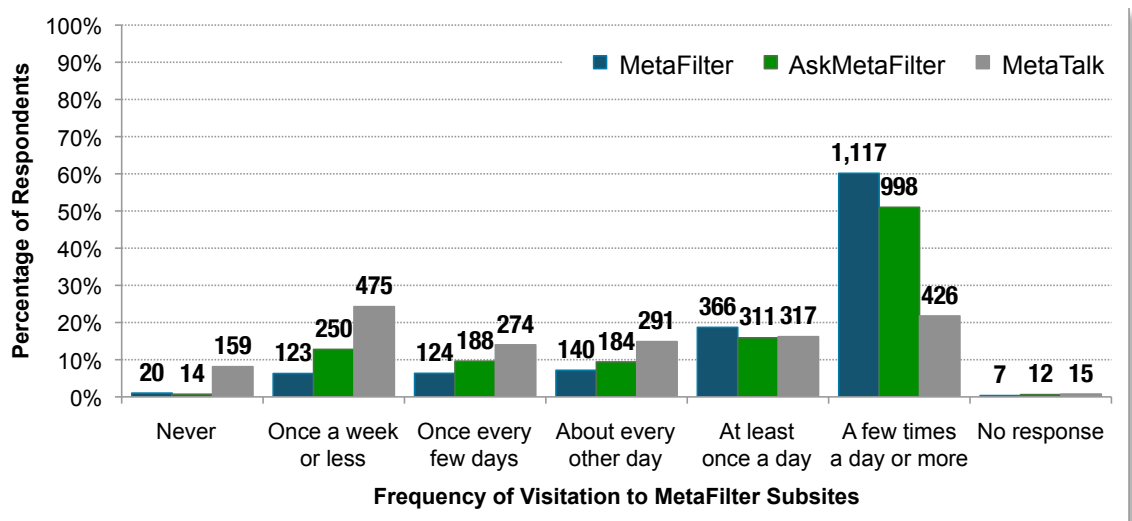


Figure 48. MetaFilter Subsite Visitation Frequency – 2012 Surveys, N=1,957



The MetaFilter front page was visited the most frequently by participants — 1,392 (56% of MeFites) out of 2,490 who answered this question in the 2010 survey reported that they visited The MetaFilter front page at least a few times a day or more; the equivalent figures were 1,177 (61% of MeFites) out of 1,923 in the 2012 survey. This was expected, given that participation on this page is and always has been the main purpose of the site in general. However, it is noteworthy that there was a 5% percentage point increase in reported daily (or more frequent) visitation to The MetaFilter front page since the 2010 survey. The frequency of visitation to The MetaFilter front page by Panel Data participants did not significantly differ from the Survey Data frequencies.

AskMetaFilter visitation was comparable to The MetaFilter front page, but occurred to a slightly lesser degree. In 2010, 45% of MeFites reported visiting AskMetaFilter at least a few times a day or more; the figure was 52% in 2012. Again, a percentage point increase (7%) in the most frequent visitation over time was observed. Panel Data participants showed significantly higher visitation frequencies to AskMetaFilter than the Survey Data participants in 2010 only ($\chi^2 = 30.229$, $df = 4$, $p < 0.0001$).

MetaTalk visitation differed from the other two main subsites, with MeFites reporting that they visited ‘once a week or less’ more than they reported visiting ‘a few times a day or more’. In 2010, 658 (26%) of survey respondents reported that they visited MetaTalk once a week or less, compared to 531 (21%) who reported that they visited at least a few times a day or more. In the 2010 survey, these percentages were 25% and 22%, respectively. Panel Data participants differed from Survey Data participants with significantly higher visitation frequencies to MetaTalk for Panel Data participants ($p < 0.001$ for both survey years).

It should be noted that overall differences in visitation frequencies across subsites was partially due to the frequency of updated content on each of the subsites. The MetaFilter front page tends to have less than half as many posts per day as AskMetaFilter, but those posts generate many more comments than those on AskMetaFilter. MetaTalk receives much fewer posts and comments than either the MetaFilter front page or AskMetaFilter.

As covered in **3.2.1 The MetaFilter Subsites, p. 48**, the focus of each subsite varies as well. The MetaFilter front page is the most outwardly-focused, with posts necessarily centering on topics external to the author of the post (even though they are likely to be topics of interest to the author). AskMetaFilter has stricter guidelines about casual chatter and personal storytelling than MetaFilter, but is at the same time more inwardly-focused, in that the content is specifically centered around issues personal to the post author. Participants in AskMetaFilter threads may respond with personal stories or information, but only in the context of helping the post author to solve the issue at hand. In this way, AskMetaFilter is more

highly regulated than the MetaFilter front page. Lastly, MetaTalk is the most inwardly-focused of the three main subsites, sometimes highly personal. Posts and comments on MetaTalk are the least regulated (so long as they still adhere to the overall guidelines of the subsite).

These differences in focus naturally result in shifts in the register across the subsites, as well as shifts in the norms about what is appropriate or encouraged behavior. The number of visitors, frequency of visitation and qualitative differences of focus for the three main subsites reveal that while the MetaFilter front page and AskMetaFilter receive the most visitors and visitations, it is MetaTalk that is the most reliable measure of social engagement. The visitors to MetaTalk — though much fewer overall — are the ones who are the most involved members of the community in the sense that these participants are aware of site changes, norms and issues; often they actively participate in shaping them. Therefore, an effect of engagement with this subsite and the pronunciation of ‘MeFi’ was expected, as visitors to MetaTalk are the people most actively and frequently discussing the topic.

Some evidence of the shift in focus and content across the subsites can be observed by the frequencies with which ‘MetaFilter’ and the M-Set are used in comments on each subsite. In the following figures, the parts-per-million (PPM) word frequency for ‘MetaFilter’, ‘MeFi’, ‘MeFite’, and ‘MeFites’ by subsite and year is given, showing a data-normalized progression of word use over time. Note that the scales the figures differ, commensurate with the frequency of the particular term analyzed.

Enregistration

Figure 49. Frequency of 'MetaFilter' by Subsite and Year

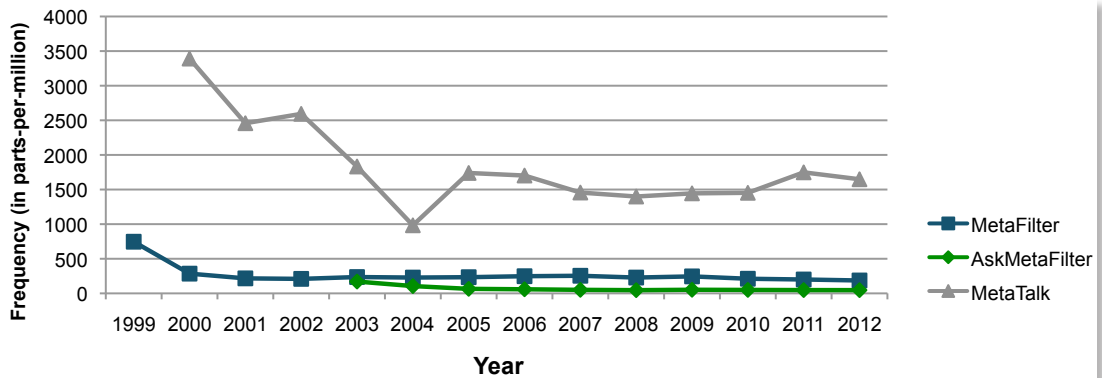


Figure 50. Frequency of 'MeFi' by Subsite and Year

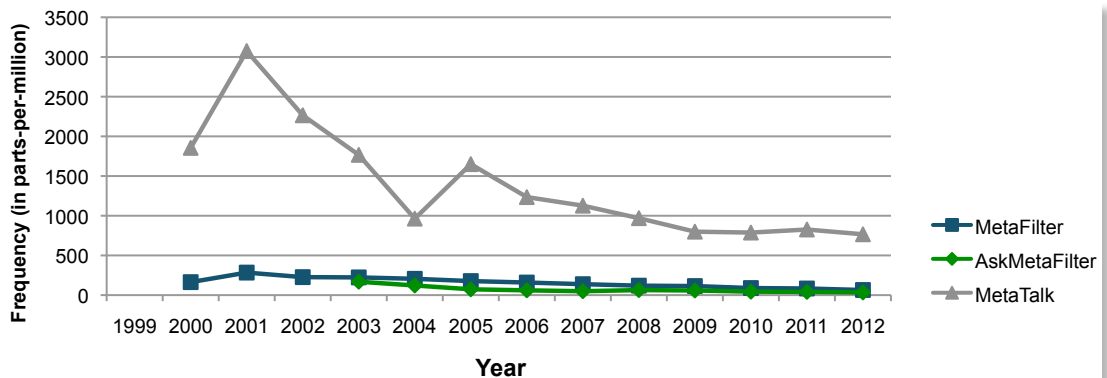
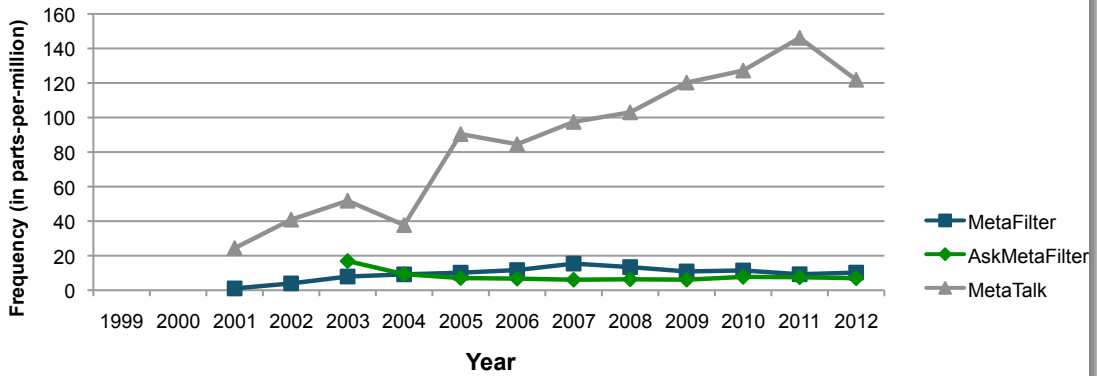
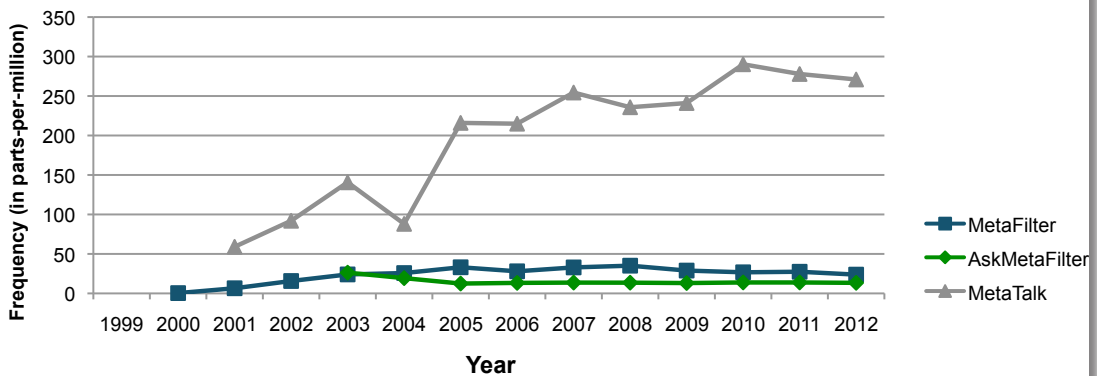


Figure 51. Frequency of 'MeFite' and 'MeFites' by Subsite and Year

PPM Frequency of 'MeFite' by Subsite and Year



PPM Frequency of 'MeFites' by Subsite and Year

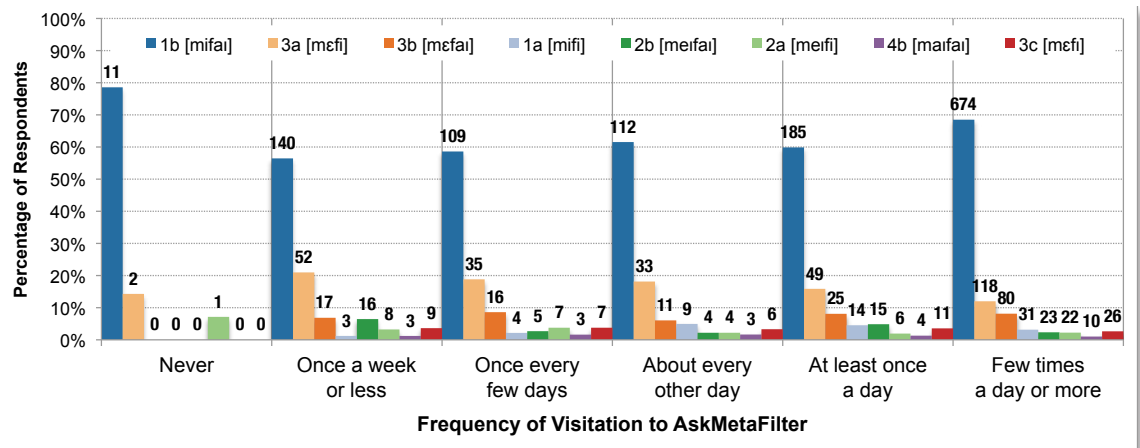


There are substantially more mentions of and meta-discussions about the M-Set on MetaTalk than on the other subsites. AskMetaFilter showed the least amount of reference to ‘MetaFilter’ and the M-Set, which was commensurate with the norms for that subsite and the nature of the discussion that occurs there.

6.4.2.1 ‘MeFi’ Pronunciation and Frequency of Visitation to MetaFilter Subsites

Visitation frequency to some MetaFilter subsites was found to be significantly correlated with ‘MeFi’ pronunciation in the 2012 Survey Data. These results were significant in two ways: the subsite that was visited, and the frequency of that visitation. Only data sets and subsites with significant outcomes are shown (only visitation to AskMetaFilter and MetaTalk were correlated with pronunciation variants).

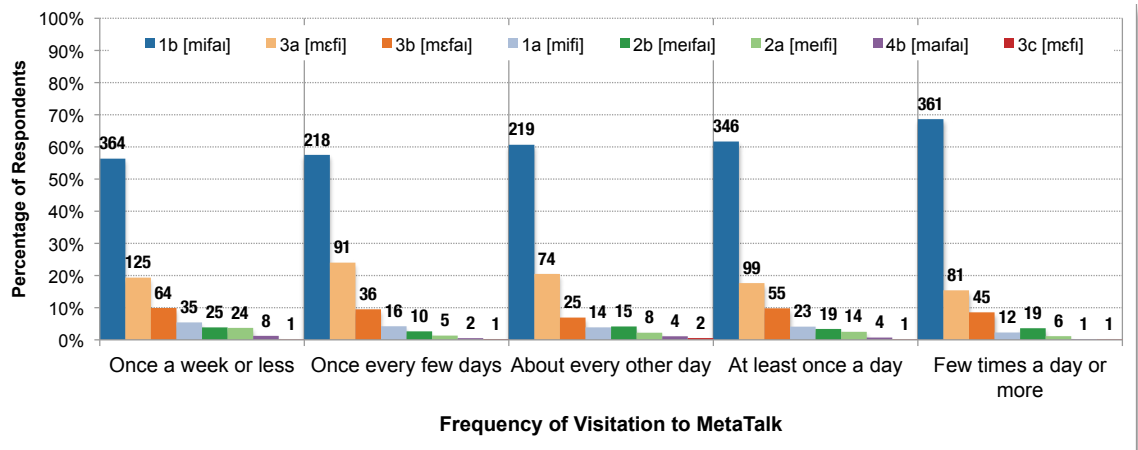
Figure 52. ‘MeFi’ Pronunciation by AskMetaFilter Visitation – 2012 Survey Data, N=2,477



$$\chi^2 = 40.1056, df = 20, p = 0.0048$$

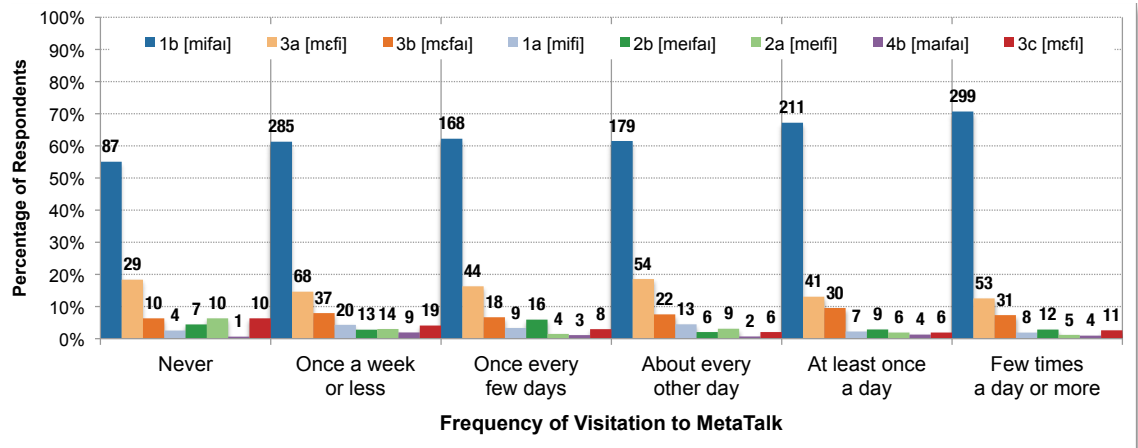
In the 2012 Survey Data only (2010 data results not shown), the most frequent visitation to AskMetaFilter significantly correlated with an increase of preference for the 1b - [mifai] variant. AskMetaFilter is more focused on the personal issues of community participants than the MetaFilter front page, but less so than MetaTalk. This shift in focus could explain why very frequent visitation to AskMetaFilter was a significantly correlated with ‘MeFi’ pronunciation, while any visitation to the MetaFilter front page was not.

Figure 53. 'MeFi' Pronunciation by MetaTalk Visitation — 2010 Survey Data, N=2,473



$$\chi^2 = 39.4683, df = 20, p = 0.0058$$

Figure 54. 'MeFi' Pronunciation by MetaTalk Visitation — 2012 Survey Data, N=1,921



$$\chi^2 = 44.7566, df = 25, p = 0.0089$$

There were statistically significant differences in the distribution of 'MeFi' pronunciations and visitation frequencies on MetaTalk in the 2010 and the 2012 Survey Data. As discussed previously, the amount of content that is updated on AskMetaFilter and MetaTalk differs, with MetaTalk showing the lowest average number of daily posts than the other two main MetaFilter subsites. MetaTalk also has seen substantially fewer on average daily comments than AskMetaFilter. Therefore, the incentive to visit MetaTalk more frequently than the incentive to visit AskMetaFilter is less, as participants will be less likely to find new content on MetaTalk at any given time. This makes analyzing visitation frequency on MetaTalk challenging, as the less frequent visitation categories may not necessarily align with less interest or involvement in MetaTalk, but rather an awareness of how likely there will be new content to read and posts to participate in. Regardless, there was a significant correlation between the most frequent visitation category and an increased

preference for the 1b – [mifaɪ] variant. That is, those who visited these two subsites the most (regardless of how much content is updated there) demonstrated an increased preference for the 1b – [mifaɪ] variant.

These findings help support the hypothesis that those who engage with MetaTalk more frequently (even though there are fewer of these participants overall), might be more biased toward the 1b – [mifaɪ(t)] variants. This hypothesis is also reinforced by the observation that, “[...] *the closer the individual’s ties to a local community network, the more likely he is to approximate to vernacular norms [...] a closeknit network has an intrinsic capacity to function as a norm-enforcement mechanism*” (Milroy and Milroy, 1985, p. 359). This also illustrates the influence and reach of online social interaction, e.g., this ‘closeness’ holds, even when participants are not in physical proximity, and are not communicating in a speech modality.

6.4.3 MetaFilter Site Participation and Enregisterment

Increased frequency of visitation to MetaFilter subsites is an indicator of deeper involvement in the community. As a result, participants who visited subsites frequently would be exposed to more content and would be more likely to be receivers of messages about the M-Set than those who visited less frequently. More frequent visitors would also generally have increased meta-linguistic awareness about the community and its norms than less frequent visitors, through repeated exposure to comments containing others’ stances on matters of language and the M-Set.

The particular subsite that was visited was correlated with preference for M-Set variants, with the more internally-focused subsites correlating with increased preference for 1b – [mifaɪ(t)] variants. It is in these more inward-looking areas — especially on MetaTalk — that discussions about the M-Set, language policy, and matters of community, identity or site norms are most likely to occur. Individuals who participate in these areas and do so with more frequency have increased likelihood of taking up roles as receivers and possibly senders in message chains, through increased exposure to messages related to the M-Set.

Unlike speech, the nature of the text-based medium also allows these message chains to be carefully planned, previewed, and edited prior to posting.⁵¹ However, also differing from speech, accents and acoustics are obscured, requiring participants to use eye-dialect spellings or other means to convey particulars of sound. This is beneficial in that broad categories of pronunciation can be referred to and identifying details of one’s voice are hidden. However, the inability to hear other participants online has its disadvantages as well. Ambiguities and confusion can arise in trying to convey pronunciations with letters. The absence of sound or

51. A MetaFilter 5-minute edit window for all comments (but not posts) was implemented in October, 2012 and allows participants to edit their comments for up to five minutes after publicly publishing them online.

visible social categories can prompt FtF contextualization cues to be conveyed through other means (e.g., punctuation, emoticons, direct explanations). Regardless, data show that MeFites have repeatedly overcome these hurdles in their ongoing discussions of the M-Set since the beginning of these pronunciation debates back in 2000.

6.4.3.1 *Message Chains Through Quoting and Response*

The common way message receivers become senders on MetaFilter subsites is by responding to, quoting, or sharing the original sender's message (e.g., on other parts of the site, on other sites, or offline). The most straightforward example of receivers becoming senders through this type of stance alignment is by directly quoting a message, where the receiver of the original message perpetuates the message chain by copying and pasting the original sender's message and optionally adding content to update, modify or otherwise respond to the original message.

Example 4. Comment #19 in 2010 Survey MetaTalk Discussion Post

Done.

None of those pronunciations were accurate for me. I think of it as My, not Meh.

Good luck with the research.

posted by 26.2 at 5:34 PM on March 24, 2010 [3 favorites +] [!]

Example 5. Comment #81 in 2010 Survey MetaTalk Discussion Post

None of those pronunciations were accurate for me. I think of it as My, not Meh.

This.

posted by Vectorcon Systems at 6:05 PM on March 24, 2010 [+] [!]

When a comment is directly quoted within the same context as the one in which the original message was read, the new message can seem repetitive to readers. However, this repetition may have an intended affect (sic), as illustrated in **Example 5** with the quoted portion pithily followed by "This." The stance containing a pronunciation preference is repeated — and perpetuates the message chain — by quoting the original author of the comment verbatim. The use of a deictic demonstrative 'this' provides a minimal response that is intentional and conveys the idea that the previous commenter already expressed the second commenter's stance entirely, i.e., that they are both in complete agreement. Furthermore, the use of "This."⁵² in response to online content has been in use in internet forums for many years and enregistered as an

52. "This." variants include, but are not limited to: this., this, and ^This. The use of the caret is intended to visually point to the content that 'this' deictically refers to.

example of internet slang since approximately 2009.⁵³

The ways in which message chain links are replicated (i.e., receivers of messages become senders of messages) through responding, quoting and sharing are manifold. In most instances, the content that is replicated or shared is done so for the purpose of allowing the receiver-turned-sender to update, modify or add to the original message. In the case of the receiver wishing to align with the original sender, the original sender's stance is highlighted in some way and the message chain is positively affected. In other instances, the receiver may wish to replicate the message in order to highlight aspects that are problematic with the original message or perhaps make an example out of the original sender. Regardless, in these latter cases, the underlying intent is to stop the message chain from continuing as is, i.e., to change the message to be in line with the receiver's perspective, and possibly to attempt to perpetuate the message chain with this new message throughout the message chain network. Repeated instances of changes in message chains over time can result in more general shifts in perception of the variables. That is, as the messages conveyed along message chains change, so do indexical associations, leading to changes in the enregisterments of variables or a variety.

6.4.3.2 *Reinforcing Message Chain Links Through Metapragmatic Activities*

Many MeFites participate in message chains indirectly through metapragmatic behaviors such as favoriting comments (or posts). The ability to engage in behaviors such as favoriting⁵⁴ are features that are available in many CMC environments and a particularly meaningful aspect of the MetaFilter CoP. These acts allow stances to be aligned with or replicated and commented upon, at differing levels of overtness and visibility to others.

Favoriting may strengthen message chain links in three distinct ways. First, favoriting is a visual notification to the message sender, letting them know that their message had been received and was supported by their peers. Heavily-favorited comments might even be automatically featured on the Popular Favorites tab of MetaFilter, where the comment is replicated, along with a display of the favorite count and a link to the comment in its original context. This may be read by a wider audience who may not have initially seen the comment in its original context, but learned about it through reading the Popular Favorites page, either on the site or through another means, such as an RSS feed reader. Secondly, favoriting is a signal to message receivers that the sender's comment had been positively evaluated by other MeFites and may be worthy of their attention. Thirdly, favoriting is a deliberate act of alignment by the message receivers who are doing the favoriting. To favorite a comment, for

53. The use of "This." as a minimal response has also been reified as such and has its own entry in the popular online meme repository, Know Your Meme, <http://knowyourmeme.com/memes/this>.

54. 'Favoriting' is MetaFilter-specific and not necessarily a universally-known term. Other CMC environments may refer to this metapragmatic behavior as 'liking', 'upvoting', 'adding', to name a few.

whatever reason(s), one must be consciously aware of a message exceeding a certain threshold according to one's own personal criteria of value and deem that comment worthy of public expression, as favorites are visible to others and MeFites are generally aware of this fact when they favorite comments or posts.

These characteristics of favoriting become a powerful mechanism in boosting the strength of message chain links (i.e., the influence and authority of the messages). While the effects of favoriting are both indirect in nature and hard to quantify in practice, the social function of metapragmatic behaviors such as favoriting should not be underestimated. This is especially important to consider in communities where such metapragmatic acts are widely recognized as serving a social function and having communicative value. For example, on MetaFilter, favoriting is perceived by many as a covertly prestigious act that may confer a modicum of social capital onto the comment or post author (see 2.3 *Capital*, p. 19). Furthermore, the ability to see and favorite posts and comments has been a repeated topic of overt commentary since its implementation on MetaFilter in May 2006, thereby raising metapragmatic awareness about the potential social meanings of favorites in general. Therefore, the metapragmatic act of favoriting may carry extra social significance in the context of its sociocultural history and current perceptions in the eyes of many MeFites.

6.4.3.3 *Summary of MetaFilter Site Participation and Enregisterment*

Even though quantitative analysis revealed that only very frequent visitation to some subsites was correlated with increased preference for 1b variants, it is the entirety of the practice that creates the environment for processes such as enregisterment to occur as such, each subsite and feature contributing to and shaping the whole. All of these features of MetaFilter onsite participation, from the persistent transcript to the ability to quote, share, edit and deliberately craft responses, comments and posts, provide mechanisms for enregisterment to occur more easily. Stances can be made, aligned with, responded to, and replicated. The public history of content available on MetaFilter means that the resulting message chains are both observable and able to continue long after their original message chain links were established.

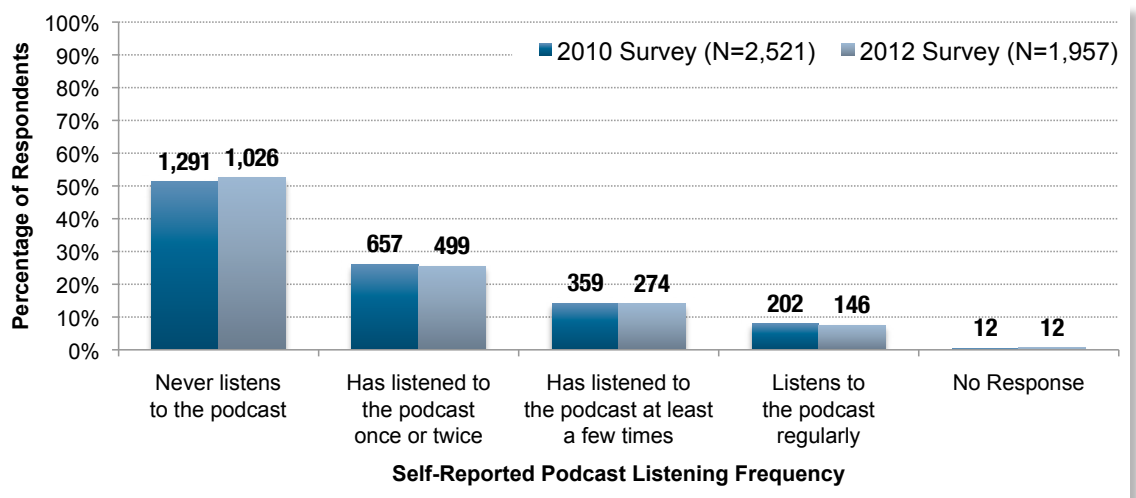
6.4.4 **Frequency of Podcast Listening**

The communication that takes place on the Podcast occurs in the medium of speech. However, the MetaFilter Podcast is a passive engagement recorded, prepared and broadcasted by authority figures within the community (MetaFilter moderators) to its audience (MeFites). Therefore, the podcast does not contain the interactive element that FtF conversation or other mediums of spoken communication might provide.

The content of the podcasts is discussion generated by notable posts and events occurring on MetaFilter. Discussions of pronunciation differences — especially as they relate to the M-Set or MeFite usernames — also frequently occur on the podcast. Additionally, some moderators prefer different pronunciations for the M-Set than others, and these variants can be heard by participants listening to the podcast.

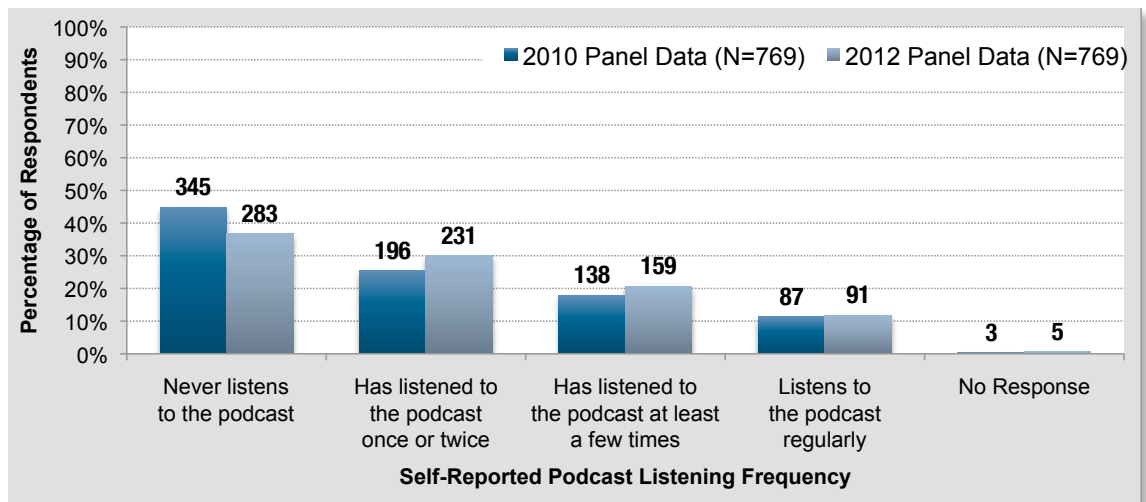
The podcast co-occurs with its own discussion post (on the Podcast subsite and also cross-posted to MetaTalk). There is usually a great amount of meta-commentary in the podcast posts, especially those concerning the M-Set pronunciation.

Figure 55. Self-Reported Podcast Listening Frequency — 2010 and 2012 Survey Data



$$\chi^2 = 1.315, df = 4, p = 0.8588$$

Figure 56. Self-Reported Podcast Listening Frequency — 2010 and 2012 Panel Data



$$\chi^2 = 11.065, df = 4, p = 0.0256$$

The majority of participants had never listened to the podcast at the time of each survey. The second largest groups comprised those participants who had listened to the podcast at

least once or twice, or at least a few times. The smallest contingent were those who listened to the podcast regularly, demonstrating the limited reach of the podcast and its perceived status as a more in-group activity than perhaps other aspects of the site.

Panel Data participants showed a significant change over time in podcast listening frequency that the Survey Data respondents did not. Panel Data participants also significantly differed from the Survey Data respondents in podcast listening patterns across both survey years (2010 significance level: $\chi^2 = 18.187$, $df = 3$, $p = 0.0004$; 2012 significance level: $\chi^2 = 56.563$, $df = 3$, $p < 0.0001$). Panel participants showed a 6% decrease in the ‘Never listened to the podcast’ category in 2010 as compared to the Survey Data, and a 16% decrease for that same category as compared to the Survey Data in 2012. Increases in all other more frequent podcast listening categories as compared to the Survey Data were observed across both survey years. This demonstrates yet another way in which the Panel Data participants are more deeply involved in the MetaFilter community, with more frequent podcast listening than the rest of the Survey Data participants.

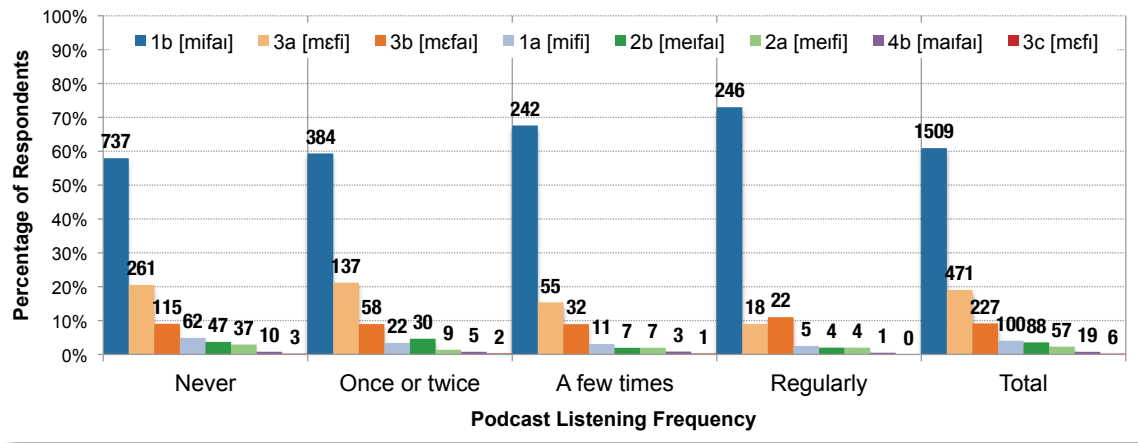
6.4.4.1 *‘MeFi’ Pronunciation and Podcast Listening*

The MetaFilter Podcast is one of the more in-group aspects of the site culture and the majority of participants do not listen to the podcast. However, those that do are made aware of current events in the community, notable posts, comments or users, in-jokes and moderators’ stances on these topics. Additionally, as it is a podcast, the discussions are spoken. This not only allows M-Set variants to be heard and discussed (in later discussions online) but those variants can be ideologically linked to authority figures (e.g., moderators) and their stances on this and other topics.

When there are guests on the podcast, one of the first questions they are usually asked is how they pronounce ‘MeFi’. Additionally, pronunciation (of the M-Set or of MetaFilter usernames) is generally at least a brief topic of conversation on almost every podcast. This results in podcast listeners’ increased meta-linguistic awareness and register socialization.

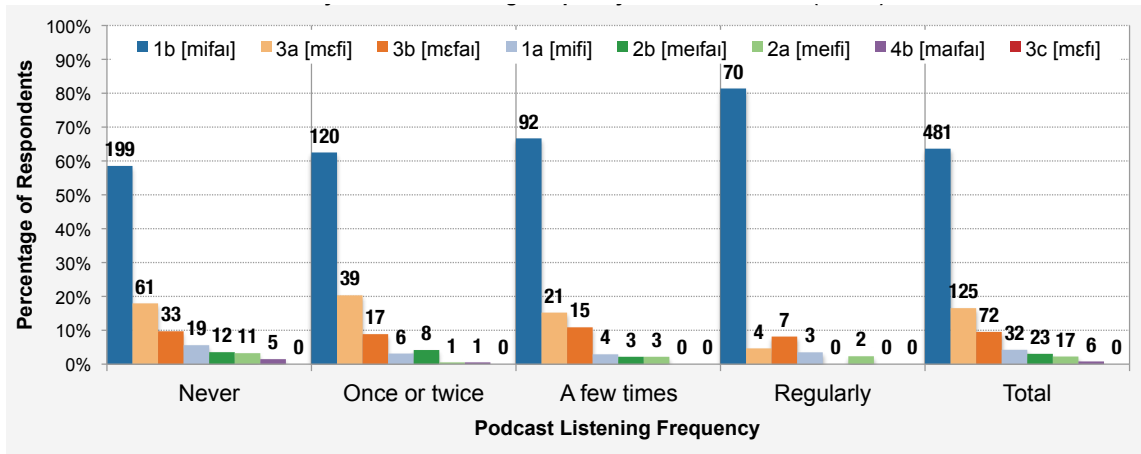
Similar to subsite visitation, it was hypothesized that those who listened to the podcast more frequently would trend toward increased preference for the 1b – [mifɑɪ] variant.

Figure 57. ‘MeFi’ Pronunciation by Podcast Listening — 2010 Survey Data, N=2,477



$\chi^2 = 24.699, df = 6, p < 0.0001$

Figure 58. ‘MeFi’ Pronunciation by Podcast Listening — 2010 Panel Data, N=756



$\chi^2 = 14.545, df = 6, p = 0.0241$

Podcast listening was significantly correlated with ‘MeFi’ pronunciation in both the 2010 Survey and Panel Data sets, but not in any of the 2012 data (not shown). Those who listened to the podcast regularly showed an increased preference for the 1b - [mifai] variant.

It not entirely clear why there was a significant positive correlation between podcast listening and ‘MeFi’ pronunciation in the 2010 Survey Data and no correlation in any of the 2012 Data. However, it is likely that this owing to a decrease in the quantity of the podcasts in recent years, as well as a decrease in the number and depth of discussions about how to pronounce the M-Set. These factors would have led to fewer message chains about the M-Set on the podcast in recent years, thereby making this type of engagement not a source of information about the debate or pronunciation variants (despite the fact that moderators on the podcast were likely to refer to MetaFilter as ‘MeFi’ as much as they did in previous podcasts).

6.4.5 Podcast Listening and Enregisterment

The audio podcast allows pronunciations and accents to be heard, possibly providing receiver-listeners with at least some information about the age, gender and geolinguistic background of authority figures in the community (specifically, moderators and their podcast guests). Additionally, other aspects of the MetaFilter register and site culture are directly addressed, allowing the message chains involving the M-Set to be contextualized and to co-occur among message chains related to other relevant matters in the community, such as notable posts, pronunciation of others' usernames, and changes to the site.

However, there are some features of the podcast that may influence enregisterment in limiting or negative ways. The podcast only occurs approximately once a month and is not advertised beyond being made available on the 'Podcast' tab on MetaFilter and published as a discussion post on MetaTalk. MeFites who do not visit those areas of the site may not be aware of the podcast or may forget to check for new audio content. As the survey findings revealed, the podcast was regularly listened to by only a very small, self-selecting audience of MeFites. This means that although the podcast was found to be a rich source of messaging about the M-Set, the reach of that messaging is very limited.

6.4.5.1 Podcast Message Chains

Links in message chains can be formed when MeFites listen to the podcast, as this is another way that messages containing stances about the M-Set are shared publicly. In these instances, site moderators (and sometimes podcast guests), refer to the M-Set or comment on the M-Set directly. Similar to commenting on the site, this is another example of a one-to-many sender/receiver relationship, whereby one or more speaker-senders on the podcast broadcast a message to an audience of listener-receivers.

In **Example 6**, MetaFilter moderators jessamyn and mathowie interview then newly-appointed moderator cortex as part of the second-ever podcast, posted to the site on February 23, 2007 and running to approximately twenty minutes in length. The excerpt in this example begins at 16:56.

Example 6. Podcast Transcript Excerpt: Episode #2

Note: parentheses denote overlapping speech; double parentheses denote metacommentary about the speech; double hyphens denote paralinguistic sounds.

- mathowie: Should we have, like, standard questions for people that are part of the site, like...
- mathowie: ("How did you find the site?")
- cortex: (I-I-I was thinkin' you should) ... that'd be cool if you did that. If you had like, you know
- cortex: (almost)
- mathowie: (yeah)

cortex: just like the three MetaFilter questions
jessamyn: well, how do you pronounce it is, uh, question number one
mathowie: (right, right)
cortex: (yeah) that's obviously going to be a big one
mathowie: (yeah)
cortex: It's-It's [mɛfaɪ]
mathowie: --audible sigh--
jessamyn: [mi:fɪts]? ((almost inaudible))
cortex: [mɛfaɪts], [mɛtəfɪltə], [æskmi]
mathowie: [mɛ...faɪts]? ((loud, drawn out pronunciation, sounds exasperated))
cortex: [mɛfaɪts] I dunno, it rolls off the tongue
mathowie: yeah, like a turd
jessamyn: (--laughter--)
mathowie: (--short laugh--)
mathowie: UM...so, how did you find the site?
all: --laughter--

Podcast URL: <http://metatalk.metafilter.com/13721/MeFi-Podcast-II-Electric-Buggaloo>

Example 7. Comment #32 in MetaFilter Podcast Post: Episode #2

Thanks for interviewing a MeFite who knows how to pronounce things properly!
posted by trip and a half at 4:01 PM on February 23, 2007 [+] [!]

The solidarity the moderators displayed regarding the questions they should ask guests implies that the pronunciation of the site name is both a well-known topic and an important “getting-to-know-you” question for MeFites. They assert that pronunciation might be the definitive question to ask MeFites first, perhaps to discern quickly where they stand on issues of mock importance.

This podcast generated a relatively large amount of commentary (101 comments in total). For many, this was only the first or second time site participants had heard the voices of their fellow MeFites. Several comments noted this and some went further to comment on pronunciation of the M-Set; see **Example 7**.

The early podcasts raised new meta-linguistic awareness about variability and pronunciation, which became the subject of commentary both on the podcast and in the related posts on the site. This continues today, although much less frequently, as both moderators and MeFites have become more accustomed to each others' voices, pronunciation preferences, and stances on a debate that has lasted over a decade and may have possibly become tired of discussing. This may help explain why the 2012 data did not show any significant correlations between podcast listening and 'MeFi' pronunciation. That is, the novelty of the podcast, discussions about variability, and the subsequent meta-linguistic awareness that was raised had effected pronunciation distributions in 2010, but those effects

had essentially “worn off” by 2012, as MeFites became interested in other topics.

Regardless, the novelty of the podcast and all the issues surrounding the translation of ideas from a text-based medium into a spoken one⁵⁵ has meant that this method of social engagement with the community is fertile ground for the establishment of message chains propagating messages about the M-Set. As **Example 6** and **Example 7** have shown, the messages from the podcast can cross modalities when the stances they contain are replicated on the site or otherwise aligned with through commentary, such as when a MeFite suggests that cortex “*knows how to pronounce things properly*”. The crossover between ideas discussed in the podcast versus online discussion posts provide a good example of how the links in a message chain are not relegated to the domains or modalities in which the receivers originally acquired their messages. For example, a person listening to the podcast may be a receiver of a moderator’s stance about the M-Set, and that person may in turn become a sender of that stance (or an opposing stance perhaps) in an entirely different domain and modality.

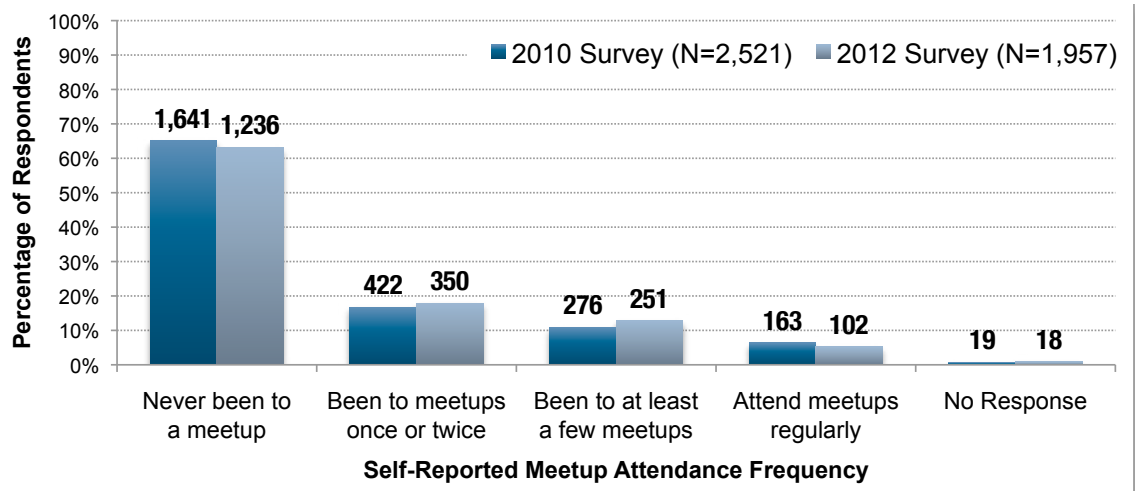
Chain links being able to cross over into other modalities (text or speech) or domains (podcasts, meetups or online, etc.) is an important feature of message chains. This crossover connects all of the ways that MeFites can participate in the community and allows the transfer of ideas to occur for those who only access certain aspects of the practice. This crossover allows messages to be disseminated to a much wider audience, which is especially relevant for deeply in-group aspects with small audiences, such as the podcast.

6.4.6 Frequency of Meetup Attendance

Results from survey questions regarding meetup attendance are shown in **Figure 59**, giving a general idea of how many MeFites attend meetups and how frequently they do so.

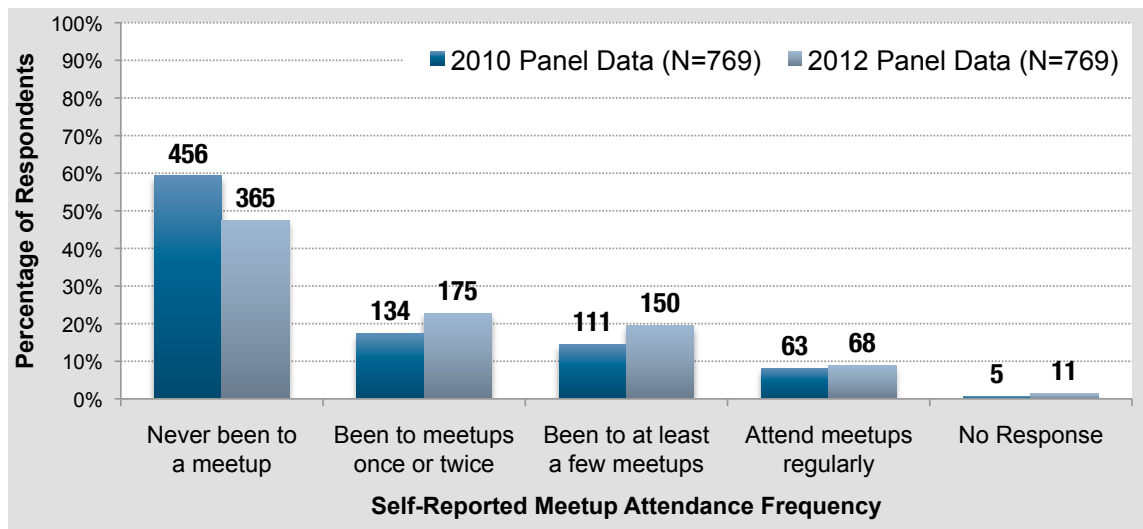
55. Podcasts almost always feature at least one moderator having pronunciation difficulty when referring to a MeFite’s username, usually when trying to give the MeFite credit for making a noteworthy post on the site.

Figure 59. Self-Reported Meetup Attendance Frequency — 2010 and 2012 Survey Data



$\chi^2 = 8.075, df = 4, p = 0.0889$

Figure 60. Self-Reported Meetup Attendance Frequency — 2010 and 2012 Panel Data



$\chi^2 = 23.795, df = 4, p < 0.0001$

In the 2010 Survey Data, 1,641 respondents (65%) reported never having been to a meetup. In 2012, 1,236 respondents (63%) reported the same. A small percentage of the population reported attending meetups regularly (6% in 2010 and 5% in 2012). The differences in meetup frequency between the 2010 and 2012 surveys was not significant.

However, Panel Data participants significantly differed from the Survey Data in meetup attendance across both survey years (2010 significance level: $\chi^2 = 13.29, df = 3, p = 0.004$; 2012 significance level: $\chi^2 = 53.913, df = 3, p < 0.0001$). Panel Data participants showed a 6% percentage point decrease in the ‘Never been to a meetup’ category in 2010 as compared to the Survey Data, and a 16% decrease for that same category as compared to the Survey Data in

2012. The Panel Data participants showed increases in all other meetup attendance categories as compared to the Survey Data, across both survey years. Similar to podcast listening results, this demonstrates another way in which the Panel Data participants differed from the overall survey sample; they were more engaged with additional, non-text-based MetaFilter activities, such as meetups, and this engagement had increased over time.

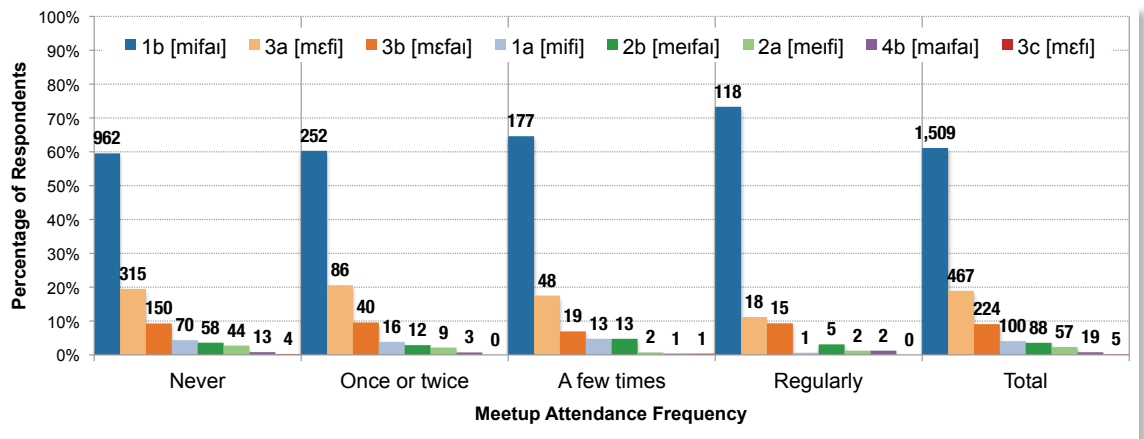
The 2012 decrease in the ‘never been to a meetup’ category by Survey Data participants (and the significant increase in meetup attendance by Panel Data participants) may be accounted for by a couple of notable events which occurred between 2010 and 2012, after the 2010 survey and before the 2012 survey. On June 24, 2010 the IRL subsite was launched, dedicated to discussing, sharing and planning meetups (see **3.2.1.4 IRL and Meetups, p. 52**). The number of MetaFilter-related events skyrocketed since the launching of IRL, allowing MeFites to engage with each other in new ways and more frequently offline (see **Figure 45. Social Engagement Factors and the Enregisterment Timeline, p. 163** for an infographic illustrating this increase). In addition to IRL, a commemorative 12-year MetaFilter anniversary website, MetaFilter Memories, was also launched (July 14, 2011), allowing MeFites to celebrate by sharing memories and stories of how they found MetaFilter.

These events have had a huge positive impact on the community, bringing people together both online and offline. While it is difficult to quantify, it is assumed that these events have fostered greater social engagement and increased bonding social capital between MeFites. Additionally, these events have likely drawn others into the community, through discussions MetaFilter members have had with members of their social networks about their interests and attendance of said events.

6.4.6.1 *‘MeFi’ Pronunciation and Meetup Attendance*

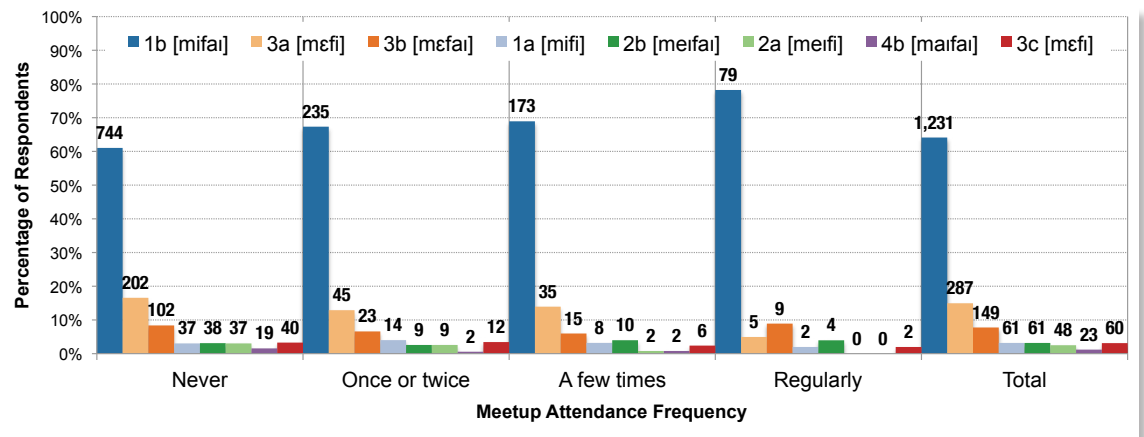
It was hypothesized that more socially engaged members of MetaFilter (i.e., those that extended the social practice by meeting with other MeFites offsite) would show a statistically significant difference in their distribution of ‘MeFi’ pronunciation preferences. This is largely due to discussions about the M-Set that often arise at meetups, where participants may hear spoken variants for the first time, or variants that differ from their own chosen pronunciation. This has a positive effect on register socialization (awareness of the linguistic forms and norms specific to MetaFilter; see **2.5.1.1 Register Competence and Register Socialization, p. 26**) and meta-linguistic awareness (knowledge of sociolinguistic variation for forms in the register).

Figure 61. ‘MeFi’ Pronunciation and Meetup Attendance — 2010 Survey Data, N=2,469



$$\chi^2 = 25.125, df = 15, p = 0.0483$$

Figure 62. ‘MeFi’ Pronunciation and Meetup Attendance — 2012 Survey Data, N=1,920



$$\chi^2 = 17.7624, df = 6, p = 0.0068$$

As demonstrated previously by the Survey Data, most MeFites do not attend meetups. However, the Survey Data also show that those who did attend meetups regularly had a higher preference for 1b – [mifai] variants. This demonstrates an effect of offline social engagement on linguistic behavior, where MeFites who engage with others ‘in real life’ are likely to hear and discuss the M-Set more than those who do not.

While meetup attendance overall was significantly higher among the Panel Data participants than in the Survey Data respondents, there were no significant correlations between meetup attendance and ‘MeFi’ pronunciation in the Panel Data.

6.4.7 Meetup Attendance and Enregistration

Similar to podcasts, FtF conversation allows pronunciations and accents to be heard. Unlike the podcast, however, the messages conveyed in this medium occur between fellow MeFites, and are not broadcast by authority figures in a one-to-many relationship. As this

type of social engagement is highly interactive, ideas about the M-Set are exchanged rapidly, as receivers become senders in quick succession through a continued dialogue containing stances, opinions and sometimes playful argumentation. Additionally, the social categories (e.g., age, location, gender) of participants are visible, allowing variants and stances about the variants to be more readily associated with social groups or speaker characteristics.

Despite these obvious advantages, however, the participants themselves are a self-selecting and mostly geographically-local group. This may introduce several biases in the exchanges that occur at meetups. Also, the information acquired through these exchanges, whereby participants can link pronunciations to observable speaker characteristics, are not easily conferrable about with the rest of the community, as even comments on the site about what occurs at meetups is more likely to refer to stances of the participants rather than information about their membership of particular social groups. Essentially, meetups provide a MetaFilter environment where message chains occur with great frequency and participant demographics are easily linkable to stances and pronunciations, but these are local distributions and not necessarily generalizable over the entire community.

Results from an earlier study by Sessions (2010), also involving MetaFilter, also confirm the positive effect of meetup interaction on social cohesiveness. Sessions demonstrated that maintaining relationships online and offline (i.e., by attending meetups) *“enhances attendees’ engagement with the online community as a whole, strengthens ties to other attendees, and contributes to the creation of bonding social capital”* (Sessions, 2010, p. 375). These advantages of the medium, in this case, resulted in greater consensus on a linguistic form — MeFites who attended meetups regularly showed increased preference for 1b – [mifaɪ], [mifaɪt] variants.

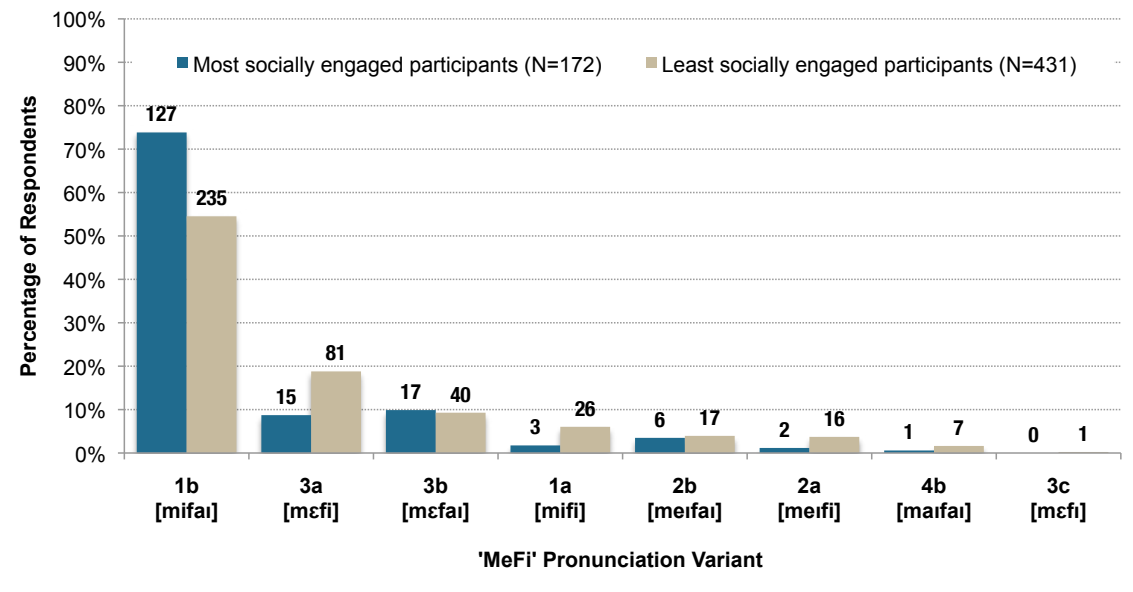
However, as suggested previously, the social information acquired and the effects of meetup attendance do not transfer easily to the community-at-large. While individuals may benefit from meetups by learning about each other, including the observation of speaker characteristics, these are effects experienced at the individual level. Community members who do not attend meetups — the majority of participants — do not directly benefit from the social bonding that occurs offline at meetups, nor are they easily able to participate in message chains that occurred there, as receivers or otherwise. This is to say, while select individuals may benefit from meetups, the group as a whole generally does not (Sessions, 2010).

6.4.8 The Effect of Social Engagement on ‘MeFi’ Pronunciation

Several social engagement factors have been analyzed separately thus far. However, a comparison between participants who engage in different types of frequent MetaFilter participation versus those who engage minimally reveals the strong effect of community influence. The 2010 Survey participants who had attended at least a few meetups or more, had

listened to at least a few podcasts or more, and visited MetaTalk at least every other day were grouped into a category labeled ‘Most socially engaged participants’. Those who had never attended a meetup or listened to a podcast and visited MetaTalk once a week or less were grouped into a category labeled ‘Least socially engaged participants’. These two groups were then compared, as shown in Figure 63.

Figure 63. Social Engagement and ‘MeFi’ Pronunciation – 2010 Survey Data, N=603



$$\chi^2 = 23.789, df = 6, p = 0.0006$$

There were no significant differences between the 2010 and the 2012 data; only the 2010 data are shown here. The effect of social engagement on pronunciation choice is most noticeable with respect to the two most popular pronunciation variants. Participants who participated the most, in varied ways, had the most exposure to message chains containing stances and other information about the M-Set. The influence of this is evident by the nearly 20 percentage point increase in preference for 1b – [mifɑɪ] by this group as compared to the least socially engaged group. The least socially engaged participants did not have access to such information, and therefore their choices were not influenced by such factors; their choices were likely largely influenced by their linguistic backgrounds and other demographic factors.

6.4.9 Summary of Social Engagement Factors

Four measures of social engagement were analyzed: year of joining MetaFilter, subsite visitation frequency, podcast listening, and meetup attendance. It was hypothesized that social engagement would be positively correlated with increased preference for the most popular variants of the M-Set.

Participants' year of joining MetaFilter had no significant bearing on pronunciation outcomes. Additionally, the data sets did not differ with respect to when participants joined the site.

Subsite visitation was correlated with 'MeFi' pronunciation, but this effect was inconsistent across the data sets, survey years, and subsites. However, a general trend was observed, whereby increased frequency of visitation to more MetaFilter- and MeFite-focused subsites (MetaTalk and AskMetaFilter, respectively) led to a stronger preference for the 1b - [mifaɪ] variant.

Podcast listening was found to be significantly correlated with 'MeFi' pronunciation in 2010, but not in 2012. Greater frequency of podcast listening was linked to an increased preference for the 1b - [mifaɪ] variant, demonstrating the potential effect of social influence and authority figures on linguistic behavior.

Meetup attendance was significantly correlated with 'MeFi' pronunciation in both survey years, but not both data sets. Greater frequency of meetup attendance was linked to an increased preference for the 1b - [mifaɪ] variant in the Survey Data, but not in the Panel Data, suggesting that meetup attendance was a consistent influence on pronunciation outcomes over time, but this influence may not necessarily effect all groups of participants in the same way.

These results show that both the type and frequency of social engagement can influence participants' pronunciation choices. In general, social engagement which was more frequent and allowed participants to observe or participate in discussions about MetaFilter culture or the M-set was associated with increased preference for the 1b - [mifaɪ] variant, while engagement which was infrequent or not inwardly focused on MetaFilter did not seem to influence outcomes either way.

6.5 M-Set Stances

It has been shown throughout this chapter that social engagement on MetaFilter has been crucial in further the process of enregistration for the M-Set variables. The accretion of particular types of stances have led to M-Set variants acquiring stereotypes and indexical associations about their use. This final section of the data presentation will address various categories of M-Set stances that have not been previously covered, to showcase the common associations MeFites have made, leading to the enregistration of the variants as having those indexical values.

Stance categories from survey rationales which have already been covered include those presented in 5.3.3 **Other Measures of Metalinguistic Awareness**, p. 124, 5.3.4 **Analogies in Survey Rationales**, p. 127, and various examples given throughout the current chapter

to illustrate the way social engagement factors influence pronunciation choice. Additional stances, made privately in survey rationales and publicly in online discussions about the M-Set, will be presented in this section. At the end of this chapter, indexical field maps for the most popular pronunciation variants of ‘MeFi’ will be shown, summarizing all of the data results and outlining some of the common associations many MeFites have about the variants.

6.5.1 Stances From Survey Rationales

The stances presented here include three categories drawn from survey rationales and concerning key attitudes shared privately about the M-Set. The first category involves evaluative stances (positive and negative), such as feeling that some pronunciations were “right” or “wrong” and “good” or “bad”, etc. The second category concerns some participants’ avoidance of the M-Set. Lastly, the third category explores the influence of authority figures on making a pronunciation choice.

6.5.1.1 *Positive and Negative Evaluations of the Variants*

Approximately one-quarter of all survey rationales contained evaluative statements about a preferred variant or another variant (or variants). These often included prescriptivist stances, such as a variant being “right” or “wrong”, “(in)correct”, “(in)accurate” and “(il)logical”. Also included in this category were expressive stances such as “cute”, “weird”, “natural”, “simple” and “easy”, to name a few.

While some rationales were straightforward to code along this dimension, the majority were not. This is because rationales may contain positive, negative, neutral and sometimes ambiguous stances about aspects of a chosen variant, as well as similar evaluations about aspects of other variants. If a rationale contained any evaluation of a variant (whether it was preferred by the participant or not), it was included in the category. Three sub-categories were then defined: rationales that contained only a positive evaluation of any variant, rationales that contained only a negative evaluation of any variant, and the third containing both rationales that were ambiguous or contained a positive and negative evaluation of any variant. In general, most rationales with positive evaluations referred to the variant(s) that the participants had chosen for themselves and negative evaluations referred to other variants. Rationales that contained both positive and negative evaluations usually highlighted features of the respondents own preferred variant(s) as well as other variants.

Table 16. Positive and Negative Evaluations of Variants

Type of Evaluation	2010 Survey (N=1,974)	%	2012 Survey (N=1,472)	%
Contained a Positive Evaluation Only	237	55%	216	56%
Contained Both Positive and Negative Evaluation	96	22%	77	20%
Contained a Negative Evaluation Only	99	23%	96	25%
Total	432	100%	389	100%
Percent of Surveys with Rationales	22%	22%	26%	26%

22% of the 2010 surveys with rationales and 26% of the 2012 surveys with rationales contained an evaluative stance (positive and/or negative). The majority of these evaluations were positive, usually referring to the participant's own choice of variant. The remainder were roughly equally split between only negative evaluations (usually referring to other variants) or mixed evaluations (both positive and negative sentiments).

There were no significant differences in the number of positive, mixed or negative evaluations given by pronunciation group or across survey years. However, the types of evaluations given by each pronunciation group varied.

Common positive evaluations for the 1a – [mifi] variant were that it “*makes sense*”, “*is right*”, “*logical*” and the “*cutest*”. Favoring rhyming syllables was also frequently mentioned.

The most common positive evaluative stances made by the 1b – [mifaɪ] group were mostly prescriptive in nature. These included statements that this variant “*looked (or seemed, felt, sounded) right (or correct, accurate, proper)*” to them. Many also made a distinction between this variant being “*right*” versus the 3 – [mɛ-] variants being more “*logical*” or “*technically correct*”, since it was claimed that they emulate the full form, ‘MetaFilter’. Negative evaluations from the 1b – [mifaɪ] group mostly consisted of statements about how the other variants seemed “*wrong*” or “*weird*”, with special attention paid to the 3 – [mɛ-] pronunciation variants as “*weak*”, “*too cute*” or other negative evaluations.

Over half of the 2a – [merfi] and 2b – [merfaɪ] pronunciation groups who made positive evaluations also cited the influence of a foreign language or variety. For example, one participant's rationale stated, “*Sounds right (for certain values of right) to my Swedish ears.*” In this instance, the positive prescriptivist stance is qualified by a linguistic context that both supports their choice and an awareness that their pronunciation is right for them but may not be the norm (i.e., recognizes that other participants may have different criteria for determining what is ‘right’).

While the majority of participants from the 3 – [mɛ-] pronunciation groups simply stated that they felt their pronunciation was “*right*”, “*correct*” or “*logical*”, many of them elaborated on this by saying that they felt their pronunciation choice was “*affectionate*”,

“cute” or favorably “diminutive”. The majority of the [mɛ-] group participants with evaluative stances in their rationales chose the 3a – [mɛfi] or 3b – [mɛfaɪ] pronunciations and referred to the 3c – [mɛfɪ] pronunciation in their negative evaluations (where negative evaluations were made). In other words, these participants often felt that basing their pronunciation on the full form ‘MetaFilter’ was “correct”, but applying that logic to the final syllable of ‘MeFi’ took ‘correctness’ too far and sounded “weird”, “awkward” or “wasn’t easy to say”. Other negative evaluations referred to their perception of the majority preference for 1b – [mifaɪ] but simply stated that this was “wrong” to them.

The 4b – [maɪfaɪ] pronunciation group was nearly categorical in their perceptions of their chosen pronunciation seeming “illogical”, “wrong” or making “no sense” to them. Negative evaluations were nearly always applied to participants’ own pronunciation choice; other pronunciations were rarely mentioned, and if they were it was almost always to cite a positive evaluation attributed to that variant. Positive evaluations for the 4b – [maɪfaɪ] variant included favorable associations with the semantics of ‘my’ (indicating possession and belonging to the community) and rhyming, making this pronunciation seem “natural” and “easy to say” for these participants, even if they could not find a linguistic justification for the vowels (which were often specifically mentioned by this group).

These positive and negative stances towards the variants lend insight to the salient motivations participants have for their choice(s). In addition, the negative evaluations reveal the confidence (or lack thereof) that participants may have in their choices. These findings also support the earlier results concerning how strongly participants felt they would use their preferred variant exclusively (see 5.3.1 *Exclusivity of Use of Preferred M-Set Variants*, p. 118).

6.5.1.2 *Avoidance of the M-Set*

Many participants stated that they never say ‘MeFi’ aloud, either preferring the form ‘MetaFilter’ or none at all (usually because they never have reason to speak the form aloud). Many also shared negative evaluations of the pronunciation variant(s) as their reason for avoidance of ‘MeFi’ in spoken environments.

In 2010, 7% mentioned that they never say ‘MeFi’ aloud; this was 8% in 2012. This increase in 2012 was not significant. Inability to make a pronunciation choice was seldom mentioned as a reason for avoidance of the short form, but dislike of the short form in general was frequently mentioned and usually related to a pronunciation variant they had heard and found displeasing. Regardless, these participants often mentioned being interested in the debate about pronunciation, at the very least demonstrated enough interest to complete a survey. This shows how even participants who avoided the term may have been aware of the ongoing discussion and took a stance in the matter, even if only to express their aversion.

6.5.1.3 *Citing an Authority in 'MeFi' Rationales*

Many participants mentioned choosing their favored pronunciation based on how they heard a moderator or another MeFite pronounce 'MeFi'. These instances were considered appeals to authority in making a pronunciation choice. These appeals also included other sources such as family or friends, hearing the M-Set on a podcast or at a meetup, and learning about others' pronunciations through MetaTalk discussions.

There was no significant bias toward any pronunciation group in citing authority and the number of authority mentions did not significantly differ between survey years (5% of rationales in 2010; 4% in 2012). The two most common authorities on 'MeFi' pronunciation were other MeFites (usually in general, but sometimes specific MeFites were mentioned) and moderators (either on the podcast or at meetups). Other sources includes people who introduced the participant to the site, family members and friends or coworkers.

This highlights the social component involved in making a pronunciation choice. Participants who listened to the podcast, interacted with other MeFites offline or were introduced to MetaFilter through friends or family seem to be overtly influenced by those interactions. Additionally, participants in 3% of both years' survey rationales stated that they would change or base their pronunciation choice on a moderator's or majority opinion, further demonstrating the importance of social influence and authority.

6.5.1.4 *Summary of Stances in Survey Rationales*

The stances covered so far in this thesis have addressed ways in which participants demonstrated meta-linguistic awareness (or lack thereof), analogies used in describing and determining pronunciations of the M-Set, positive and negative evaluations of the variants, avoidance of using the M-Set in speech, and the influence of other authority figures on pronunciation choices. While there was much overlap in the types of stances given in these private survey rationales with those made in public comments online in posts discussing the pronunciation of the M-Set, many important differences were observed as well. The next section addresses some of the common stances in public discussions about the M-Set.

6.5.2 **Public Stances From MetaTalk Posts about the M-Set**

Qualitative data from MetaTalk posts about the M-Set and surveys will be presented in this section, demonstrating some of the many types of stances MeFites display via public comments online and private comments in surveys. Through these stances, MeFites express a range of opinions, often using humor to make points, assert positions or characterize the M-Set, MeFites and the MetaFilter community.

Two MetaTalk posts are drawn upon here. The first, posted in 2006 by a MeFite with

an idle curiosity about the pronunciation of MeFi⁵⁶, garnered 102 comments by 63 different MeFites. The second post was made by myself on March 24, 2010. The purpose of this MetaTalk post was to supplement the survey and to allow MeFites to publicly discuss matters concerning the M-Set. This post generated 472 comments from 187 different MeFites. A third MetaTalk post also explicitly centering around the pronunciation of the M-Set was made on August 22, 2012, by myself, to supplement the 2012 data from the survey. The discussion that resulted from this post was mostly concerned with previous discussions, surveys and results about the research. It did not contain a substantial amount of new information to inform the data analysis here, and is therefore not included in this presentation of results.

It should also be noted that MetaFilter favorites were not added as a feature to the site until approximately two months after the 2006 MetaTalk post was made. Therefore, any favorites received on comments or the post itself would have likely been added by users after the post had closed. Due to this, favorites cannot be reliably used as a metric for the 2006 post to gauge agreement or any of the other ways favorites are used by MeFites during the period of time when posts are open.

Comments from the 2006 and 2010 MetaTalk posts were coded and will be discussed in the next pages as follows:

1. Comments in which MeFites stated their preference for a pronunciation variant.
2. Comments in which MeFites took a stance of 'correctness' about pronunciation.
3. Comments in which MeFites made reference that indexed people, places or events in pop culture relating to the pronunciation of the M-Set.
4. Comments in which MeFites referenced a MetaFilter in-joke.

6.5.2.1 *Comments Referencing a Pronunciation Variant*

In the 2006 MetaTalk post, 57% of the comments stated a variant choice for 'MeFi', as compared to 14% in the 2010 MetaTalk post. This difference was expected, as the 2006 post specifically asked MeFites about their pronunciation, whereas the 2010 post was made as a general discussion about the surveys and the M-Set. It is still notable that 14% of commenters in the 2010 post offered their pronunciation choice, despite not being asked to do so. This allowed public commentary about the variants to ensue.

Comments stating preferences for particular variants frequently employed analogies or 'eye-dialect' spellings to convey pronunciation. The following two examples show some of the different ways in which this can be achieved.

56. MeFite heatherann posted the MetaTalk thread on March 2, 2006; she was studying linguistics at the time, but was not specifically focused on MetaFilter or the pronunciation of the M-Set in her research.

Example 8. “meh”-fee versus mee-fie

everyone around me here says “meh”-fee. But I don’t feel so “meh” about MeFi and much prefer the mee-fie pronunciation.

posted by whatzit at 6:23 PM on March 2, 2006 [+] [!]

In this data example from the 2006 MetaTalk post, eye dialect spellings were used not only to convey pronunciation, but also to signal evaluative stances based on semantic associations. The 3a - [mɛfi] pronunciation was written as ‘meh’ plus ‘fee’, allowing MeFite whatzit to eschew the pronunciation based on the connotations of ‘meh’ and in spite of it being what “*everyone around me here says*”.

Example 9. Analogous Words Plus Evaluation

Me? Fie!

Me fight anyone who pronounce it differently.

posted by Eideteker at 6:12 AM on March 2, 2006 [+] [!]

In **Example 9**, analogous words were used to explain pronunciation and to give contextualized justification for those choices. Additionally, the semantic associations of ‘me’, ‘fie’ and ‘fight’ helped express how strongly MeFite Eideteker may have felt about those variants, i.e., “*me fight anyone who pronounce it differently*”.

6.5.2.2 Comments Containing a ‘Correctness’ Stance

A substantial portion of the public commentary in the MetaTalk posts contained an evaluation about the ‘correctness’ (or lack thereof) of a particular variant. In 2010, 21% of the comments made in the post exemplified this type of prescriptivist stance. In the 2006 post, 15% of the comments alluded to such a stance.

This difference can be partially explained by the context in which these posts occurred. In 2006, the debate over the pronunciation of the M-Set was less visible than in 2010 — in fact, it was this post which raised community awareness in a public, site-wide way for the first time (as opposed to passing commentary in other posts previously). Therefore, while pronunciations were expressed in the post, judgments about those choices were not passed in the same way or to the same degree that it had been in the years leading up to that point.

Example 10. ‘MeFi’ by Association

cortex is absolutely wrong on this issue.

posted by desjardins at 5:35 PM on March 24, 2010 [7 favorites +] [!]

In **Example 10**, MetaFilter moderator cortex is referred to as an advocate for the 3b – [mɛfɑɪ] pronunciation (cortex has stated in various posts and podcasts that the 3b – [mɛfɑɪ] variant is the one he prefers). By stating “*cortex is absolutely wrong*”, judgement is passed on his chosen variant as well.

Example 11. Indirect ‘correctness’

I always thought “Meh-fee”

Then I started going to meetups and was roundly corrected to “Mee-fie”

posted by scarabic at 6:05 AM on March 2, 2006 [+] [!]

In some instances, the correctness stance was indirect, meaning that the commenter did not overtly state an opinion on whether they personally felt a variant was ‘right’ or ‘wrong’, but a prescriptivist-style stance was conveyed nonetheless. In **Example 11**, MeFite scarabic gives a very short narrative about ‘MeFi’ pronunciation in which he relates how he was told at a meetup that his preferred pronunciation was incorrect. It is unclear whether scarabic agreed with those who corrected him, but the act of sharing this story indirectly reports the stances of others at the meetup (a multi-modal message chain), i.e., that the 3a – [mɛfi] pronunciation is ‘wrong’.

6.5.2.3 Comments Referencing a Person, Place or Thing of Cultural Significance

The number of cultural references made in the two posts were about even; they accounted for 53% of comments in the 2006 post and 52% in the 2010 post. This equated to just over half of the comments in either post making mention of a person, place or thing of cultural significance.

Example 12. “Me Phi Me” Hip-Hop Performer and Musical Group: 2006 Post

Me-phi (long i sound).

Remember Me Phi Me, the “fraternity of one” hip hop guy?

posted by mathowie at 6:12 AM on March 2, 2006 [+] [!]

In the comment in **Example 12**, posted by owner and MetaFilter creator Matt Haughey (username mathowie), the early 1990’s American rapper Me Phi Me was referenced to explain mathowie’s pronunciation of ‘MeFi’. While the orthography of <phi> and <fi> differ, the analogy was made through the similar possible phonetic mappings of those letters to the segments [fɑɪ]. This comment may also confer authority on the 1b – [mɪfɑɪ] variant by the fact that the site owner was publicly asserting his pronunciation preference.

Example 13. “Let’s Call the Whole Thing Off” Song by the Gershwins: 2006 Post

So if you go for oysters and I go for ersters, I’ll order oysters and cancel the ersters

posted by Astro Zombie at 7:15 AM on March 2, 2006 [+] [!]

In **Example 13**, MeFite Astro Zombie referenced a classic song “Let’s Call the Whole Thing Off”, popularized by the lyrics “*You say toe-may-toe / I say toe-mah-toe.*” In this comment, a pronunciation variant was not asserted, but an evaluative stance about the debate over the M-Set was made through lighthearted allusion. In this way, the comment served to help further establish the pronunciation of the M-Set as a community shibboleth, rather than aiding the enregisterment of any particular variant.

Example 14. “Raymond Luxury Yacht” in Monty Python’s Flying Circus: 2006 Post

It’s pronounced “me throat-warbler mangrove”.

posted by Mr. Bad Example at 5:39 PM on March 24, 2010 [7 favorites +] [!]

Similar to **Example 13**, **Example 14** showcases an evaluative stance about the pronunciation debate. In the referenced Monty Python skit, a talk show host interviews a pretentious dermatologist character named Raymond Luxury Yacht, who has a comically oversized nose. During introductions, Mr. Yacht interrupts the host to explain that his name is spelled ‘Raymond Luxury Yacht’, but pronounced ‘Throat Warbler Mangrove’. At this point, the host remarks, “*You are a very silly man and I’m not going to interview you.*”

The use of this skit to make a joke about ‘MeFi’ pronunciation functions by drawing a parallel between a ridiculous character and discussions about the identity of MetaFilter. This allowed the evaluations associated with the characters in the skit to be temporarily ‘borrowed’, for reappropriation onto the ‘characters’ involved in the pronunciation debate (e.g., an anthropomorphized ‘MeFi’ figure and those who identified with it). Therefore, through this cultural comparison, evaluations of the debate as ‘humorous’ or possibly even ‘ridiculous’ are implicitly asserted. The comment also received seven favorites, indicating that some people likely recognized the reference and/or agreed with the comment’s sentiment.

Example 15. “Family Circus” Comic Strip in the Sunday Funnies: 2010 Post

I’ve always been a MEE-FAI, but I heard someone calling it Meffy, rhyming with that kid from Family Circus, and I think that’s funny and charming.

But wrong.

posted by dirtidirt at 5:32 PM on March 24, 2010 [30 favorites +] [!]

Example 15 illustrates the complexity of stance-taking, as several types of evaluative statements are expressed in a single comment. Besides asserting his own pronunciation variant (i.e., “*I’ve always been a MEE-FAI*”), MeFite dirt dirt references a popular US-based Sunday comic strip from the 1960’s called “Family Circus”, which features a child named Jeffy (among several others). The tone of the Family Circus comics is usually light and amusing, and often touches on religious themes and wholesome family values.

The comment in **Example 15** capitalizes on the innocent and lighthearted tone of Family Circus, drawing a parallel between the comic strip and the ‘Meffy’ pronunciation (3a - [mɛfi]) as “*funny*” and “*charming*”. dirt dirt then humorously subverts his own assertion with a prescriptivist stance, by stating that this comparison (and/or the ‘Meffy’ pronunciation) is also “*wrong*”.

This demonstrates another way in which cultural references can be drawn into discussions to prove points, make jokes, and otherwise assert stances. The indexical associations of the referent (in this example, “*funny*” and “*charming*” characteristics of Family Circus characters) are reappropriated, allowing further comparisons to be made.

Additionally, the comment garnered 30 favorites, most of which were given within days of the comment being posted. Although MeFites favorite comments for a variety of reasons, the comment count itself is visible to all readers by default (although the count can be filtered or hidden) and can provide meta-data to readers concerning that comment’s reception by others (see 6.4.3.2 **Reinforcing Message Chain Links Through Metapragmatic Activities**, p. 174 for further analysis of favoriting).

6.5.2.4 *Comments Referencing a MetaFilter In-joke*

A special type of cultural reference — community specific in-jokes — were found in 3% of comments in the 2006 thread and 14% of comments in the 2010 thread. This disparity was not unexpected, as the community has had several additional years to develop a history of in-jokes and community-specific memes by 2010 as compared to 2006.

Additionally, the concept of the ‘internet meme’ was a relatively new idea in 2006; the Know Your Meme project, with the aim to document all internet memes, was only created in 2007. Internet memes have achieved increasing popularity in online communities in recent years and MetaFilter has been no exception — the MetaFilter Wiki has a page specifically devoted to in-jokes and memes, which can be found here: http://mefiwiki.com/wiki/In_Jokes.

Two popular MetaFilter memes and in-jokes referenced in comments about the M-Set are showcased in **Example 16** and **Example 17**.

Example 16. MetaFilter: Taglines

MetaFilter: Meh. Fie.

posted by DevilsAdvocate at 5:36 PM on March 24, 2010 [3 favorites +] [!]

‘MetaFilter taglines’ originate as far back in MetaFilter history as the year 2000 (see 3.4.1 **Elements of the MetaFilter Register**, p. 71). This in-joke is made when a participant quotes or paraphrases an earlier comment in a post and prepends “MetaFilter:” to the comment, as if to suggest that the earlier quote epitomized the community in some way (akin to the taglines under the logos in the top left corner of the subsites).

In **Example 16**, the tagline format is used to assert that “*Meh. Fie.*” is an apt characterization of MetaFilter. Through personal correspondence with DevilsAdvocate, it was revealed that the comment was made as a clever riff on an in-joke, but did not actually reflect his preferred pronunciation. This comment was in no way deceptive, but the incongruity between the stance the comment contains and the linguistic preferences of the participant who made the comments highlights some of the complexities of online communication, especially where topics of language use and pronunciation are concerned.

The comment garnered only three favorites, which is typical for a MetaFilter tagline comment, as the meme is very prevalent and MeFites have occasionally expressed that they tire of seeing examples of it propagated throughout the site. Despite this, some MetaFilter taglines do receive more than a few favorites, such as when the application of the meme is especially apt or salient in some way, and therefore stood apart from other instances like it.

Example 17. Special Snowflakes

I am a special snowflake, apparently, because in conversation I would never call someone a MeFite. It would be a Metafite.

posted by anastasiav at 4:42 AM on March 25, 2010 [+] [!]

Although the idea of one being a “unique snowflake” had its origins in the 1996 novel (and later 1999 film) “Fight Club” the use of the term ‘snowflake’, and especially ‘special snowflake’ has become an in-joke in its own right on MetaFilter since the beginning of the site. MeFites often use the phrase to refer to themselves, usually somewhat ironically, as a way of acknowledging special circumstances that may only apply to them. The use of ‘snowflake’ and related terms however has become overused in the eyes of many participants and there have been several MetaTalk posts over the years requesting that the phrase ‘special snowflake’ and all of its variants no longer be referenced on the site. However, others still find the in-joke apt and endearing and will continue to use it, or at the very least not be bothered by it, despite the protests of others.

In **Example 17**, the statement “*I am a special snowflake*” prefaces the assertion that the term ‘MeFite’ is dispreferred by anastasiav and the extremely rarely seen ‘Metafite’ is preferred instead. Therefore, despite negative views some MeFites have about ‘special snowflake’, referencing this bit of in-group knowledge to assert a stance demonstrates awareness of site culture and confers some in-group authority on what follows. This authority is further established through recognition that anastasiav’s preferred pronunciation of ‘MeFite’ goes against the norm, making the use of the in-joke even more appropriate in these circumstances.

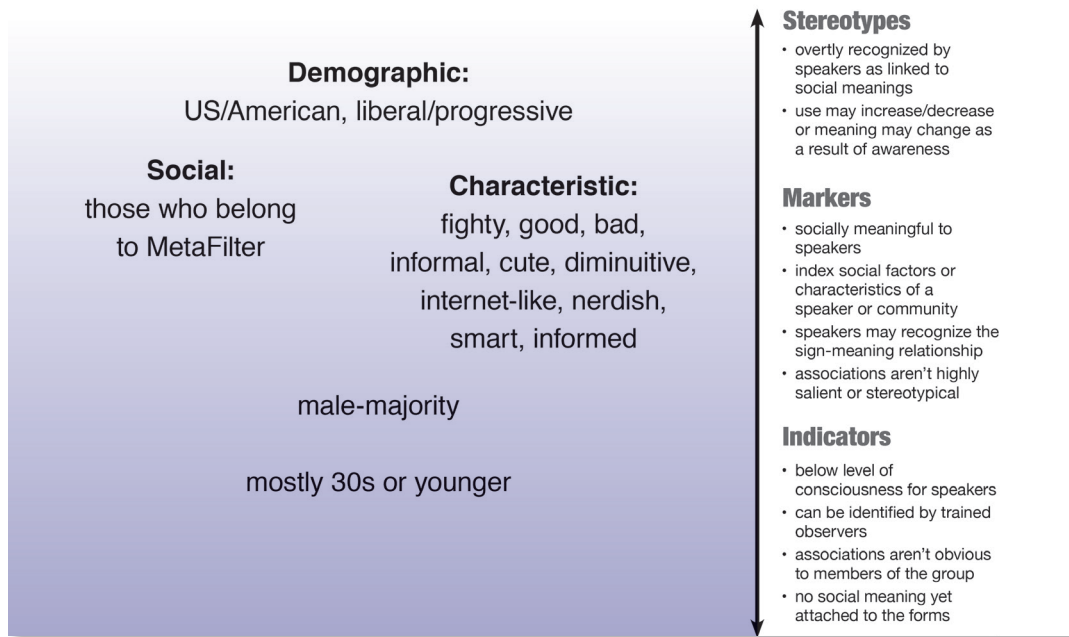
6.5.3 Indexicalities of ‘MeFi’ and Popular Pronunciation Variants

Four categories of qualitative stance data have been presented and discussed. The examples chosen here have highlighted the myriad ways stance can be expressed, as well as the complexity of the values contained within. These examples demonstrate the range of indexical associations the M-Set can have, leading to their enregisterment as such.

In the next several figures, the stance data has been combined with previous data results to demonstrate the field of common indexical associations that currently exist for the abbreviation ‘MeFi’, the three most popular pronunciation variants, and a fourth variant, 4b – [mɔɪfɔɪ]; the latter was included because of its saliency as an outlier, and the strong, consistent associations participants had about the variant. The indexical associations shown here are placed according to my assessment of their classification as an indicator, marker, or stereotype (Labov, 1972; see 2.5.3 **Indexicality**, p. 28 for more on this model). Labov’s (1972) model is not typically applied in this way, with the associations being the subject of ‘indicator’, ‘marker’, or ‘stereotype’ labels, rather than the variant itself having such a designation. However, in this research, I’ve found that applying the label to the associations provides a more nuanced description of how the variant might be categorized, as it is only the associations in the higher categories (i.e., non-indicators) that are the visible indexicalities; this format allows the awarenesses of the indexical links to be presented in the field.

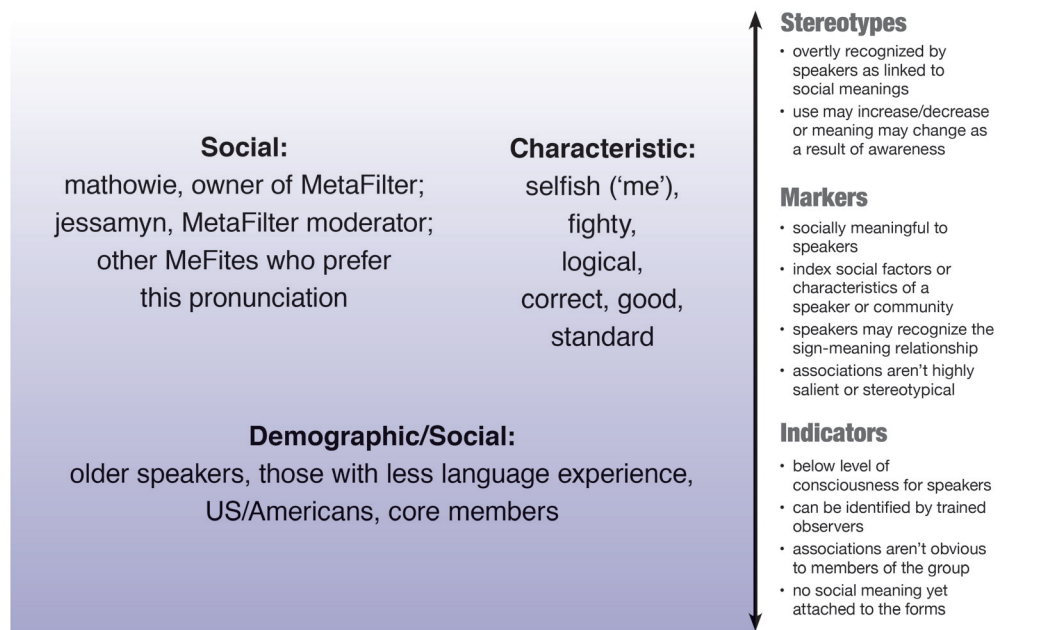
The placements (and the indexical associations that are presented in each field) are estimations and do not represent the perceptions of all individuals; they are based on the findings in the data and my own experience with the community, including over eight years of participation. Additionally, the maps represent indexicalities for variants for a particular point in time; it is possible and likely that attitudes and perceptions have shifted in the years since the most recent data collection (2012).

Figure 64. Indexical Field For 'MeFi'



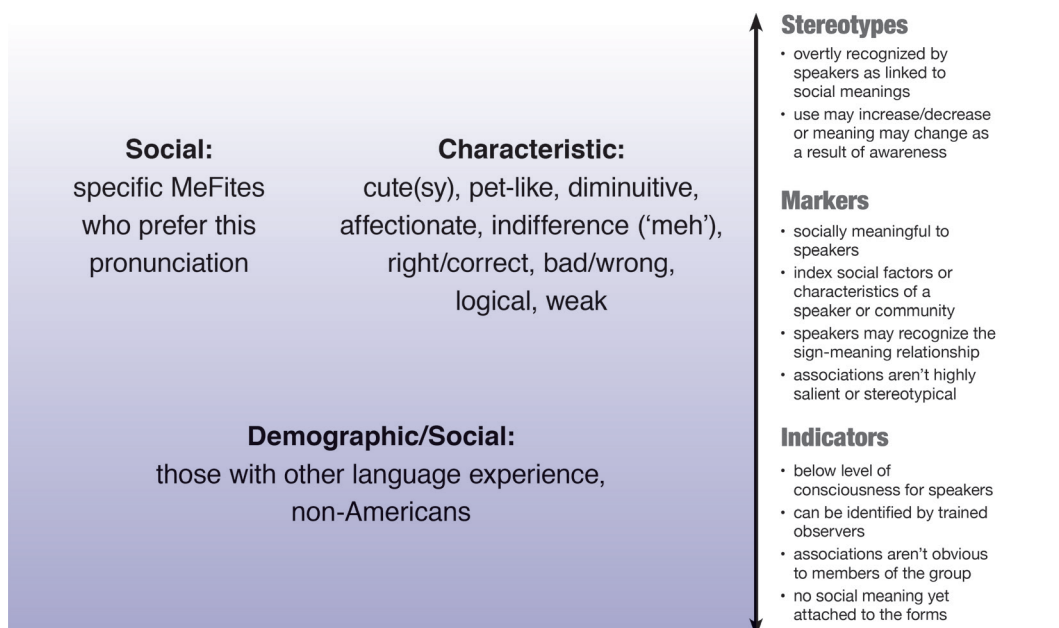
Based on observation of years of community discussion, there seems to be a commonly held belief for many participants (and outside observers) that MetaFilter has strong US/American and liberal biases. These are indexed in the use of 'MeFi', as the nickname itself denotationally references the community and its people (e.g., 'MeFi' indexes 'MeFites'; those who belong to MetaFilter). Less known demographic associations that some, but not all participants may have with 'MeFi' include ideas about MetaFilter's gender and age distribution. Lastly, character-based indexicalities for 'MeFi' in general include more popular stances participants have about the variants, as well as ideas about MetaFilter itself (e.g., nerdish, informed, representative of internet culture); these characteristics may be stereotypes for some, but only markers for others.

Figure 65. Indexical Field For 1b - [mifaɪ]



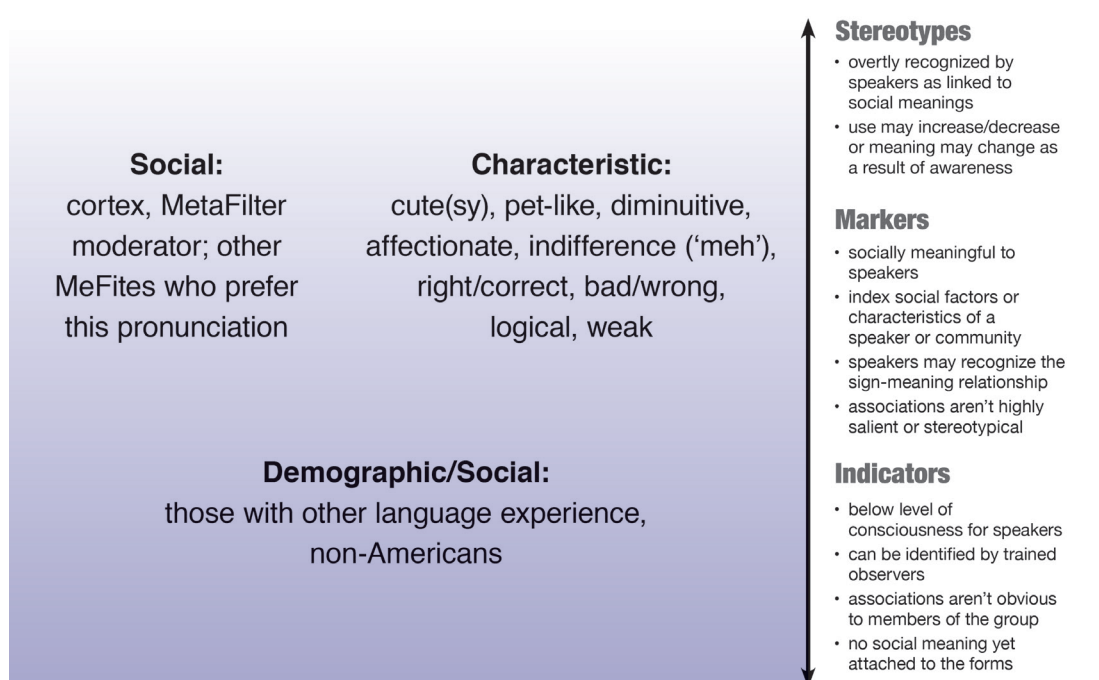
Associations participants may have about this variant at the stereotype or marker level include connections with the site owner and moderator(s) who prefer 1b - [mifaɪ], characteristics such as being selfish, fighty, correct, etc., and some ideas about the popularity or emerging “standardness” of the variant. Factors that are not easily accessible to participants but are nonetheless indexically linked to this pronunciation at the indicator level are age (older speaker), language experience (those with less), geography (US), and level of participation (core).

Figure 66. Indexical Field For 3a - [mɛfi]



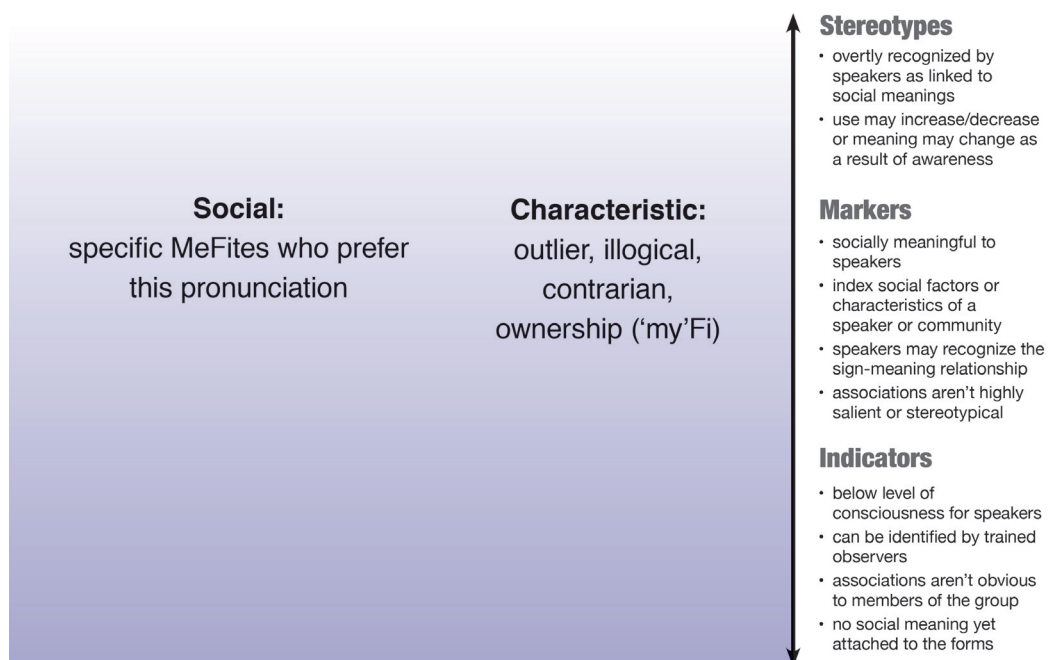
The most common indexical associations participants make about this variant are owing to sound-symbolism associations, e.g., that the variant sounds cute(sy), like a pet name, or weak and indifferent ('meh'), etc. Participants may also associate this variant with participants they know (personally or otherwise) who prefer this pronunciation. At the indicator level, connections were observed in the data results linking this variant to demographic categories having to do with language experience (having other, non-English language experience) and geography (non-American).

Figure 67. Indexical Field For 3b - [mɛfaɪ]



Indexical associations for this variant were almost identical to that of 3a - [mɛfi], as both share a common first vowel. However, this variant has additional connections to MetaFilter moderator cortex, who is forthcoming about his chosen pronunciation and is also very present on the site in several ways, including countless appreciated music contributions. As such, cortex has a lot of influence and capital and therefore many of his personal characteristics may be indirectly linked to his choice of variant.

Figure 68. Indexical Field For 4b - [maɪfaɪ]



While this variant was one of the least popular choices for MeFites, the indexical associations that participants have about the variant were aligned and strong, across all pronunciation groups. Participants consistently felt that the 4b - [maɪfaɪ] pronunciation was an outlier and seemed illogical. Those who preferred it felt a sense of ownership, i.e., 'my'Fi). There were not enough data to make accurate assessments about demographic/social categories indexically linked to this variant at the indicator level.

6.6 Summary of Enregisterment

Assessing the enregisterment of any feature or variety requires an understanding of the social context and important local factors in which the enregisterment occurs. Social factors such as age, gender or class, influences of which on sociolinguistic variability are well understood through decades of well-grounded sociolinguistic research, do not influence the enregisterment process in universally predictable ways and may be superseded by local factors which could be more visible and salient to participants (Bucholtz and Hall, 2010:20; Squires, 2010). For example, in this case study of MetaFilter, demographic factors such as age or geography were highly correlated with M-Set variants, yet the M-Set was being ascribed indexical associations (i.e., being enregistered) along dimensions related to other, more visible factors to participants (e.g., well known MeFites who preferred particular variants, sound-symbolic associations participants could make on their own, comments on the site, etc.). These associations were derived mostly from stances that MeFites took concerning the M-Set and instances of use of the M-Set, either read about on the site or heard at meetups

and on the podcast. This differs from most previous studies of enregisterment, where indexical associations speakers had about the features of the variety were linked to stances other speakers made, instances of use *and* the demographic categories correlating with those speakers (the latter of which could be easily observed or otherwise verified).

With MetaFilter, the primary mechanism for the diffusion of cultural values associated with forms (and contributing to the overall identity of the group) is the involvement with the practice, especially focused on the features of the practice that facilitate the discussion of pronunciation and fostering of meta-linguistic awareness. This facilitating and fostering does not need to occur only online, as shown by the influence of meetup attendance on M-Set pronunciation. Enregisterment does not necessarily require an offline component either, as can be seen with the influence of online discussion of the M-Set, where eye-dialect spellings were employed and references to in-jokes, culture, previous commentary and prescriptivism or grammar were used to take stances and assert opinions. This fostering does not even need to be interactive, as evidenced by podcast listening, which is a passive form of engagement and that was found to be significantly correlated with M-Set pronunciation. From this, it can be seen that MetaFilter provides several overlapping ways that participants can engage with the community, and, as a result, multiple paths in which enregisterment can occur. It can therefore be argued that a seemingly simple community blog like MetaFilter, where the landscape mostly consists of wall-like blocks of white text on colored background, actually provides a rich array of enregisterment opportunities. This allows the register to thrive, and promotes community-building and bonding social capital through several means, all of which have been made as persistent and accessible to participants as possible.

For those who were heavily involved in the community, ideas about pronunciation and the M-Set were reinforced through these multiple channels. This is reflected in the data results showing how increased participation (medium and frequency) not only correlated with M-Set pronunciation, but also MeFites' stronger preference for saying that they would use their preferred variants exclusively. Qualitative data showed that these participants in turn voiced stronger opinions about the M-Set and with more frequency, thereby actively participating in and shaping message chains through various means in the message chain network.

At the micro level, these various mediums, discussions, and debates become instances where individuals can use linguistic resources (such as variability in the M-Set) to take stances and align (or distance) themselves with others. These assertions allow MeFites to express who they are and what they care about, even if it is only for the brief moment it takes to publish a short comment online. In the aggregate, these comments and stances add up to more concrete ideas about who participants are, and how they comprise the community that

Enregisterment

is MetaFilter. The sum of all these parts contribute to the general identity of MetaFilter as a place where such nuanced debates can occur, by participants who are brought together not by shared demographic features, but rather a common interest in discussing any and every topic imaginable.

Chapter 7: Discussion & Conclusion

7.1 Introduction to Discussion & Conclusion

This chapter begins with an overview of the research presented in this thesis, followed by a more detailed summary of the research findings. Explanations and interpretations are provided in cases where the findings did not confirm the original hypotheses. Next, a more general interpretation of the results is presented, addressing the main objectives of the research. A discussion of some of the methodological hurdles encountered while conducting this research is given, including an assessment of the validity of the results. The potential impact of this study on various sociolinguistic topics and general areas of research will be covered, followed by possibilities for future research and concluding remarks.

7.2 Overview of Research

This research set out to give a multi-faceted account of the ongoing enregisterment of a sociophonetically variable nickname used within an online community of practice. This case study was the first in-depth analysis of its kind, and involved a community which was ideal for this purpose — a naturally occurring instance of enregisterment in which there has been a written record of the entire process over time, in the modality in which the process primarily occurred (i.e., a text-based medium). Additionally, the outcomes are reflective of how English-speaking CMC participants might pronounce new names or other neologisms, where often no widely recognized or standardized pronunciations exist.

The methodology employed in this research aimed to find the statistically significant social factors that correlated with pronunciation variants of the name ‘MeFi’ and the term used for the people who belong to the MetaFilter community, ‘MeFites’; the variables have been collectively referred to as the ‘M-Set’. Pertinent to this enquiry was an exploration of the mediums and channels through which these new forms acquired their indexical associations with the community over time (i.e., became enregistered). Lastly, this research was concerned with how the forms in this process, are linguistic resources used in establishing group and individual identities.

Extensive community participation allowed me to understand and describe the context in which sociophonetic variation can occur on MetaFilter, and the ways that conventions, memes, and other forms of messaging operate within that medium. Both quantitative and qualitative methods allowed data to be collected from a variety of sources — these data could be reformulated for use in several types of analysis. Results from these procedures revealed the sociophonetic variation that exists within the community, and the range of social factors which correlated with that variation.

7.3 Summary of Data Results

The summary of results is divided into four categories: data populations, change over time, the pronunciation of the M-Set, and the process of enregisterment.⁵⁷ These divisions are necessary in establishing a context for interpreting all of the findings, with each category building upon the results of the previous one. Therefore, the process of enregisterment can be understood within the framework of populations which vary in their social behaviors and correlate with differences in M-Set distributions over time.

7.3.1 Summary of Findings Related to Data Populations

The two survey populations—the Survey Data and the Panel Data⁵⁸ — differed in meaningful ways. Panel Data participants were found to represent a more balanced gender distribution, as well as being slightly older, on average, than the overall 2010 Survey population. The most notable difference between the two populations, however, was that the Panel participants, on average, were more socially engaged in the MetaFilter community. These participants attended more meetups, listened to the podcast more frequently, and read more areas of the site and did so with greater frequency. Probably owing to this increased engagement, Panel Data participants also felt more strongly about exclusive use of their preferred variant(s) and had given more thought to the matter of M-Set pronunciation than had the overall 2012 Survey population.

Although the Survey and Panel Data differed with respect to social engagement and metalinguistic awareness factors, other factors such as geography, native language, and language experience, did not significantly differ between the populations. These were all demographic factors, which were the only measures that were nearly consistently equal between the populations.

From this analysis, it can be seen that there were two demographically comparable but behaviorally distinct data populations in this research. These data sets could be reliably compared over time, across several measures relating to pronunciation choice.

7.3.2 Summary of Findings Related to Change Over Time

Conducting two surveys over a two-and-a-half year span allowed measures which might have changed over time to be assessed and their outcomes considered in light of M-Set pronunciation distributions. The M-Set showed significant change over time, with participants preferring more 1b – [mifaɪ(t)] variants in 2012 than in 2010. A general trend in switching variants was observed, whereby the most infrequently preferred variants became

57. All outcomes discussed in this section have corresponding results tables in **Appendix G**.

58. See 4.3 **Sampling Design**, p. 96 for a thorough explanation of the data populations.

increasingly less popular, as the people who originally preferred them tended to switch to the more popular variants. MeFites also reported that they felt more strongly about exclusive use of their preferred variant(s) in 2012 than they did in 2010.

Of all the demographic factors, only native language and gender showed significant change over time. The 2012 Survey population showed a higher percentage of non-native English speakers, as well as a more balanced gender representation. The increased gender balance was at least partially the result of improved survey methodology.

Analysis of social engagement factors revealed that only the Panel Data participants showed significant changes over time, through their increased levels of podcast listening and meetup attendance. Owing to changes in the 2012 survey methodology, site visitation over time could not be accurately measured. Regardless, it can be said that the Panel Data participants crucially differed from the overall Survey Data participants in that they were more engaged in the community, and that this disparity between the two populations increased over time.

7.3.3 Summary of Findings Related to the Pronunciation of the M-Set

The overall distribution of the M-Set revealed the highest preference for 1b - [mifɑɪ(t)] variants, followed by 3a - [mɛfi] or 3b - [mɛfɑɪ] for 'MeFi', and 3b - [mɛfɑɪt] for 'MeFite'. The 1b, 3a, and 3b variants accounted for 86% of all Survey Data.

How strongly participants felt they would use their preferred variant exclusively (as opposed to using other variants) was highly correlated with pronunciation choice, with increased levels of exclusive use for the most popular 'MeFi' variant, 1b - [mifɑɪ]. MeFites also felt more strongly about exclusive use of their preferred variant of 'MeFi' than they did for 'MeFite', suggesting more metalinguistic awareness and interest in the pronunciation of 'MeFi' than 'MeFite'.

Many MeFites had given the matter of M-Set pronunciation much thought and demonstrated this as various forms of metalinguistic awareness in survey rationales. The most common justifications cited were grammatical rules or linguistic processes. The act of citing linguistic or grammatical rules was not biased toward any pronunciation group, but the types of rules that were cited were skewed in this way. The 1b - [mifɑɪ] pronunciation group frequently mentioned knowledge of phonetics or prosody, especially as it applied to the first vowel of 'MeFi', while the 3a - [mɛfi] and 3b - [mɛfɑɪ] groups often cited abbreviation processes consistent with a shortening of 'MetaFilter'. Several variations of citing the visual appearance of 'MeFi' were used to justify pronunciation choices as well, and these were heavily skewed toward the 1b pronunciation group.

Participants' native English speaking status did not show a correlation with 'MeFi'

pronunciation. However, having non-English language experience was positively correlated with more group diversity of pronunciation preference; lack of language experience was positively correlated with preference for 1b – [mɪfɑɪ].

Geography was highly correlated with pronunciation choice across both data sets and survey years, with the US and Australia showing a strong bias toward 1b – [mɪfɑɪ(t)] variants, while the UK and Canada preferred these variants significantly less. Age of respondent was positively correlated with 1b – [mɪfɑɪ(t)] variants.⁵⁹ Gender was not correlated with pronunciation in any data set or survey year.

Correlations between individual social engagement factors and pronunciation choice were mixed, with no consistent measure of social engagement correlating with pronunciation across both survey years and both data sets. However, the overall trend for increased engagement to positively correlate with a preference for 1b – [mɪfɑɪ(t)] variants was borne out in every measure for which a significant outcome was observed. When participants who demonstrated little to no social engagement were compared to those who engaged in multiple ways with MetaFilter and did so with greater frequency, a highly significant positive correlation between social engagement and preference for 1b – [mɪfɑɪ(t)] variants was observed.

7.3.4 Summary of Findings Related to the Process of Enregistration

Enregistration was shown to take place via message chains located only on the site at first, but then becoming multi-modal as other opportunities for interaction became available to a wider audience (e.g., the existence of the podcast, the increased popularity of meetups). Various types of stances contained in these message chains were explored, as well as meta-pragmatic activity such as favoriting, which served to reinforce these positions.

While message chains on the site could not easily be quantified, the frequency of ‘MeFi’ use itself was tabulated and a pattern that is typical of the diffusion of a new innovation was observed. Key points in the spread of ‘MeFi’ over time were highlighted, such as the first instances by an early adopter of the innovation, and the first two MetaTalk posts about pronunciation. Examples of authoritative influence on the pronunciation of the M-Set were presented, e.g., a transcribed excerpt of moderators discussing ‘MeFi’ pronunciation on the podcast, and popular comments about the pronunciation of the M-Set.

This exploration gave necessary context to the quantitative results from the Survey Data. For example, while social engagement factors were correlated with pronunciation choice, the qualitative data were necessary to understand how these relationships actually operated. It was not just any social engagement that mattered and influenced opinion; it was the type of interaction that involved sustained interest in aspects of the site where policy, norms, and

59. However, there were no age correlations in the 2010 Panel Data.

community matters were discussed (e.g., MetaTalk, or the early years of the podcast). This gave participants access to information about the M-Set directly, as well as social knowledge about the types of stances MeFites took on topics such as language use, group ethos, and other issues that might come into play in matters of individual and collective identity.

7.4 General Interpretation of Findings

A pronunciation standard for the M-Set has been emerging within the MetaFilter community, which is observable as a significant increase in preference for the 1b – [mifaɪ(t)] variants over time. This increase coincided with higher levels of exclusive use of those variants, as well as with other types of expressions of metalinguistic awareness about language and the M-Set.

The pronunciation of the M-Set was found to be initially influenced by the demographic factors of participants (e.g., their geographic location, age). However, participants who became more involved in the community through increased social engagement were more strongly influenced by that social activity, where discussions about the M-Set served to communicate the stances and associations participants make about the variants. It is from this interaction that the convention emerges, including participants' awareness of it as such.

These data results were the outcomes of many different forms of message chains that participants had contributed to over time. These chains allowed for all manner of ideas about the M-Set to be disseminated and they resulted in different distributions of pronunciation preferences along dimensions related to the influence and reach of those message chains. For example, MeFites who were the most actively involved in the community (e.g., through regular site participation, podcast listening, and meetup attendance) were receivers of and/or were participants in multi-modal message chains more than those who were less engaged. The more engaged participants had more opportunities to hear instances of the M-Set in use, as well as increased exposure to evaluations of their use (and those who used particular variants).

As well as participants' increasing knowledge about the M-Set, social involvement heightened these participants' metalinguistic awareness in general. Social knowledge and awareness is a form of register competence, which can translate to social capital. As a result, participants who demonstrate this knowledge may be recognized as an influence or even an authority on matters of importance to the community, which further establishes these participants' place within the community.

However, it is the accretion of several types of stances and interactions, over time, that give form and vitality to a particular entity — whether that entity is the identity of an individual, the collective ethos of a group, the group's register, or a particular form or feature in the register. While this research showcases just one example of the process by which an

entity establishes its place within a community, the concepts which underlie the formation of indexical relationships between the M-Set and the MetaFilter community (i.e., the enregisterment process) can also be applied to these other entities.

For the M-Set, the pronunciation of the variables is partially influenced by the demographic characteristics of those who use them. However, the most influential factors on M-Set pronunciation, and the most commonly-recognized evaluations of the variants (i.e., their enregistered values), are the amounts and types of social engagement of the participants. This engagement is shaped by the modality and the medium — text-based interaction in CMC. Therefore, when considering how enregisterment occurs in online spaces, it is necessary to consider both demographic factors and socio-structural ones. These both contribute to enregisterment in different ways, and at different points in the process.

In CMC-based enregisterment processes, demographic factors come first and foremost; they influence enregisterment at the individual level, and bias participants' pronunciation outcomes prior to (indexical and stance) information from others that they might receive through social engagement. This is one way in which participants bring their backgrounds with them into CMC spheres, even if those characteristics are not visible or otherwise observable to others.

Social engagement reinforces, augments, or changes the indexical associations individuals have about entities. In the pronunciation of the M-Set, social engagement contributes to the enregisterment of the variables in different ways, with each type of interaction having distinct advantages and disadvantages in that process. For example, participants on MetaFilter subsites were receivers in message chains more often than they could be observers of correlational indices (such as age or geography linked to pronunciation choice), regardless of how frequently those participants engaged with the community. Podcast listeners were participants in multi-modal message chains, receiving additional indexical information from the audio streams of those in positions of influence and authority (i.e., MetaFilter moderators). Meetups were potentially the most influential in terms of enregisterment, but they had the most limited reach. At meetups, participants were very likely to hear instances of the M-Set, and presumably were able to easily link those uses to the characteristics of the speakers who used them. M-Set pronunciations are often a topic of discussion at meetups, and also therefore provided opportunities for those who attended meetups to increase their metalinguistic awareness about the M-Set and language variation in general. However, the overwhelming majority of participants did not attend meetups, and the meetups themselves were not accessible to most participants due to their geographic locality or other constraints. Even so, meetups generally consisted of a local distribution of a relatively

small number of participants, and so MeFites may have recognized that the interaction that occurred may not have been generalizable to the rest of the community.

Regardless, participants at any level or type of social involvement were more able to acquire social information from stances about the M-Set than they were able to make connections between M-Set variants and the social categories linked to pronunciation outcomes (because they could not easily see or learn about the demographic features of participants). The message chains that participants were exposed to or took part in were shown to influence their choices, and in many instances, the chains persuaded them to switch toward the emerging standard.

More generally, knowing the pragmatic norms of the community and elements of its register allows participants to manipulate and play with the language to achieve social effects. This is evident on MetaFilter when participants reformulate common tropes, quote others to perform parodies, and use in-jokes. This linguistic play is so productive and favorably received on MetaFilter that the very idea of doing it is entrenched in the identity and ethos of the practice of the community itself.

As such, the variability in the pronunciation of 'MeFi' is, again, a linguistic resource that MeFites may use to express their individual identities. This is similar to other linguistic forms and features that allow for social positioning within a group (Barton and Lee, 2013, p. 87). That is, 'MeFi' can be used to define in- or out-group members or to define stances or identities within a group.

The accumulation of stances — especially frequent or salient stances — also help to design the co-created identity of MetaFilter, via the self-created identities of its participants. The M-Set is just one of many linguistic resources used in achieving these objectives, but it is one that is emblematic of group and individual identity at various levels of meaning (e.g., denotationally, connotationally). As the stances about the M-Set change and evolve over time, so does the discursive record of the site, each informing the other's history.

7.5 Research Hypotheses and Inconsistent Outcomes

Several hypotheses about which factors were correlated with pronunciation choice were confirmed by the results of the survey, but some of these results were mixed or inconsistent. While geography was consistently correlated with pronunciation, the factors found to have inconsistent correlations either between data populations or across survey years included the following: having other language experience, age, AskMetaFilter and MetaTalk subsite visitation, podcast listening, and meetup attendance. These outcomes will be individually reviewed in this section.

Having non-English language experience was more strongly correlated with pronunciation outcomes in 2010 than in 2012, and more so for the Survey Data than the Panel Data (having other language experience was not significantly correlated with pronunciation in the 2012 Panel Data). Assuming that the Panel Data participants did not significantly increase their language knowledge in the two and a half years between surveys, and given that the methodology for this measure was considerably improved in the 2012 survey, the results for 2012 have much greater statistical validity than those for 2010. As such, the 2012 Survey Data showed a correlation between language experience and pronunciation, whereas the 2012 Panel Data did not. This difference may be explained by the increased social engagement of Panel participants as compared to the overall Survey population. The Survey population, being less socially engaged overall, were more likely to make pronunciation decisions along lines relating to personal experience and history that they brought with them into their social environment (i.e., demographic factors, such as language background). Conversely, the Panel participants were more apt to consider additional influences that were acquired through their interactions with others, such as others' stances, heard instances of the M-Set, etc., which may not have been available to the less socially-engaged participants (e.g., they were probably not exposed to this information). This outcome highlights the differences in data populations, as well as the effect of social interaction on pronunciation variation.

The mixed results for age across data populations, however, cannot be easily explained by the same logic. Unlike other demographic factors such as gender, age was treated rather simplistically in this survey methodology, with participants reporting the number of years they had acquired, rather than socially- or self-defined age groupings, which consider life stages or other criteria in determining age status. Age was treated as ratio scale data in collection, then sorted into roughly equal groups in data normalization. This rather broad treatment of the age data meant that all participants who took both surveys (i.e., Panel Data participants) had aged by approximately two years and five months between surveys, and were not necessarily in the same age group in the 2012 data that they were in 2010 data. Additionally, some Panel Data participants gave their age in one survey but not in the other. These two factors may have contributed to inconsistencies over time in the Panel Data. However, the Survey Data participants showed a consistent trend of age positively correlating with 1b – [mifar] pronunciation across both survey years.

Further analysis was conducted which revealed that removing the 25-29 age group from statistical analyses resulted in no significant bias in 2012 Panel Data results, and these data were therefore no longer showing correlations between age and pronunciation choice across both survey years.

It is likely that the reasons for the significant outcomes in the overall Survey Data but not the Panel Data are due to the influence of message chains, which are more readily available to more engaged participants, such as those that yielded the Panel Data. That is, age was a greater influence for pronunciation to the Survey Data participants than it was for the Panel Data participants, who had additional social engagement factors to consider, which in all likelihood influenced their linguistic behavior more than demographic factors did.

These varied results across age groups and between data sets provide an opportunity for further research. Employing a more complex survey methodology in collecting information about participants' age, including qualitative data, would go a long way toward giving a more thorough account of how social factors such as age might influence enregisterment, and how age more generally biases linguistic behavior in a text-based medium (where participants cannot easily assess the ages or life stages of their peers, but where participants are nonetheless influenced by such factors).

Findings pertaining to social engagement and linguistic choices were not consistent across population data sets nor survey years. For example, with regards to subsite visitation, only the 2012 Survey Data showed a correlation between pronunciation choice and AskMetaFilter visitation frequency. Both the 2010 and 2012 Survey Data showed a correlation between pronunciation choice and MetaTalk visitation frequency. The Panel Data showed no correlations between pronunciation choice and visitation frequency on any subsite or survey year.

To interpret these inconsistent findings, two previously reported results need to be considered. First, the Panel Data participants, on the whole, were shown to be significantly more socially engaged than the Survey Data participants. From a data perspective, this means that an overwhelming majority of the Panel Data participants' data points were skewed toward more frequent or regular social interaction categories. Comparatively, the Survey Data showed a greater and more balanced range of interaction frequencies and types across the data table categories.

The second thing to consider is that the measures being described here involve whether or not there was a pronunciation bias for those with increased engagement compared to those with decreased engagement *within* the population being sampled. Therefore, where the population being sampled was less diverse in their engagement levels and/or where the numbers of participants for categories within the measure were small, the likelihood of observing statistically meaningful differences would be reduced.

This likelihood appeared to be the case with measures of social engagement and these data populations. The Survey Data participants represented a more balanced diversity of all

available combinations of measure categories and pronunciation choices, whereas the Panel Data were both smaller in terms of sample count and skewed toward increased engagement categories within the measure. Therefore, in the Panel Data, the difference between the amount of engagement for those who engaged with MetaFilter frequently and those who did not was not as great as it was in the Survey Data, and did not result in any significant differences in the Panel Data, except in cases where the bias in pronunciation choice was extreme (e.g., podcast listening prior to the 2010 survey).

This rationale also helps explain similar outcomes for meetup attendance frequency and pronunciation choice — significant effects were achieved in the Survey Data, but not in the Panel Data. Again, the Panel Data were based upon a smaller sample size and were significantly skewed toward the more frequent meetup attendance categories as compared to the Survey Data, which were more numerous and represented a greater range of involvement. This allowed the relationship between meetup involvement and linguistic behavior to be more prominent, achieving significant levels in the Survey Data.

A correlation between podcast listening and pronunciation choice was found in both the Survey and Panel Data in 2010, but in neither data set from 2012. Unlike the other social engagement factors, podcast listening frequency had an influence across all of the data, but that reach was not consistent over time. This may be explained by the fact that there were fewer podcasts in 2011 and 2012⁶⁰ than there were in the years leading up to the first survey in 2010. Additionally, participatory research revealed that the discussion surrounding the M-Set on the podcast had qualitatively changed in recent years, occurring with less frequency and depth, as the topic had been covered numerous times previously and the moderators were familiar with each others' preferences and stances about the M-Set. This may have resulted in the topic becoming less prominent for podcast listeners who knew of the debate over pronunciation, and potentially seeming unimportant or inaccessible to those who were hitherto unaware of the debate. That is, it was most likely a change in the quantity and quality of the message chains available in the medium (i.e., the podcast), and the decrease in availability of the medium itself (i.e., fewer podcasts) which led to a lack of correlation between podcast listening and pronunciation outcomes, and not a change over time in the data populations or the methodology that primarily accounts for this outcome.

To summarize the hypothesis and outcomes pertaining to social engagement, the hypotheses regarding the existence of positive correlations between measures of social engagement and the most preferred pronunciation variant in the Survey Data have been confirmed (see 6.4.8 **The Effect of Social Engagement on 'MeFi' Pronunciation**, p. 185 for

60. See Figure 46. Social Engagement Factors and the Enregisterment Timeline, p. 163.

combined results pertaining to these measures). However, the hypothesis was only partially confirmed for the Panel Data; these participants were significantly more involved in the community, but the positive effects of increased social engagement were either inconsistent or did not achieve significance for some measures.

This disparity between the populations may also suggest a saturation point with respect to the sustained influence of social engagement. MeFites who have participated on the site regularly, listened to the podcast, and attended meetups have likely been, at the very least, receivers of message chains about the M-Set, if not active participants in the debates. Many of these participants may have settled on their pronunciations and are less likely to be swayed by others' stances and rationales, as this information is no longer novel and may have already been taken into consideration.

Future research on enregisterment should consider the complexity of data populations with respect to hypotheses about social engagement and their potential correlations with variables. Similar to S-curves representing change over time, in which a point of stasis is reached after a rapid uptake, the effect of social engagement also has such limits. Participants in a community may take great interest in novel topics, and be heavily influenced by the message chains that are exchanged in that initial engagement. However, over time the frequency with which those same participants will re-engage may stabilize or significantly diminish. Therefore, hypotheses about the effects of social engagement should not simply assume a consistent linear relationship between interaction and outcomes. Rather, hypotheses should factor in the possibility of saturation and its potential effects on outcomes for those within the community who may have reached that threshold; this was a likely explanation for the effect of social engagement on the Panel Data's pronunciation results in this research.

7.6 Research Limitations

This section aims to outline some of the limitations of this work, from study design to implementation and its potential application to future work. It is my hope that the limitations outlined in this section will be a useful guide to researchers pursuing knowledge and planning research in the areas of sociolinguistics and CMC studies.

The work I have undertaken here has necessitated the invention of a methodology which can account for this new domain in which enregisterment can occur, as well as how to account for and interpret the results from this unique data collection process. This methodological journey has led to many new insights, but it has also consumed a fair amount of effort and time in the process when unfruitful avenues were pursued. Conversely, some results of the research, while fortunately turning out favorably, could have benefitted from additional foresight and planning. I would like to address all these aspects openly and

candidly, so that the information shared can be of use to others who might adopt some of the methodology into their own research, in addition to providing some necessary balance to the results and analysis I have presented thus far.

7.6.1 Internal Validity

The internal validity of the results in this study concerns the reliability of the measures employed to collect data. In some cases, the measures included in this research were too narrow in scope or otherwise did not fully capture the information they were intended to. In other cases, the measures were too broad or ambiguous, resulting in copious amounts of data which required extensive normalization and interpretation, which potentially introduced bias.

However, the majority of the measures that were implemented were done so in ways which resulted in high internal validity, and in most cases the responses were verifiable by some methods relating to the study or survey design. For example, the majority of participants corroborated their pronunciation choices in survey rationales by providing additional information, such as words which rhymed with their choices. These survey rationales also often corroborated the responses concerning the amount of thought given to pronunciation and how strongly participants felt they would use their preferred variants exclusively.

Additionally, the Panel Data population provided a measure of control in some respects, as these were data from the same participants over time. This allowed the amount of error to be assessed for factors which were fixed or proportional over time (e.g., year of joining MetaFilter, age) as well as the amount of variation over time for factors which were more mutable (e.g., country of residence, language experience). Also, for factors such as gender, which was elicited using a list of options in the 2010 survey but with a free-form answer box in 2012, the effects of changing the measure's implementation could be assessed and reflected upon. As a result, it was found that measures such as gender greatly benefitted from this change in implementation for several reasons. Participants valued the opportunity to self-define, which resulted in their answers being a more accurate reflection of the measure overall. This change also allows the researcher to categorize responses as appropriate for the participant community, instead of relying on imposed structures (such as a binary gender paradigm, in the case of gender results).

Issues of internal validity involving measures which were possibly too narrow in scope included the following survey questions:

1. Country of residence — As geographic background was one of the greatest and most consistent influences on pronunciation choice, it would have been useful to have been able to make a distinction between participants' country of residence versus their country of origin.

2. Gender, Dialect, and Ethnicity — These questions were not free-form choices in the 2010 survey, resulting in some participants necessarily assigning themselves to ill-fitting categories, or choosing not to respond to the questions.

Issues of internal validity involving measures which were too broad or ambiguous in scope included the following survey questions:

3. Exclusive use of preferred variant — While this measure proved to be a significant influence on pronunciation choice across both populations and survey years, the wording of the question resulted in possibly subjective interpretations of what was being asked. For future research, this measure would benefit from being split into two or more clearer questions about usage preferences, including an option to select whether or not the term was ever spoken aloud by the participant.
4. Ethnicity and Nationality — these questions were free-form in 2012 and not well-defined. Answers ranged from self-defined perceptions of ethnicity, long explanations of family heritage, or non-comparable responses based on potentially country-specific meanings of the terms. In future, global census-style options should be provided and/or further clarifications given of what is being asked (e.g., “*What is your current nationality? Your nationality refers to the country (or countries) that would be eligible to issue you a current passport.*”).

Additionally, some measures should have been included in both surveys but were not, for reasons primarily stemming from a desire to keep the surveys as short and minimally time-consuming for participants as possible. This approach likely led to a higher response and completion rate, but also meant that some potentially significantly correlated data were not collected. For example, the results of this research revealed that the ways in which participants were able to access the site may have been an influence on their frequency and style of participation, which could have ultimately been an influence on their linguistic behavior. Therefore, future research will include a question regarding the means by which participants access MetaFilter (e.g., laptop/computer, smartphone, other device), including whether or not they access the site directly or through a proxy (e.g., RSS feed, Twitter feed, other website, other means).

Additional speech data — either recorded at meetups or from conducting sociolinguistic interviews — would be a benefit to future work as well. The approvals to collect speech data were obtained for this study, but it was found that there were many practical and ethical issues with gathering a representative sample of high-quality speech data from social events at various global locations. The possibility of conducting sociolinguistic interviews using Skype or other voice-over-IP applications was considered, but again eschewed owing to recording limitations and the time and effort necessary to transcribe and analyze large amounts of qualitative speech data.

Lastly, one of the main methodological hurdles in this research involved obtaining accurate knowledge regarding participants' pronunciation preferences. Recordings of the participants' pronunciations of the M-Set in various speech styles would have been optimal, but highly unfeasible given the nature of the communicative environment. Ultimately, synthesized recordings of the variants were used, which had the benefit of providing all participants with the same stimuli to judge. Unavoidably, this may have introduced some potential new issues, such as participants' not being able to access, play, or hear the recordings, or differences in participants' perceptions of the recordings.

7.6.2 External Validity

The external validity of this study concerns how generalizable these data are over the rest of the MetaFilter community, and other populations. Fortunately, the study design—which included several years of personal community participation over time—yielded a participation rate over five times as large as what is generally considered to be a representative sample. This means that the sample size for the Survey populations can be generalized over the entire MetaFilter population. However, the Panel Data differed and were not generalizable over the entire population. These participants, while still comprising 5% of the active MetaFilter userbase, were qualitatively different from the majority of the community, owing to their increased social engagement.

While the path to enregisterment for the M-Set variables cannot be generalized to all other netologisms, or all other CMC communities, the case study presented here provides possibilities for how these processes might potentially work. It is likely that CMC communities with similar social structures and aims would be subject to the same influences. That is, other primarily text-based CMC community weblogs may develop their registers through similar processes to those of MetaFilter, with demographic factors initially correlating with linguistic behavior, but community-specific measures of social engagement proving to be a greater influence for those seeking inward trajectories into the community.

7.6.3 Statistical Analysis

The high response rate to both surveys resulted in the sample size was more than sufficient for statistical analysis. However, for some measures, the small number of responses for some categories within the measure meant that the data needed to be restructured (e.g., countries other than the US, Canada, the UK, or Australia were grouped together), or the statistical power was not as great (even if the effect size was highly significant). There were no cases reported here in which the data were completely underpowered (making the results unreliable or invalid), but the Panel Data set was smaller than the overall Survey Data (as a

subset of the Survey Data), and they were skewed in some measures, resulting in less statistical power than that found in the Survey Data results. This is noted for completeness, but is not a cause for great concern in the results or analysis.

Initially, binary logistic regression models were implemented as the statistical analysis method, with the dependent variables across all measures being the [mi-] or [mε-] pronunciation groups for 'MeFi'. However, regression models are a form of predictive analytics, and are not necessarily suitable for assessing the basic correlations between factors in the data.⁶¹ Therefore, after much work, the methodology was changed and basic chi-square tests across all data were implemented and their outcomes analyzed. Future analysis could potentially include a log-linear analysis of the data, as this is basically an extension of the chi-square test, but is designed for multiple factors and is in the form of a model (Field, Miles, and Field, 2012).

Regardless of the above consideration, the chi-square tests were sufficient and fitting for the type of data involved in this research. Additionally, the results were simple and straightforward to interpret alongside the qualitative data. In the end, the methodology relied upon throughout this thesis provided a good balance for a thorough mixed-methods approach.

7.7 Implications of Findings

It is necessary to return to some of the concepts presented in the literature review section and to reflect on their relationship to the outcomes of this study. In some cases, the theories, models, and ideas of others have been supported by this research. In others, the work has been modified or expanded, as the outcomes and the environment in which the research takes place have necessitated changes to earlier models.

7.7.1 Implications for the Sense of 'Community'

"...different people in cyberspace look at their virtual communities through differently shaped keyholes. In traditional communities, people have a strongly shared sense of place — the room or village or city where their interactions occur. In virtual communities, the sense of place requires an individual act of imagination" — Rheingold, 2000, p. 53

The issue of MetaFilter as a potentially intangible, unlocatable place is also part of the debate over the pronunciation of the M-Set. Participants recognize that the textual medium presents linguistic hurdles for self-definition and phonetic standardization of their community name. In this sense, the "keyhole" of which Rheingold (2000, p. 53) speaks lends itself not only to different conceptions of place, but also different conceptions of grammar and language use.

61. Results from the binary logistic regression models are provided in **Appendix H**.

Each participant has their individual mental representation of both meanings of ‘MeFi’ (its referential sense of ‘community’ and the pronunciation of it as the name of the community). While there may be overlap with other participants’ imaginations of each, the degree to which there can be variation in conceptions may be much greater without the feedback that a visible geographic location of the community or spoken exchanges can provide.

This variation in the referential meaning of ‘MeFi’ confers advantages and disadvantages for the community. On the positive side, the range of conceptions about what ‘MeFi’ is means that a wider range of possibilities exist and that participants are free to pursue them. This fosters creativity and freedom of expression, and creates a more egalitarian landscape. On the negative side, participants may feel more isolated by this ambiguity and diversity in the meaning and phonetic representation of ‘MeFi’, as they may not easily have the means to assess whether their conceptions are shared by others, or whether conventions and standardized forms exist. Regardless of the perception of variation in conception of the M-Set as a “good” or “bad” feature, the reality of this sense of place exists, and must be reconciled as one feature of this new experience of community in online space.

More generally, this research has shown the origination of a linguistic innovation, how it has spread through an online community, and the means through which it became enregistered as a meaningful marker of the group. I’ve shown how the structure and social organization of community can foster positive attributes such as engagement and enregisterment, which is not necessarily attributable to the “richness” of the medium.⁶² Evidence for this is seen in the many years prior to the existence of the MetaFilter podcast, where the establishment of ‘MeFi’ on the site — and all of its indexicalities — occurred in a text-based medium, and with great recurrence and frequency.

7.7.2 Implications for the Community of Practice (CoP) Model

This case study of MetaFilter has greatly benefitted from a CoP approach, which focuses on what it is that unifies a community. MetaFilter provided a unique example of a constellation of practices which exist primarily online, and whose practitioners do not rally around an easily identifiable goal or single unifying theme, but rather the idea of discussing any and all topics imaginable.⁶³

Several concepts from the CoP model have been borrowed and applied to this research. Modes of belonging (i.e., engagement, imagination, alignment) and levels of participation (i.e., core, active, occasional, peripheral, transactional) have been particularly helpful

62. See 2.4.1 Classification of CMC Research Areas, p. 21 for background info on Herring’s 2004 study.

63. It should be noted that the existence multi-themed online CoPs is not unique to MetaFilter, as many CMC-based communities exist under a similar framework, e.g., internet forums such as Reddit or microblogging sites such as Tumblr.

concepts in analyzing data and accounting for the varied social and linguistic behavior of MetaFilter participants. These concepts were combined with ideas about enregisterment and capital where possible, which proved to be a multi-faceted way to interpret the results. Further and more thorough application of the CoP model in studies of enregisterment may continue to prove fruitful, and could potentially advance understanding of the enregisterment process, while also providing more case studies of CoPs.

7.7.3 Implications for Enregisterment

In the eleven years since the first study of enregisterment (Agha, 2003), the concept has become very popular in sociolinguistics, with notable advancements to theory by Johnstone (2009, 2010), Beal (2010), and Squires (2010). Similarly, in the last decade or so, researchers in CMC have also made major strides (Baron, 2003; Herring, 2004, 2007; boyd and Ellison, 2008; Crystal, 2008).

The findings of the present study build an important bridge between sociolinguistics and internet research. These outcomes provide enregisterment researchers with insights about new environments to explore and various socio-structural factors to consider, and give CMC researchers new sociolinguistic tools that can be applied to their research endeavors.

Previous studies of enregisterment have been smaller in scope and depth, and efforts were focused on the communities and variables involved in the enregisterment of forms or varieties, rather than in the transference mechanism itself. The scope of this research has enabled an expansion of the message chain concept (formerly, 'speech chain'; Agha, 2003, 2005). This expansion was necessary to make the concept more universally adaptable to account for enregisterment occurring through non-spoken modalities, as allowed me to provide further clarification to the minimal requirements of a message chain.

Using the flexibility of this updated message chain model, it became possible to describe and compare chains across a variety of dimensions. In doing so, several questions were implicitly asked and answered in the data, revealing patterns that would not be visible were the data not delineated in such ways. For example, the distinction between text-based versus multi-modal message chains made it possible to see that the introduction of 'MeFi' and its eventual enregisterment as a meaningful marker of the group occurred primarily in text for several years, where it had a wide reach to participants of many levels. The benefits of this could be compared to those of multi-modal message chains, which occurred with more frequency once the podcast was introduced and as meetups became more popular.

With the definition of a message chain further elucidated, and the requirements circumscribed more fully, it becomes possible to more productively apply the concept in future research on enregisterment. To carry this further, researchers can use the flexibility of

the model and then tailor it descriptively, according to the particulars of their communities under investigation. Specifically, the following questions could be addressed (where they are relevant to the data population and research):

- Are the chains multi-modal? If so, which modalities do they employ?
- Are the chains typically one-to-one, one-to-many, or some other combination?
- How easily can chains be received and/or their messages interpreted?
- How easily can chains be replicated? Through what means?
- What types of associations and stances are transferred through the chains?
 - Are these aligned with (i.e., visible alongside) speaker characteristics?

These are a few of the questions that helped inform this research, and could be of equal benefit to others. Additionally, these research lines could potentially make the works that result comparable along particular dimensions, and allow us to continue to build upon the message chain concept, thereby adding to the body of work on enregisterment.

This case study has also added to our understanding of variables and their role in the process of enregisterment. For example, geography is a universal variable in the sense that we all bring our geolinguistic backgrounds with us into our communicative spheres. However, the amount of influence this factor has seems quite variable, relative to the importance of geography to the communicative environment in which enregisterment is occurring. For online communities such as MetaFilter, geography is not and does not need to be publicly indexed by the variants; other factors may have stronger and more meaningful associations.

This study has supported the notion that social engagement factors are more directly relevant to enregisterment processes than other non-participation-based factors such as geography or age, which may bias choices but usually are not features which actively shape outcomes. That is, in every case of enregisterment, it is the act of participating in the community, taking stances, and allowing message chains to occur which advances the process. The means by which message chains are exchanged (i.e., the specific types of social engagement) may vary, relative to the structure of the community and the medium, but the act of participation is a universal in the process. As such, the concept of having a 'local identity' is relative; it is not always the case that enregistered forms which are indexical of a local identity are also aligned with a geographically identifiable location.

Furthermore, the enregisterment of a variable is not simply a function of increased frequency of use within a community. While it is true that the variable must reach a certain level of prominence (i.e., use, measurable by achieving a threshold of frequency) for it to be viable to undergo enregisterment processes, it is not this use alone that ensures its enregisterment. Many features and word forms do propagate throughout a community, as

evidenced by their increased use by disparate members of the community across several social dimensions. However, it is necessary for these features and forms to acquire indexical associations that are emblematic of characteristics, qualities or behaviors (e.g., stereotypes) of the group. These indexicalities must also spread and achieve a certain level of recognition. It is only then can we see how the enregisterment of a variable takes place, through the spread of stances and attitudes about the features and forms. As such, our study of linguistic behavior must always return to what we as participants think and feel about the world around us, and all the tools at our disposal to help create it as we wish.

The broader implication of this study for our understanding of enregisterment is that it is a highly contextual process. Generalizations about how enregisterment works beyond the basic functioning of message chains or the influence of at least some type of social engagement are unlikely to apply to all communicative environments where enregisterment can occur. It is this variability which allows enregisterment to be dynamic and innovative, and uniquely reflective of every environment in which it occurs.

7.7.4 Implications for Indexicality

The outcomes of this research bring an interesting challenge to current models of indexicality. The M-Set variants are approaching ‘stereotypes’ (according to the Labovian model, see 2.5.3 **Indexicality**, p. 28) from their current status as ‘markers’ (in the eyes of many, but not all, MeFites). The ideas that are perpetuated about the M-Set are influenced by factors that are the most salient to the participants. It therefore matters how we measure indexical associations for the purpose of assessing the enregisterment of a variable. If we as researchers assess the status of a variable solely by what we perceive participants to be saying (i.e., what is conveyed via message chains), we might miss important demographic factors (e.g., geography, age) which can bias actual distributions. If we assess the status by those demographics alone, we miss out on the realities of social ascriptions in the community (e.g., what participants actually think, feel, and express about the variables, including the stereotypes they create and perpetuate).

The possibility of a distinction between the influential factors which are visible to participants and influential factors which are not visible in this way should be built into current models of indexicality, so that variables are not miscategorized along the hierarchy. For example, in researching the M-Set, it is conceivable that an inaccurate assessment could have been made of the variables as ‘indicators’, based on the observed correlations with age and the absence of social ascriptions about age, i.e., connections between age and pronunciation are not noticeable to participants and/or are not the subject of overt social commentary.

All influential social factors need to be included in the assessment of the indexical

status of a variable, and the presence or lack of overt social commentary or social ascriptions related to influential factors need to be explained not just in terms of their presence or absence, but in the context of the social environment, where the social commentary that occurs is partially shaped by the features of the modality in which it occurs. From this, the lack of overt social commentary involving demographic categories may not be expected in environments where those demographic factors are not easily observable by participants, and the indexicality status of a variable is assessed along other dimensions which are more likely to be the subject of overt social commentary (because participants can easily make those connections through observation). For example, some participants' high levels of involvement within the MetaFilter community made them more aware of community matters and more visible to others. These two factors translate to increased social capital, authority and influence for those participants (assuming their contributions were not negatively perceived by the community). The comments they made and the stances they took about the M-Set may be quoted and commented upon more frequently, may receive more favorites, and be more memorable to others. The indexical associations that result from this are therefore more likely to involve the stances in their comments and the perception of the involved participants' roles and authority within the community, rather than their ages, genders or geographic locations, etc.

Lastly, non-engagement participation modes (i.e., imagination and alignment) should be taken into account in enregisterment processes. The stereotypes or other ideas participants have about variables or varieties do not always manifest in directly observable interaction or communication, and yet they still may be very real and reified to the participants who hold those values. The current research did not take this directly into account, and as such, I did not ask participants about their attitudes towards the variables. This would have gone a long way toward more accurate indexical field maps. However, the stance data and survey rationales provided substantial data toward this purpose.

7.7.5 Implications for Onomastics

The exploration of naming practices and its link to identities and ideologies is increasingly important. Much of today's decision-making involving names—from playful banter concerning the pronunciation of 'doge'⁶⁴ or highly-sensitive sociopolitical decisions about how to refer to groups or territories in political unrest (e.g., the pronunciation of 'Ossetia/Ossetians', or the omission or addition of 'the' preceding 'Ukraine')—is conveyed through or influenced by CMC environments, while those very environments present both

64. An internet meme referring to pictures of Shiba Inu dogs (i.e., 'doges'), usually wearing scarves and shown with captions that represent the emerging grammar of 'dogespeak'.

new challenges and advantages in that mode of communication.

This has implications for how we create identities, how we ‘other’ individuals or groups, how we signal belonging, and other acts of evaluation and positioning. These acts are not bound by spoken environments, even if the features which comprise the entities in question concern spoken representations. As this case study of MetaFilter has shown, an entire history of social knowledge was built over many years, primarily in CMC text-based exchanges — sociophonetic variation does not necessarily need to have a community of FtF speakers for the variation to exist, for names to become shibboleths, or to be used as resources in creating identities. The variation has meaning in the minds of people, and its reach can go beyond what can be heard. As such, the impact of mass media and other non-FtF sources in the process of naming, forming conventions concerning the phonetic representations of names, and attaching evaluations to those names and representations, may often be underestimated or undervalued. This study has shown how all of these things can occur through even the most seemingly passive forms of community participation (although they are escalated by more active involvement), suggesting that a greater amount of social information may be being exchanged than we’ve previously accounted for.

7.8 Future Research Directions

The ways we might be able to generalize the results of this research over other linguistic variables (especially netologisms), or over other populations (especially CMC-based communities) leave many research paths yet to be explored. Additionally, the functioning of and relationship between macrosocial demographic factors versus social engagement factors in online communities needs to be investigated further to advance knowledge about enregisterment, and language variation and change in general. Possibilities for these topics will be explored further in the next section.

7.8.1 Future Research on Netologisms

In most written Englishes today, the relationship between graphemes to sounds is not a direct 1:1 correlation. As such, little is still known about how speakers from different English-speaking backgrounds might pronounce netologisms or nonce words that they experience primarily in text (if they have experienced them at all). A pronunciation study involving ambiguously-pronounceable word forms could lend insight to this process, advancing research from areas of cognitive perception and processing to sociolinguistic and sociocultural biases.

The selection of a range of speakers who may or may not have familiarity with these forms is an important area to explore, as it helps us to measure the effect of demographic factors, such as age or geography, from experiential factors, such as the influence of the

environments where those forms are encountered (and who uses them there). Therefore, for some variables and speakers, social engagement factors are somewhat removed (because they aren't experienced in any way), allowing the outcomes to be assessed independently from those potential influences and providing a baseline to measure the effects of social interaction. This would provide a better understanding of the starting point in pronunciation choices where social meaning is involved.

7.8.2 Future Research on Enregisterment

The examination of an enregistered item, e.g., 'MeFi', does not account for the enregisterment of an entire style or variety; many forms are needed to justify this claim. The M-Set here is just one case study, showing how a single lexical item's pronunciation gets enregistered. This process is not necessarily the same for other variables, even if they exist and are used within the exact same context.

Continued study which examines the enregisterment of other forms on MetaFilter would be worthwhile for assessing which features in the process of enregisterment are universal to that process (for the MetaFilter community), and which are unique or particular to specific variables. To expand this beyond the enregisterment of the pronunciation of forms would be useful as well, as it is likely that the enregisterment of phrases or ways of doing things on MetaFilter are subject to other internal (linguistic) and social constraints than the M-Set.

Focusing on the function of message chains in the process of the enregisterment of other forms, phrases, and ways of doing things on MetaFilter could lend further insight on the mechanisms of action in information dissemination and linguistic awareness. Message chains carrying messages about various variables could be compared in terms of their content, frequency or other discernible patterns over time.

Lastly, the study of the enregisterment of the MetaFilter variety as a whole, how it came to be, and how it is perceived as such, would be a nice complement to previous studies of enregisterment in FtF environments (Agha, 2003; Johnstone, Andrus, and Danielson, 2006; Beal, 2009). This could include analysis of additional aspects of variety enregisterment, such as commodification (Johnstone, 2009; Beal, 2009) and de-enregisterment (my term), which were researched for this thesis, but unable to be included owing to space constraints.

7.8.3 Future Research on Sociolinguistic Behavior in CMC Environments

It is worthwhile investigating how extendable the generalizations made about the M-Set variables and their enregisterment are to other netologisms used frequently in various CMC social spaces, where similar pronunciation debates may arise. For example, the pronunciation

of netologisms such as ‘doge’, ‘.gif’⁶⁵, ‘Linux’⁶⁶, and SNS’s such as ‘Imgur’, create opportunities for stances and evaluations to become indexical of identities, thereby allowing participants to position themselves using these variables as linguistic resources in achieving those aims.

More research is needed to understand how and why factors such as age and geography might influence linguistic behavior online, and how different geographic backgrounds may influence outcomes in different ways. This could be achieved with continued research, not just focusing on the environments in which the forms are used, but also the backgrounds of the participants, as they are conceivably in two ‘places’ at the same time. More attention needs to be paid to this interaction between spaces, as they are not distinct spheres and are therefore experienced simultaneously (Jurgenson, 2012). It is possible, and indeed highly likely, that many individuals feel or express identities which are in some sort of conflict—that is, their identity as a local to an online environment is somehow different or incompatible with the local identity they project in physical, bounded spaces, such as their workplace (usually) or other FtF setting. These avenues are worthwhile exploring, for a better understanding of how the personas are reconciled and of the reasons which underpin their perceptions.

Conducting a traditional ethnography may also reveal attitudes about variation, prescriptivism, community, etc., which this research was unable to investigate further (the scope of the study, the methodology chosen, and other constraints did not allow it). Understanding these influences better could have a range of potential benefits, depending which level of linguistic research is of interest. For example, discourse analysts may be interested in the way these attitudes influence the types of stances that participants make and how those stances are responded to in environments where the interaction is spontaneous but the language can be carefully planned. Researchers in onomastics and sociophoneticians would be able to better understand variation in place names, especially when political or economic consequences are at stake.

7.8.4 Summary of Future Research Directions

Regardless of which research strands may be pursued through future study, netologisms are entering the registers of communities online and offline at an increasing pace, commensurate with the integration of CMC technologies into our day-to-day lives. This research not only sheds light on how that integration occurs, but the ways in which we share perceptions and attitudes about those netologisms. I have put forth just one example of a successful methodology for obtaining answers to such research questions, and I present it in the hope that it will be improved upon and modified as needed. There are numerous other

65. Short for ‘Graphics Interchange Format’.

66. The name of an operating system, based on the first name of its creator, Linus Torvalds.

possibilities for continuation of the project or its themes and the suggestions made in this section are just a few ideas of what may be accomplished in the discovery of new insights in the area of language variation and change.

7.8.5 The Future of the M-Set

The M-Set — and the debate about the pronunciation of the M-Set variables — is not immune to fluctuations in interest. Granted, the debate about pronunciation will likely always be a part of MetaFilter's history and playful arguments about pronunciation of the M-Set variables will continue to occur as new members enter the community, attend meetups, listen to the podcast, etc. However, the community appears to have reached its peak with respect to intense focus and discussion on the subject. This was of course partially due to the current study, which brought direct attention to the topic for a focused period of time.

Given that new participants are continually joining the community (and some older users leave or stop participating) and those participants' start off with limited social engagement with other participants, especially FtF, it is likely that the variation will continue to be present, and that MeFites will continue to discuss the M-Set. This is especially probable given that discussing variation and matters of speech in general is something MeFites enjoy doing as part of their practice, especially on the MetaTalk subsite of MetaFilter.

7.9 Concluding Remarks

As our communication becomes increasingly computer-mediated, it is important to understand how social values, direct and indirect, come to be associated with variants of linguistic variables in seemingly unconventional ways. Relatedly, it is necessary to examine how these associations inform our identities, and the co-creation of our group identity. In the case of MetaFilter, the prominence of some features of our identities shift with online engagement — MeFites may be less concerned with the demographic characteristics of their online peers (which are hard for them to perceive or verify), but may be rather more focused on the opinions, attitudes, and ideas that are asserted by the individuals who participate. It is the latter which can be more readily capitalized upon, expanded, and identified with in this context.

The main outcomes of this study demonstrate how the ways one can engage socially within an online community exert influence on language behavior. These social measures transcend physical boundaries and are yet another way in which the purpose of the practice itself is central to understanding the linguistic behavior that results. The approach undertaken here has allowed the process of enregisterment to be observed, quantified and assessed, thereby providing a successful model for future mixed-methods research involving language variation and change in CMC environments.

Appendices

Appendix A: MetaTalk Posts About the Pronunciation of the M-Set

Three of the six selected MetaTalk posts included in this research were posted by MeFites in years prior to the pronunciation surveys. These posts made direct enquiries to the MetaFilter community about their pronunciation of the M-Set. The content of these three posts is pasted below.

“I’m sorry if this has been posted before, but I couldn’t find any reference to it. I’ve been wondering, how is one supposed to pronounce “MeFi”? Most abbreviations exist in a pronounced form as well as written. But “meffy” just sounds really weird to me. Does this ever come up when people talk about the site face to face? :-)”

posted by caveday in June, 2001 (40 total comments)

“How do you pronounce the shortened “MeFi”? “meh fee,” “me-figh” or something else?”

posted by insomnyuk in December, 2001 (36 total comments)

“MeFi: may-fee [mefi] or mee-fie [mifaɪ]? Mefite: mee-fight or may-fight? I was a little disconcerted at a meet-up to realize that some people pronounce these terms differently than I do. Is there a dominant pronunciation [sic]? Any other variants? (idle curiosity of a linguistics student)”

posted by heatherann in March, 2006 (102 total comments)

Appendix B: Survey MetaTalk Posts

Figure 69. 2010 Survey MetaTalk Post

MetaFilter Poll & Research Discussion Thread
March 24, 2010 5:12 PM [Subscribe](#)

Hi MeFites! This is the discussion thread for the **MetaFilter Poll**, appearing on the front page of MetaFilter, available for 5 days to logged-in site users. If you would like to take the poll, I would ask that you please do so BEFORE you read or participate in the discussion here. This is to help ensure that your responses to the poll questions are entirely your own decisions and are not influenced by the opinions of those participating in this thread. If you have already taken the poll or are not interested in taking the poll but would like to discuss it, please do so here.

If you have taken the poll already, thank you! I really appreciate you taking the time to contribute to what should be a really interesting data set about the MetaFilter community. Please share any thoughts, questions and concerns here. I am happy to discuss any aspect of this study with you, and I look forward to reading all of your comments!

About the Research: This poll is part of an ongoing sociolinguistic research project involving online communities and mediums of discourse. I am conducting this particular stage of research towards my completion of a Masters degree in Linguistics, and with the full consent of the MetaFilter moderators, San Francisco State University and its Institutional Review Board (IRB). The exact purpose of this poll will be revealed after all the data have been collected and analyzed (early May, 2010).

About the Poll: The poll is open to all logged-in site members who are over 18 years of age. Please submit the poll only once, using your primary MetaFilter account (not a spare or 'sockpuppet' account). There are 16 multiple choice and 2 short fill-in questions. The first several questions will be about your participation on the site. The remaining questions will be about your demographic background. The last questions are about participation in the poll, concluding with an opportunity to share any concerns or comments directly with the researcher (me).

Participation in any part of this study is entirely voluntary and you may opt-out at any time. You may also email me directly at mefistudy@gmail.com.

I hope you've enjoyed being a part of this, and thank you!!
posted by [iamkimiam](#) to **MetaFilter-Related** at 5:12 PM (472 comments total) [[add to favorites](#)]
[7 users marked this as a favorite](#) [!]

Figure 70. 2012 Survey MetaTalk Post

2012 MetaFilter Survey & Research Discussion Thread
 August 22, 2012 5:01 PM [Subscribe](#)

Hi MeFites! This is the discussion thread for the **2012 MetaFilter Pronunciation Survey**, available for 5 days to all logged-in site users. Even if you already took the 2010 survey, it'd be valuable to have you do it again. Additionally, this time around all survey participants will have the option to have a word frequency table of their MetaFilter posting history generated and emailed to them. **If you would like to take the 2012 survey, I would ask that you please do so BEFORE you read or participate in the discussion here (as well as before you view the previous MetaTalk thread about the 2010 survey).** This is to help ensure that your responses to these survey questions are entirely your own decisions and are not influenced by the opinions of those participating in this thread. If you have already taken the 2012 survey or are not interested in taking the survey but would like to discuss it, please go ahead and do so here.

If you have taken the 2012 survey, thank you! I really appreciate you taking the time to contribute to this ongoing research about language and MeFi. Please share any thoughts, questions and concerns here. I am happy to discuss any aspect of this study with you (including the **2010 survey**). As this second survey will be the last (and because I've been able to devote 3 years to this project so far), there's a lot more that I can share with you all. I'm looking forward to whatever you have to say!

About the Research: This survey is part of an ongoing sociolinguistic research project focused on language variation in an online community. I am conducting this final stage of research towards the completion of my PhD degree in Linguistics, and with the full consent of the MetaFilter moderators, the **University of York** and its Ethical Review Board (HSSEC).

About the Survey: The survey is open to all logged-in site members who are over 18 years of age. Please submit the survey only once, using your primary MetaFilter account (not a spare or 'sockpuppet' account). There are 20 questions in total. The first several questions will be about your participation on the site. The remaining questions will be about your demographic background. The survey concludes with a request to generate a word frequency table of all of your MetaFilter contributions, which you can choose to have emailed to you as well. Lastly, an opportunity to share any concerns or comments directly with the researcher (**Kim Witten, iamkimiam**) is provided.

Participation in any part of this study is entirely voluntary and you may opt-out at any time. You may also email me directly at mefistudy@gmail.com.

I hope you've enjoyed being a part of this, and thank you!!
 posted by [iamkimiam](#) to **MetaFilter-Related** at 5:01 PM (306 comments total) [[add to favorites](#)]
 14 users marked this as a favorite [!]

Appendix C: Survey Information & Consent Pages

Figure 71. 2010 MetaFilter Survey Information and Consent Page

Please participate in a MetaFilter poll!

Hi MetaFilter members! I, iamkimiam, am conducting a short poll, open to all site members who are over 18 years of age. Please submit the poll only once, using your primary MetaFilter account (not a spare or 'sockpuppet' account). If you would like to discuss any aspect of this poll and research, please see the related MetaTalk thread. You may also email the researcher (iamkimiam) directly at mefistudy@gmail.com. Please refer to the information below for additional details about voluntary participation in this research poll. Thanks!

Implied Consent to Participate in Research: All demographic information, as well as specific information associated with individual user profiles will be kept strictly confidential. Actual identities (usernames, user numbers and real names) will not be used in published reports. Data collected from this online poll will be used for continuing research in a sociolinguistic study involving online communities and mediums of discourse. This research is being conducted by a MetaFilter member, and with the full consent of the MetaFilter moderators, the academic institution that the researcher is affiliated with, and that institution's IRB (San Francisco State University). The exact purpose of this study will be revealed after all stages of data collection have been completed.

You have been invited to participate because you are a member of the site MetaFilter.com. There will be 16 multiple choice and 2 short fill-in questions to answer, and at the end of the poll there will be a checkbox allowing you to opt-in for possible future participation in an independent study to take place at a later date. Participation in any part of this study (including any related study in the future) is entirely voluntary and you may opt-out at any time. If you submit data, but later decide that you do not want some or all of your data to be included in the study, please email me at mefistudy@gmail.com.

You must be at least 18 years of age or older to participate in this portion of the research. Please use your primary MetaFilter account when submitting your poll. Your account must be active (not disabled or banned) and created prior to the appearance of the poll on the site. There are no risks or benefits to you in participating. You may choose to participate or not. If you do not wish to participate, you may simply move on, without completing or submitting the form. If you do participate, **completion and submission of the poll indicates your consent to the above conditions.**

There will be an opportunity to discuss any aspect of this poll and research in a MetaTalk thread, to appear shortly after the poll goes live. **Please complete the poll before participating in the discussion on MetaTalk. This is to help ensure that your responses to the poll questions are entirely your own decisions and are not influenced by the opinions of those participating in the MetaTalk discussion.**

If you have any concerns or questions regarding any aspect of this study, please contact the researcher (iamkimiam) at mefistudy@gmail.com. Please do not contact the moderators about this poll. They have given their full consent for this research, including the manner in which it is being conducted.

Please read and check all of the following (required):

- I have read the implied consent and I agree to the terms and conditions outlined here.
- I am over 18 years of age and I am participating through my primary MetaFilter account.
- I am the person associated with the user account that is currently logged into MetaFilter.

[Click HERE](#) to begin your participation in the MetaFilter Poll.

Figure 72. 2012 MetaFilter Survey Information and Consent Page

Please participate in the second MetaFilter survey!

Hi MetaFilter members! Once again, I am conducting a short survey about your MetaFilter and pronunciation preferences. Some of you may recall an earlier survey very similar to this one, back in March 2010. Even if you participated then, I'd like to ask that you join in one more time as the 2012 survey is improved and slightly expanded (but still should take no more than 5-10 minutes to complete). This survey is one page, 20 short questions (multiple choice, fill-ins and short answer) and you may answer as little or as much as you'd like. The survey is open to all site members who are over 18 years of age.

Regardless of whether or not you participated in this research before, your 2012 survey will help make the previous survey results better, allowing for comparison of data from two different points in time and gathering new data and word frequency tables (a new feature this time around-your word frequency table will be emailed to you as well if you wish to provide your email address at the end of the survey).

Please submit the survey only once, using your primary MetaFilter account (not a spare or 'sockpuppet' account). If you would like to discuss any aspect of this survey and research, see the related MetaTalk thread. You may also email the researcher ([Kim Witten](mailto:iamkimiam), iamkimiam on MetaFilter) directly at mefistudy@gmail.com. Please refer to the information below for additional details about voluntary participation in this research survey. Thanks!

Implied Consent to Participate in Research:

All demographic information, as well as specific information associated with individual user profiles will be kept strictly confidential. Actual identities (usernames, user numbers and real names) will not be used in published reports. Data collected from this online survey will be used for continuing research in a sociolinguistic study involving online communities and mediums of discourse. This research is being conducted by a MetaFilter member (iamkimiam), and with the full consent of the MetaFilter moderators, the academic institution that the researcher is affiliated with (University of York, UK), and that institution's ethical review board (HSSEC).

You have been invited to participate because you are a member of the site MetaFilter.com. There will be 20 questions to answer, and at the end of the survey there will be a checkbox allowing you to opt-in to have MetaFilter moderator cortex generate a word frequency table of your publicly available commenting history on MetaFilter. This text file can be emailed to you and will be sent to the researcher for inclusion in her ongoing research of word usage trends in text-based mediums. Participation in any part of this study (including any related study in the future) is entirely voluntary and you may opt-out at any time. If you submit data, but later decide that you do not want some or all of your data to be included in the study, please email the researcher (Kim Witten) at mefistudy@gmail.com.

You must be at least 18 years of age or older to participate in this portion of the research. Please be logged in using your primary MetaFilter account when submitting your survey (not a spare or 'sockpuppet' account). Your account must be active (not disabled or banned) and created prior to the appearance of the survey on the site. There are no risks or benefits to you in participating. You may choose to participate or not. If you do not wish to participate, you may simply move on, without completing or submitting the form. If you do participate, **completion and submission of the survey indicates your consent to the above conditions.**

There will be an opportunity to discuss any aspect of this survey and research in a MetaTalk thread, to appear shortly after the survey goes live. **Please complete the survey before participating in the discussion on MetaTalk, reviewing the previous MetaTalk survey thread, or visiting the researcher's website. This is to help ensure that your responses to the survey questions are entirely your own decisions and are not influenced or informed by the opinions of those participating in the MetaTalk discussion(s) or by the researcher.**

If you have any concerns or questions regarding any aspect of this study, please contact the researcher (Kim Witten, iamkimiam on MetaFilter) at mefistudy@gmail.com. Please do not contact the moderators about this survey. They have given their full consent for this research, including the manner in which it is being conducted.

Please read and check all of the following (required):

- I have read the implied consent and I agree to the terms and conditions outlined here.
- I am over 18 years of age and I am participating through my primary MetaFilter account.
- I am the person associated with the user account that is currently logged into MetaFilter.

[Click HERE](#) to begin your participation in the MetaFilter Survey.

Appendix D: The MetaFilter Surveys

Figure 73. The 2010 MetaFilter Survey

The MetaFilter Poll

You may stop participating at any time. Your poll data will not be included until you click 'submit' at the bottom of this page. Remember, all of your poll data is completely confidential and will not be shared with anyone at any time.

The first several questions will be about your participation on the site. The remaining questions will be about your demographic background. The last question will be about your participation in this poll, concluding with an opportunity to share any concerns or comments directly with the researcher.

1. How often do you read or participate in MetaFilter or any of the subsites? (For each column, please check the answer box that best describes your activity for that area of MetaFilter.)

<p>MetaFilter</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less 	<p>Ask MetaFilter</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less 	<p>MetaTalk</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less
<p>MetaFilter Projects</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less 	<p>MetaFilter Jobs</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less 	<p>MetaFilter Music</p> <ul style="list-style-type: none"> <input type="radio"/> At least once an hour <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less

2. How often do you listen to the following? (Check the answer box that best describes your activity.)

MetaFilter Podcast

- I listen to the podcast regularly
- I have listened to the podcast at least a few times
- I have listened to the podcast once or twice
- I have never listened to the MetaFilter podcast

MetaFilter Music Contributions

- I listen to MetaFilter Music tracks regularly
- I have listened to at least a few tracks on MetaFilter Music
- I have listened to a MetaFilter Music track once or twice
- I have never listened to any track on MetaFilter Music

3. How do you interact with other MetaFilter members outside of MetaFilter? (Check all that apply.)

- In person (excluding meetups)
- On MetaChat
- Through Instant Messaging
- On IRC
- Other (please list all related sites or ways you interact with other MetaFilter members)
- I do not interact with MetaFilter members outside of MetaFilter

4. Have you ever been to a MetaFilter Meetup? (Check the answer box that best describes your activity.)

- Yes, I go to Meetups regularly
- Yes, I have gone to at least a few Meetups
- I have attended Meetups once or twice
- No, I have never been to a Meetup

5. Think about the word 'MeFi'. Say it out loud if you need to. Which audio recording below best matches the way you currently say 'MeFi'? (Click on each option to hear digital audio samples. Next to each word is an approximated transcription, written in brackets, using the International Phonetic Alphabet (IPA) corresponding to each audio sample. If you do not know IPA, do not worry; just choose by using the audio samples associated with each option and completely ignore the transcriptions. The order of these audio samples is random.)

<input type="radio"/> MeFi - [<input type="radio"/> mifi]	<input type="radio"/> MeFi - [<input type="radio"/> msfɪ]	<input type="radio"/> MeFi - [<input type="radio"/> mɛsfɪ]
<input type="radio"/> MeFi - [<input type="radio"/> mɛfi]	<input type="radio"/> MeFi - [<input type="radio"/> mɛrfɪ]	<input type="radio"/> MeFi - [<input type="radio"/> mɪfai]

6. How strongly do you feel that the pronunciation of MeFi that you have chosen directly above is the one you (would) use exclusively (as opposed to the other options listed above)?

Indifferent — I (might) use other forms ● ● ● ● ● Very Strongly — I (would) only use my chosen form

7. Please share your reason(s) for pronunciation that you have chosen. Also, if you used to pronounce MeFi differently, please describe how and try to explain what you believe might have influenced your pronunciation change:

8. Think about the word 'MeFite'. Say it out loud if you need to. Which recording below best matches the way you currently say 'MeFite'? (Click on each option to hear digital audio samples. Next to each word is an approximated transcription, written in brackets, using the International Phonetic Alphabet (IPA) corresponding to each audio sample. If you do not know IPA, do not worry; just choose by using the audio samples associated with each option and completely ignore the transcriptions. The order of these audio samples is random.)

<input type="radio"/> MeFite - [<input type="radio"/> mɛfɪt]	<input type="radio"/> MeFite - [<input type="radio"/> mɛrfɪt]	<input type="radio"/> MeFite - [<input type="radio"/> mɛfɪt]
<input type="radio"/> MeFite - [<input type="radio"/> mɛfɪt]	<input type="radio"/> MeFite - [<input type="radio"/> mɪfɪt]	<input type="radio"/> MeFite - [<input type="radio"/> mɪfɪt]

9. How strongly do you feel that the pronunciation of MeFite that you have chosen directly above is the one you (would) use exclusively (as opposed to the other options listed above)?

Indifferent — I (might) use other forms ● ● ● ● ● Very Strongly — I (would) only use my chosen form

Figure 74. The 2010 MetaFilter Survey (continued)

10. Please share your reason(s) for pronunciation that you have chosen. Also, if you used to pronounce MeFite differently, please describe how and try to explain what you believe might have influenced your pronunciation change:

11. How much thought you have given to the pronunciation of MeFI or MeFite, prior to this poll:

- I have given this some considerable thought prior to this poll
- I have thought about this only briefly before now
- I have never thought about this before now

12. Please state your age and your sex/gender:

Age:

- Male
- Female
- Transgender
- Other (please explain)

13. Is English your native language?

- Yes
- No

14. Do you speak any other languages proficiently?

- Yes. If so, what other languages have you spoken or studied:
- No

15. Are you currently living in the US?

- Yes
- No

If no, what is your current country of residence?

16. Please select the dialect of English that best describes the one you currently speak. (Check all that apply. Elaborate if you can, using the blank provided to specify an accent, geographic region or other affiliation you have with your currently spoken variety of English.)

- American
- Australian
- British
- Canadian
- Other English dialect
- I do not know the dialect of English that I speak.

17. What is your Race/Ethnicity?

- African American
- Asian
- Caucasian
- Hispanic
- Native American
- Decline to State
- Other

18. Would you be interested in participating in further independent research about MetaFilter, conducted by iamkimiam?

- Yes
- No

If yes, Email's in the profile, or Please use this email address instead

19. Please share any comments or questions you have for the researcher here (I will read this section and respond to any concerns you have directly. You may also email me at mefistudy@gmail.com if you wish):

Figure 75. The 2012 MetaFilter Survey

The MetaFilter Poll

You may stop participating at any time. Your survey data will not be included until you click 'submit' at the bottom of this page. Remember, all of your survey data is completely confidential and will not be shared with anyone at any time. Any and all survey questions that you do not wish to answer may be left blank.

The first several survey questions will be about your participation on the site. The remaining questions will be about your demographic background. The last question will be about generating a word frequency table of your comment history. Lastly, the survey will conclude with an opportunity to share any concerns or comments directly with the researcher.

1. How often do you read or participate in MetaFilter or any of the subsites? (For each column, please check the answer box that best describes your activity for that area of MetaFilter.)

<p>MetaFilter</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never 	<p>Ask MetaFilter</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never 	<p>MetaTalk</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never
<p>MetaFilter Projects</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never 	<p>MetaFilter Jobs</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never 	<p>MetaFilter Music</p> <ul style="list-style-type: none"> <input type="radio"/> More than a few times a day <input type="radio"/> A few times a day <input type="radio"/> At least once a day <input type="radio"/> About every other day <input type="radio"/> Once every few days <input type="radio"/> Once a week or less <input type="radio"/> Never

MetaFilter IRL

- More than a few times a day
- A few times a day
- At least once a day
- About every other day
- Once every few days
- Once a week or less
- Never

2. How often do you listen to the following? (Check the answer box that best describes your activity.)

MetaFilter Podcast

- I listen to the podcast regularly
- I have listened to the podcast at least a few times
- I have listened to the podcast once or twice
- I have never listened to the MetaFilter podcast

MetaFilter Music

- I listen to MetaFilter Music tracks regularly
- I have listened to at least a few tracks on MetaFilter Music
- I have listened to a MetaFilter Music track once or twice
- I have never listened to any track on MetaFilter Music

3. Do you ever interact with other MetaFilter members outside of MetaFilter? If so, how? (Check all that apply.)

- In person (excluding meetups)
- On MetaChat, MeFight Club or other MetaFilter-related spin-off sites
- On social networking sites (including Twitter, Google+, Facebook or other similar sites)
- Through Instant Messaging
- On IRC
- Other (please list all related sites or ways you interact with other MetaFilter members)
- I rarely or never interact with MetaFilter members outside of MetaFilter

4. Have you ever been to a MetaFilter meetup? (Check the answer box that best describes your activity.)

- Yes, I go to Meetups regularly
- Yes, I have gone to at least a few Meetups
- I have attended Meetups once or twice
- No, I have never been to a Meetup

5. Think about the word 'MeFi'. Say it out loud if you need to. Which audio recording below best matches the way you currently say 'MeFi'? (Click on each option to hear digital audio samples. Next to each word is an approximated transcription, written in brackets, using the International Phonetic Alphabet (IPA) corresponding to each audio sample. If you do not know IPA, do not worry; just choose by using the audio samples associated with each option and completely ignore the transcriptions. The order of these audio samples is random.)

<input type="radio"/> MeFi - [ɹ̥ mɛfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɛfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɛfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɛɪfɪ]
<input type="radio"/> MeFi - [ɹ̥ mɛfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɪfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɪfɪ]	<input type="radio"/> MeFi - [ɹ̥ mɛɪfɪ]

6. How strongly do you feel that the pronunciation of MeFi that you have chosen directly above is the one you (would) use exclusively (as opposed to the other options listed above)?

Indifferent — I (might) use other forms Very Strongly — I (would) only use my chosen form

● ● ● ● ● ●

Figure 77. The 2012 MetaFilter Survey (continued)

19. What is your ethnicity?
 (This is free-form, go nuts!)

20. What is your nationality?
 (This is free-form, go nuts!)

Word Frequency Table — Consent

Please click the checkbox below if you give your consent to have a word frequency table generated of all of your publicly available posting history on MetaFilter.

- Yes, I will allow a word frequency table to be generated and emailed to the researcher.
 - I would like to receive a copy of this text file at the following email address:
- No, don't generate a word frequency table.

Please share any comments or questions you have for the researcher here.
(I will read this section and respond to any concerns you have directly. You may also email me at mefistudy@gmail.com if you wish):

Appendix E: The MetaFilter Register

This MetaTalk post related to the research enquired about norms, in-jokes and memes on MetaFilter, to directly elicit qualitative data about the MetaFilter register, and enregisterments that were specific to or salient in the community. The main text of that post is repasted below (the somewhat informal wording of the post is in keeping with the style of other MetaTalk posts on MetaFilter):

“Hi Everybody! I’m trying to come up with a “MeFi glossary of terms / phrases / behaviors” to include in my PhD dissertation about linguistics and MetaFilter. If you could help me with this, that would be fantastic. I’m looking to create a list of the types of things that say to you (or to others) “Oh, that’s so MetaFilter!” or “That sounds like something a MeFite would say.” This could be anything from the use of a single character to a phrase or meme, to a way of doing things here. If you can think of something to add, please share it in the comments. Thanks!”
posted by iamkimiam in February, 2012 (331 total comments)

This post resulted in a list of items that exemplify enregistered terms or other features, memes, etc. that are considered to be part of the MetaFilter register:

. [the obituary dot]	hope me
[X]filter	Hurf durf butter eater
Are you friends of Matt/Jessamyn?	I hope your head falls off
Ask vs. Guess culture	I’m sorry, that won’t be possible
banhammer	I’VE BEEN SILENCED ALL MY LIFE
beans, beanplate, beanplating, beanplaters	IANAL/IANAD
DTMFA	maroon
Everyone needs a hug	MeFi/MeFite
favorite	MetaFilter: taglines
FIAMO	nerd thunderdome
flag	Special snowflake
flameout	taters
FPP	the blue, the gray, the green
Go die in a fire	We are from the internet
GRAR	We have cameras
grilled cheese sandwich	WTF MATT
hamburger, {}	Zamboni
Hon	

Appendix F: Orthographically Similar Forms to ‘MeFi’

Table 17 lists relevant spelling forms that match the CVCV template of ‘MeFi’. Regex Dictionary, a searchable online dictionary based on The American Heritage Dictionary of the English Language (4th ed.), was used to find analogous orthographic forms (column 2). In column 3, the first 40 results of each query from the Contemporary Online Corpus of American English (COCA) can be found. COCA is a 400 million word balanced corpus of spoken, online, and printed text, spanning 10 years and continually updated. In the final column, place names and other found neologisms and abbreviations that fit the pattern specified for that row are listed.

Table 17. Analogy by Orthography

Search String	Regex Dictionary (all results; alphabetical)	COCA Corpus (top 40 results; sorted by frequency)	Other words (alphabetical)
C<e>C<i>	cedi, deli, deni, kepi, peri, semi, yeti	Levi, deli, Teri, demi, Lexi, SETI, semi, Jedi, Desi, Ceci, Keri, deci, Leni, Debi, Jeri, Geri, Meri, peri, yeti, Negi, Ceti, Seri, Devi, neri, Jeni, Remi, Beni, Deri, Teti, heli, hemi, ceri, Celi, femi, dedi, Feri, Pepi, medi, Nemi, Keli	Redi, refi
<me>CV	-mere, Mede, meme, memo, menu, mere, mesa, mete	mere, menu, memo, mesa, mega, meme, meta, melo, mete, mero, mepa, mele, meno, mema, medi, meza, meca, meru, meso, mera, mela, mene, meze, mese, meti, mesi, meja, meco, meda, meto, mego, meli, medu, MeWe, mewa, medo, mede, megu, mevo, meka	MePa
C<e>CV	-cene, -gene, -geny, -mere, -pede, Beja, bema, beta, bevy, ceca, cede, cedi, cere, cero, cete, deco, defy, deke, dele, deli, deme, demo, demy, Dene, dene, deni, deny, dewy, feme, fere, feta, fete, gene, geta, helo, heme, here, hero, jete, keno, kepi, Leda, leno, Leto, levo, levy, Mede, meme, memo, menu, mere, mesa, mete, nene, neve, pepo, pere, peri, peso, rede, redo, rely, repo, rete, sego, seme, semi, sene, sera, sere, seta, sexy, tegu, tepa, tete-a-tete, Tewa, Veda, Vega, vela, vena, very, veto, weka, were, yeti, zebu, zero, zeta	were, very, here, hero, gene, mere, zero, pete, rely, menu, deny, sexy, veto, Peru, Reno, memo, beta, levy, Vera, mesa, Levi, FEMA, Lena, Rene, defy, deli, Leno, Vega, demo, pena, Remy, Zeke, deco, Teri, Neha, sera, bela, weve, feta, pero	Devo, Lego, MePa, Nemo, PETA, Rena, Teva, Zena, Zeta
CVC<i>	bani, bidi, cami, cedi, deli, deni, divi, foci, haji, hara-kiri, hari-kari, Hopi, kaki, kami, kepi, kiwi, lari, lati, loci, loti, magi, mahi, maxi, midi, mini, muni, Nazi, nisi, nori, Pali, peri, pili, puli, puri, Rabi, ragi, raki, rami, rani, saki, sari, sati, semi, sori, Sufi, tabi, tali, taxi, tiki, tipi, titi, topi, tori, Tupi, vagi, wadi, wiki, yagi, yeti, yogi, yoni, ziti, zori, Zuni	Nazi, taxi, mini, Lori, Toni, Levi, Yuri, coli, deli, MUNI, mimi, Bali, Hopi, Kofi, Mali, Teri, Fuji, MIDI, Tori, sami, sari, Jimi, Dali, kari, Jodi, yogi, demi, Laci, Lexi, Lani, SETI, Joni, tiki, kiwi, semi, Judi, lili, Fiji, Jedi, bibi	HiFi, lofi, Lodi, Redi, refi, Rudi, WiFi, Wiki

Note: Words in this chart are likely to fit in more than one category. Words are orthographically represented here as they are most commonly seen. Words from COCA whose meaning or use were severely limited in scope (one rare source for all instances) were removed from this list and replaced with the next available word.

Appendix G: Summary of Findings Tables

Table 18. Summary of Panel Data as Compared to Survey Data by Various Measures

Measure	Significance of Data Set Comparison		Notes
	2010	2012	
Overall Distribution of 'MeFi'	n/s	n/s	
Overall Distribution of 'MeFite'	n/s	n/s	
Exclusive Preference	n/s	*	Panel participants felt more strongly than overall Survey participants.
Amount of Thought Given	n/s	***	This research likely influenced the 2012 outcome.
Native Language	n/s	n/s	
Language Experience	n/s	n/s	
Geography	n/s	n/s	
Age	*	n/s	The 2010 Panel participants were slightly older than the 2010 overall Survey participants, on average.
Gender	***	n/s	The 2010 Survey was skewed male, whereas the 2010 Panel participants and all 2012 data were more gender balanced, and to equal degrees.
Year of Joining MetaFilter	n/s	n/a	
MetaFilter Visitation Frequency	n/s	***	2012 Panel participants read MeFi more frequently than the overall 2012 Survey population.
AskMetaFilter Visitation Freq.	***	n/s	2010 Panel participants read AskMe more frequently than the overall 2010 Survey population.
MetaTalk Visitation Frequency	***	***	Panel participants read MeTa more frequently than the overall Survey populations.
Podcast Listening Frequency	***	***	Panel participants listened to the podcast more frequently than overall Survey populations.
Meetup Attendance Frequency	***	***	Panel participants attended meetups more frequently than overall Survey populations.

Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, n/s = no significant differences between the Survey Data and the Panel Data were observed, n/a = comparisons between the Survey Data and the Panel Data were not possible for the measure.

Table 19. Summary of Change Over Time by Various Measures

Measure	Significance of Change Over Time		Notes
	Survey Data	Panel Data	
Overall Distribution of 'MeFi'	**	n/s	See 5.2.4.3 Change over time for the M-Set – Panel Data, p. 115 for a full explanation of these results.
Overall Distribution of 'MeFite'	*	**	Increased preference for 1b – [mifaɪ(t)] variants over time.
Exclusive Preference	**	**	Participants felt more strongly about exclusive use of their variant over time.
Amount of Thought Given	***	***	This research likely influenced these outcomes over time.
Native Language	*	n/a	Significant increase in non-native English speakers in Survey populations over time.
Language Experience	n/a	n/a	See 4.5.1.1 Changes Made in the 2012 MetaFilter Survey, p. 104 for an explanation of why the surveys were not comparable by this measure.
Geography	n/s	n/s	
Age	n/s	n/s	
Gender	***	n/s	Significant shift toward more equal gender balance in Survey populations over time.
Year of Joining MetaFilter	n/a	n/a	This is a fixed measure which cannot change over time.
MetaFilter Visitation Frequency	n/a	n/a	See 4.5.1.1 Changes Made in the 2012 MetaFilter Survey, p. 104 for an explanation of why the surveys were not comparable by these measures.
AskMetaFilter Visitation Freq.	n/a	n/a	
MetaTalk Visitation Frequency	n/a	n/a	
Podcast Listening Frequency	n/s	*	Panel participants showed significant increase in podcast listening frequency over time.
Meetup Attendance Frequency	n/s	***	Panel participants showed significant increase in meetup attendance frequency over time.

Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, n/s = no significant change over time was observed for the measure, n/a = change over time calculations were not applicable to the measure, or change over time was not analyzed for the measure for other reasons.

Table 20. Summary of Pronunciation Outcomes by Various Measures

Measure	Significance (by Year and Data Set)				Hypothesis Met?
	2010		2012		
	Survey	Panel	Survey	Panel	
Overall Distribution of 'MeFi'	***	***	***	***	Yes
Overall Distribution of 'MeFite'	***	***	***	***	Yes
'MeFi' Pronunciation by ...					
Exclusive Preference	***	***	***	***	Yes
Amount of Thought Given	n/s	n/s	n/s	n/s	No
Native Language	n/s	n/s	n/s	n/s	No
Language Experience	***	**	*	n/s	Partially
Geography	***	***	***	***	Yes
Age	***	n/s	***	*	Partially
Gender	n/s	n/s	n/s	n/s	Yes
Year of Joining MetaFilter	n/s	n/s	n/s	n/s	No
MetaFilter Visitation Frequency	n/s	n/s	n/s	n/s	No
AskMetaFilter Visitation Freq.	n/s	n/s	**	n/s	Partially
MetaTalk Visitation Frequency	**	n/s	**	n/s	Partially
Podcast Listening Frequency	***	*	n/s	n/s	Partially
Meetup Attendance Frequency	*	n/s	**	n/s	Partially

Significance levels: * = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$, n/s = the result was not significant

Appendix H: Regression Models

Table 21. Binary Logistic Regression Model Results – 2010 Model Data

2010 Model Data	B	SE	Sig.	Odds Ratio	95% CI for odds ratio	
					Lower	Upper
(Constant)	-0.153	0.292		0.858	0.483	1.517
Age Group (baseline: 19–24)						
25-29	0.037	0.197		1.037	0.706	1.531
30-34	-0.200	0.203		0.819	0.551	1.220
35-39	-0.430	0.223	.	0.651	0.420	1.007
40-44	-0.552	0.253	*	0.576	0.349	0.941
45-49	-0.653	0.327	*	0.521	0.269	0.973
50-54	-0.888	0.432	*	0.412	0.166	0.920
55+	-0.186	0.442		0.830	0.334	1.919
Country (baseline: United States)						
Canada	1.118	0.178	***	3.058	2.154	4.337
United Kingdom	0.772	0.217	***	2.163	1.404	3.299
Australia	0.640	0.376	.	1.896	0.880	3.890
Year Joined MetaFilter (baseline: 1999–2003)						
2004–2007	-0.350	0.162	*	0.704	0.513	0.971
2008	-0.580	0.221	**	0.560	0.362	0.861
2009	-0.451	0.219	*	0.637	0.413	0.977
2010	-0.287	0.366		0.750	0.357	1.512
Podcast Listening (baseline: Never Listened to Podcast)						
Listened Once or Twice	-0.046	0.145		0.955	0.718	1.268
At Least a Few Times	-0.238	0.178		0.788	0.554	1.113
Listened Regularly	-0.679	0.256	**	0.507	0.301	0.825
Amount of Thought Given to Pronunciation (baseline: No Thought Prior)						
Brief Thought Prior	0.367	0.144	*	1.444	1.091	1.922
Considerable Thought Prior	0.498	0.202	*	1.646	1.108	2.444
Strength of Preference for Chosen Variant (baseline: Level 1: Indifferent)						
Level 2	-0.074	0.234		0.929	0.587	1.467
Level 3	-0.339	0.234		0.713	0.450	1.126
Level 4	-0.718	0.195	***	0.488	0.333	0.715
Level 5: Very Strongly	-1.300	0.192	***	0.272	0.187	0.397

$R^2 = 0.083$ (Hosmer & Lemeshow), 0.09 (Cox & Snell), 0.132 (Nagelkerke). Model $\chi^2(23) = 159.321$
 $p < 0.001$ ***

Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

Table 22. Binary Logistic Regression Model Results – 2012 Model Data

2012 Model Data	B	SE	Sig.	Odds Ratio	95% CI for odds ratio	
					Lower	Upper
(Constant)	0.613	0.424		1.846	0.801	4.233
Age Group (baseline: 19–24)						
25-29	0.219	0.268		1.245	0.740	2.122
30-34	-0.331	0.268		0.718	0.426	1.223
35-39	-0.644	0.285	*	0.525	0.301	0.923
40-44	-0.458	0.292		0.633	0.357	1.126
45-49	-0.845	0.345	*	0.430	0.216	0.841
50-54	-0.658	0.398	.	0.518	0.232	1.113
55+	-0.820	0.433	.	0.440	0.183	1.006
Country (baseline: United States)						
Canada	0.811	0.204	***	2.249	1.505	3.347
United Kingdom	0.965	0.214	***	2.625	1.719	3.985
Australia	-0.303	0.398		0.738	0.319	1.544
Year Joined MetaFilter (baseline: 1999–2003)						
2004–2007	-0.340	0.195	.	0.712	0.487	1.046
2008	-0.505	0.265	.	0.604	0.358	1.011
2009	-0.753	0.268	**	0.471	0.277	0.792
2010	-0.732	0.268	**	0.481	0.283	0.810
2011	-0.632	0.275	*	0.532	0.309	0.908
2012	-0.668	0.357	.	0.513	0.250	1.019
Subsite Visitation – AskMetaFilter (baseline: Visits once a week or less)						
A few times a day or more	-0.451	0.216	*	0.637	0.419	0.976
At least once a day	-0.236	0.247		0.790	0.487	1.284
About every other day	-0.248	0.273		0.780	0.455	1.331
Once every few days	0.154	0.265		1.166	0.694	1.960
Never	-0.540	0.847		0.583	0.081	2.627
Subsite Visitation – MetaTalk (baseline: Visits once a week or less)						
A few times a day or more	0.626	0.220	**	1.871	1.217	2.886
At least once a day	0.611	0.219	**	1.843	1.200	2.835
About every other day	0.609	0.214	**	1.839	1.209	2.801
Once every few days	0.310	0.218		1.363	0.887	2.088
Never	0.280	0.261		1.323	0.789	2.202
Podcast Listening (baseline: Never Listened to Podcast)						
Listened Once or Twice	-0.117	0.160		0.890	0.650	1.215
At Least a Few Times	-0.301	0.205		0.740	0.493	1.100
Listened Regularly	-0.774	0.292	**	0.461	0.255	0.803
Meetup Attendance (baseline: Never Been to a Meetup)						
Been Once or Twice	-0.358	0.180	*	0.699	0.489	0.990
Been at Least a Few Times	-0.444	0.207	*	0.641	0.424	0.956
Attends Regularly	-0.612	0.332	.	0.542	0.273	1.013
Amount of Thought Given to Pronunciation (baseline: No Thought Prior)						
Brief Thought Prior	0.160	0.181		1.174	0.826	1.682
Considerable Thought Prior	0.543	0.223	*	1.720	1.114	2.670
Strength of Preference for Chosen Variant (baseline: Level 1: Indifferent)						
Level 2	-0.206	0.271		0.814	0.478	1.383
Level 3	-0.636	0.251	*	0.530	0.323	0.865
Level 4	-1.045	0.218	***	0.352	0.229	0.540
Level 5: Very Strongly	-1.946	0.223	***	0.143	0.092	0.221

$R^2 = 0.125$ (Hosmer & Lemeshow), 0.132 (Cox & Snell), 0.194 (Nagelkerke). Model $\chi^2(38) = 222.842$ $p < 0.001$ ***
 Significance codes: *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$, . $p < 0.1$

References

- Adams, M., 2009. Power, Politeness, and the Pragmatics of Nicknames. *Names: A Journal of Onomastics*, 57(2), pp. 81–91.
- Agha, A., 2003. The social life of cultural value. *Language & Communication*, 23(3-4), pp. 231–273.
- Agha, A., 2004. Registers of Language. In A. Duranti, ed. *A Companion to Linguistic Anthropology*. Oxford: Blackwell, pp. 23–45.
- Agha, A., 2005. Voice, Footing, Enregisterment. *Journal of Linguistic Anthropology*, 15(1), pp. 38–59.
- Aitken, A.J., 1984. Scottish Accents and Dialects. In P. Trudgill, ed. *Language in the British Isles*. pp. 94–98.
- Ali-Hasan, N., 2005. *MetaFilter: an analysis of a community weblog*. University of Michigan School of Information.
- Altendorf, U. & Watt, D., 2008. The Dialects in the South of England: Phonology. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 194–222.
- Androutsopoulos, J.K., 2008. Potentials and Limitations of Discourse-Centred Online Ethnography. *Language@Internet*, 5(9), pp. 1–20.
- Androutsopoulos, J.K. & Georgakopoulou, A., 2003. *Discourse Constructions of Youth Identities*, John Benjamins Publishing Company.
- Armstrong, A. & Hagel, J., 2009. The Real Value of Online Communities. In E. Lesser, M. Fontaine, & J. Slusher, eds. *Knowledge and Communities*. Routledge, pp. 85–98.
- Auerbach, D., 2014. MetaFilter layoffs: Why has Google forsaken the legendary Internet forum? *Slate.com*. Available at: http://www.slate.com/articles/technology/bitwise/2014/05/metafilter_layoffs_why_has_google_forsaken_the_legendary_internet_forum.html [Accessed June 2, 2014].
- Aura, S. & Hess, G.D., 2004. *What's In A Name?*, Available at: http://www.cj-resources.com/CJ_Female_Offenders_pdfs/whats_in_a_name_women_black_feminism_and_beyond_-_Collins_1996.pdf [Accessed February 15, 2012].
- Bacon, I., 1993. Out on the internet. *MacUser*, 9, pp. 209–214.
- Bailey, G., 2008. Real and Apparent Time. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes, eds. *Handbook of Language Variation and Change*. Oxford, UK: Blackwell Publishing Ltd, pp. 312–332.
- Banks, D., 2013. #TtW13 Presentation Preview: On the Political Origins of Digital Dualism. *Cyborgology*. Available at: <http://thesocietypages.org/cyborgology/2013/02/23/ttw13-presentation-preview-on-the-political-origins-of-digital-dualism/> [Accessed August 14, 2014].
- Baron, N.S., 1998. Letters by phone or speech by other means: the linguistics of email. *Language and Communication*, 18(2), pp. 133–170.
- Baron, N.S., 2003. Language of the Internet. In A. Farghali, ed. *The Stanford handbook for language engineers*. Stanford, California: CSLI, pp. 59–127.
- Baron, N.S., 2004. See you Online: Gender Issues in College Student Use of Instant Messaging. *Journal of Language and Social Psychology*, 23, pp. 397–423.
- Barton, D. & Lee, C., 2013. *Language Online: Investigating Digital Texts and Practices*, London, New York: Routledge.
- Bauer, L., 1998. Is there a class of neoclassical compounds, and if so is it productive? *Linguistics*, 36(3), pp. 403–422.
- Bauer, L., 2008. Derivational Morphology. *Language and Linguistics Compass*, 2(1), pp. 196–210.
- Bauer, L. & Warren, P., 2008. New Zealand English: Phonology. In K. Burridge & B. Kortmann, eds. *Varieties of English: The Pacific and Australasia*. Berlin and New York: Mouton de Gruyter, pp. 39–63.
- Beal, J.C., 2008. English Dialects in the North of England: Phonology. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 122–144.
- Beal, J.C., 2009. Enregisterment, Commodification, and Historical Context: “Geordie” versus “Sheffieldish.” *American Speech*, 84(2), pp.138–156.
- Bell, A., 2001. Back in style: reworking audience design. In P. Eckert & J. R. Rickford, eds. *Style and Sociolinguistic Variation*. Cambridge University Press, pp. 139–169.
- Blommaert, J., 2007. On scope and depth in linguistic ethnography. *Journal of Sociolinguistics*, 11(5), pp. 682–688.
- Boberg, C., 1997. *Variation and Change in the Nativization of Foreign (a) in English*. Diss. Philadelphia, PA: University of Pennsylvania.
- Boberg, C., 2000. Geolinguistic diffusion and the U.S.-Canada border. *Language Variation and Change*, 12(1), pp. 1–24.
- Boberg, C., 2004. Real and Apparent Time in Language Change: Late Adoption of Changes in Montreal English. *American Speech*, 79(3), pp. 250–269.

References

- Boberg, C., 2008a. English in Canada: phonology. In E. W. Schneider, ed. *Varieties of English: The Americas and the Caribbean*. Berlin and New York: Mouton de Gruyter, pp. 144–160.
- Boberg, C., 2008b. Regional Phonetic Differentiation in Standard Canadian English. *Journal of English Linguistics*, 36(2), pp. 129–154.
- Boberg, C., 2010. *The English language in Canada: Status, history, and comparative analysis*, Cambridge, UK: Cambridge University Press.
- Boberg, C., 2013. Surveys: The Use of Written Questionnaires in Sociolinguistics. In C. Mallinson, B. Childs, & G. Van Herk, eds. *Data Collection in Sociolinguistics: Methods and Applications*. New York, NY, USA: Routledge.
- Bourdieu, P., 1986. The forms of capital. In J. G. Richardson, ed. *Handbook of Theory and Research for the Sociology of Education*. Greenwood Press, pp. 241–258.
- Boyd, D.M. & Ellison, N.B., 2008. Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1), pp. 210–230.
- Bucholtz, M., 1996. Geek the girl: Language, femininity, and female nerds N. Warner et al., eds. *Gender and belief systems*, pp. 119–131.
- Bucholtz, M., 1999. “Why be normal?”: Language and identity practices in a community of nerd girls. *Language in Society*, 28(2), pp. 203–223.
- Burridge, K. & Kortmann, B., 2008. Introduction: varieties of English in the Pacific and Australia. In K. Burridge & B. Kortmann, eds. *Varieties of English: The Pacific and Australasia*. Berlin and New York: Mouton de Gruyter, pp. 23–36.
- Cameron, D., Frazer, E., Harvey, P., Rampton, M.B.H., Richardson, K. 1992. *Researching language: Issues of power and method*. New York: Routledge.
- Cannon, G., 1987. Abbreviations and Acronyms in English Word-Formation. *American Speech*, 64(2), pp. 99–128.
- Carney, E., 1994. *Survey of English Spelling*, Taylor & Francis.
- Chambers, J.K., 2002. Patterns of Variation including Change. In J. K. Chambers, P. Trudgill, & N. Schilling-Estes, eds. *The Handbook of Language Variation and Change*. Malden, MA: Blackwell Publishers, pp. 358–361.
- Chomsky, N., 1965. *Aspects of the Theory of Syntax*, Cambridge, MA: MIT Press.
- Clark, U., 2008. The English West Midlands: Phonology. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 145–177.
- Corbeil, J.P., 2012. French and the francophonie in Canada. Language, 2011 Census of Population. Statistics Canada. Catalogue no. 98-314-X2011003. Available at: http://www12.statcan.gc.ca/census-recensement/2011/as-sa/98-314-x/98-314-x2011003_1-eng.pdf.
- Crystal, D., 2006. The medium of Netspeak. In *Language and the Internet*. Cambridge: Cambridge University Press, pp. 26–65.
- Crystal, D., 2011. *Internet linguistics : a student guide*, New York NY: Routledge.
- Danesi, M., 2011. What’s in a Brand Name? A Note on the Onomastics of Brand Naming. *Names: A Journal of Onomastics*, 59(3), pp. 175–185.
- David, J., 2011. Commemorative Place Names — Their Specificity and Problems. *Names: A Journal of Onomastics*, 59(4), pp. 214–228.
- Davies, B., 2005. Communities of practice: Legitimacy not choice. *Journal of Sociolinguistics*, 9(4), pp. 557–581.
- Dickens, C., 2004. *David Copperfield*, London: Penguin Classics.
- Duranti, A., 1997. *Linguistic Anthropology*, Cambridge: Cambridge University Press.
- Eckert, P., 1988. Adolescent social structure and the spread of linguistic change. *Language in Society*, 17(2), pp. 183–207.
- Eckert, P., 1989. *Jocks and Burnouts: Social Categories and Identity in the High School*, Teachers College Press.
- Eckert, P., 1996. Vowels and Nail Polish: The Emergence of Linguistic Style in the Preadolescent Heterosexual Marketplace. In N. Warner et al., eds. *Gender and Belief Systems: Proceedings of the Fourth Berkeley Women and Language Conference*. Berkeley, pp. 183–190.
- Eckert, P., 2000. *Linguistic variation as social practice: the linguistic construction of identity in Belten High*, Blackwell.
- Eckert, P., 2008. Variation and the Indexical Field. *Journal of Sociolinguistics*, 12(4), pp. 453–476.
- Eckert, P. & McConnell-Ginet, S., 1992. Think practically and look locally: Language and gender as community-based practice. *Annual review of anthropology*, 21, pp. 461–490.
- Foulkes, P. & Docherty, G., 2006. The social life of phonetics and phonology. *Journal of Phonetics*, 34(4), pp. 409–438.

References

- Giles, H., Taylor, D.M. & Bourhis, R., 1973. Towards a theory of interpersonal accommodation through language: some Canadian data. *Language in Society*, 2(2), pp. 177–192.
- Gittell, R. & Vidal, A., 1998. *Community Organizing: Building Social Capital as a Development Strategy*, Thousand Oaks, CA: SAGE Publications.
- Gladwell, M., 2010. Why the revolution will not be tweeted. *The New Yorker*. Available at: http://www.newyorker.com/reporting/2010/10/04/101004fa_fact_gladwell?currentPage=all [Accessed July 17, 2014].
- Gordon, M.J., 2008. New York, Philadelphia and other northern cities: phonology. In E. W. Schneider, ed. *Varieties of English: The Americas and the Caribbean*. Berlin and New York: Mouton de Gruyter, pp. 67–86.
- Granovetter, M., 1973. The Strength of Weak Ties. *American Journal of Sociology*, 78(6), pp. 1360–1380.
- Granovetter, M., 1983. The Strength of Weak Ties: A Network Theory Revisited. *Sociological Theory*, 1(1), pp. 201–233.
- Greiffenstern, S., 2010. *The Influence of Computers, the Internet and Computer-Mediated Communication on Everyday English*, Berlin: Logos Verlag Berlin GmbH.
- Gries, S.Th., 2004. Shouldn't it be breakfunch? A quantitative analysis of blend structure in English. *Linguistics*, 42(3), pp. 639–667.
- Gumperz, J.J., 1968. The Speech Community. In *International Encyclopedia of the Social Sciences*. London: Macmillan, pp. 381–386.
- Guy, G.R., 2011. Variation and change. In W. Maguire & A. McMahon, eds. *Analysing Variation in English*. Cambridge: Cambridge University Press, pp. 178–198.
- Hall-Lew, L., Coppock, E. & Starr, R.L., 2010. Indexing Political Persuasion: Variation in the Iraq Vowels. *American Speech*, 85(1), pp. 91–102.
- Hall-Lew, L. & Stephens, N., 2011. Country Talk. *Journal of English Linguistics*, 40(3), pp. 256–280.
- Haughey, M., 1999. MetaFilter | Community Weblog. Available at www.MetaFilter.com
- Hay, J. & Drager, K., 2007. Sociophonetics. *Annual Review of Anthropology*, 36(1), pp. 89–103.
- Herring, S.C., 1993. Gender and democracy in computer-mediated communication. *Electronic Journal of Communication*, 3(2), pp. 476–489. Available at: <http://portal.acm.org/citation.cfm?id=233411>.
- Herring, S.C., 1999. Computer-Mediated Discourse D. Schiffrin, D. Tannen, & H. E. Hamilton, eds. *Discourse*, 612634(January), pp. 1–24. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.99.6319&rep=rep1&type=pdf>.
- Herring, S.C., 2004. Computer-Mediated Discourse Analysis: An Approach to Researching Online Behavior. In S. A. Barab, R. Kling, & J. H. Gray, eds. *Designing for Virtual Communities in the Service of Learning*. New York: Cambridge University Press, pp. 338–376.
- Herring, S.C., 2007. A Faceted Classification Scheme for Computer-Mediated Discourse. *Language@Internet*, 4(1), pp. 1–37. Available at: http://www.languageatinternet.de/articles/2007/761/index_html [Accessed January 31, 2011].
- Herring, S.C., 2008a. Language and the Internet. In W. Donsbach, ed. *International Encyclopedia of Communication*. Blackwell Publishers.
- Herring, S.C., 2008b. Questioning the Generational Divide: Technological Exoticism and Adult Constructions of Online Youth Identity. In D. Buckingham, ed. *Youth, Identity, and Digital Media*. The John D. and Catherine T. MacArthur Foundation Series on Digital Media and Learning. Cambridge, MA: The MIT Press, pp. 71–92.
- Herring, S.C., 2010. Web Content Analysis: Expanding the Paradigm. In J. Hunsinger, M. Allen, & L. Klastrup, eds. *The International Handbook of Internet Research*. Berlin: Springer Verlag, pp. 233–249.
- Herring, S.C. & Paolillo, J.C., 2006. Gender and genre variation in weblogs. *Journal of Sociolinguistics*, 10(4), pp. 439–459.
- Herring, S.C., Scheidt, L.A., Bonus, S. & Wright, E., 2004. Bridging the gap: a genre analysis of Weblogs. *Proceedings of the 37th Annual Hawaii International Conference on System Sciences*, pp. 1–11. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1265271>.
- Herring, S.C., Scheidt, L.A., Wright, E. & Bonus, Sabrina, 2005. Weblogs as a bridging genre. *Information Technology & People*, 18(2), pp. 142–171.
- Hickey, R., 2008. Irish English: Phonology. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 71–104.
- Holmes, J. & Meyerhoff, M., 1999. The Community of Practice: Theories and methodologies in language and gender research. *Language in Society*, 28(2), pp. 173–183.

References

- Holmes, J. & Woodhams, J., 2013. Building interaction: The role of talk in joining a community of practice. *Discourse & Communication*, 7(3), pp. 275–298.
- Horsman, S., 2006. The politics of toponyms in the Pamir mountains. *Area*, 38(3), pp. 279–291.
- Horvath, B.M., 2008. Australian English: Phonology. In K. Burridge & B. Kortmann, eds. *Varieties of English: The Pacific and Australasia*. Berlin and New York: Mouton de Gruyter, pp. 89–110.
- Hutchby, I., 2001. *Conversation and technology: from the telephone to the Internet*, Cambridge UK; Malden MA: Polity Press.
- Hymes, D., 1974. *Foundations in Sociolinguistics: An Ethnographic Approach*, University of Pennsylvania Press.
- Johnstone, B., 2007. Linking Identity and Dialect through Stancetaking. *Stancetaking in discourse: Subjectivity, evaluation, interaction*, pp. 49–68.
- Johnstone, B., 2009. Pittsburghese Shirts: Commodification and the Enregisterment of an Urban Dialect. *American Speech*, 84(2), pp. 157–175.
- Johnstone, B., 2010a. Enregistering Style. In *Rhetoric Society of America conference*. Minneapolis. Available at: http://works.bepress.com/cgi/viewcontent.cgi?article=1053&context=barbara_johnstone [Accessed January 5, 2012].
- Johnstone, B., 2010b. Making Pittsburghese: Communication technology, expertise, and the discursive construction of a regional dialect. *Language & Communication*, pp. 1–13.
- Johnstone, B., Andrus, J. & Danielson, A.E., 2006. Mobility, Indexicality, and the Enregisterment of “Pittsburghese.” *Journal of English Linguistics*, 34(2), pp. 77–104.
- Jurgenson, N., 2012. When Atoms Meet Bits: Social Media, the Mobile Web and Augmented Revolution. *Future Internet*, 4(4), pp.83–91. Available at: <http://www.mdpi.com/1999-5903/4/1/83/>
- Kahn, D., 1976. *Syllable-based generalizations in English phonology*. Indiana University Linguistics Club, Vol. 156.
- Kasof, J., 1993. Sex Bias in the Naming of Stimulus Persons. *Psychological Bulletin*, 113(1), pp. 140–163.
- Kortmann, B. & Upton, C., 2008. Introduction. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 23–34.
- Kretschmar, W.A., 2008. Standard American English pronunciation. In E. W. Schneider, ed. *Varieties of English: The Americas and the Caribbean*. Berlin and New York: Mouton de Gruyter, pp. 37–51.
- Labov, W., 1963. The Social Motivation of a Sound Change. *Word*, (19), pp. 273–309.
- Labov, W., 1972. *Sociolinguistic Patterns*, Philadelphia, PA: University of Pennsylvania Press.
- Labov, W., 2001. *Principals of Linguistic Change Volume 2: Social factors* P. Trudgill, ed., Oxford, UK: Wiley-Blackwell.
- Ladefoged, P., 2006. *A Course in Phonetics*, Boston, MA: Thomson, Wadsworth.
- Lave, J. & Wenger, E., 1991. *Situated Learning: Legitimate Peripheral Participation*, Cambridge University Press.
- Lawson, R., 2011. Patterns of linguistic variation among Glaswegian adolescent males. *Journal of Sociolinguistics*, 15(2), pp. 226–255.
- Lawton, P., 2005. *Capital and stratification within virtual community: a case study of MetaFilter.com*. Masters Thesis. Lethbridge, Alta.: University of Lethbridge, Faculty of Arts and Science.
- LeBlanc, T.R., 2010. *1337 \/\4YZ of \$p34King: Toward a unifying pragmatic theory of virtual speech community building*. Diss. Baton Rouge, La.: Louisiana State University.
- Levon, E., 2013. Ethnographic Fieldwork. In C. Mallinson, B. Childs, & G. Van Herk, eds. *Data Collection in Sociolinguistics: Methods and Applications*. New York, NY, USA: Routledge.
- Lin, N., 1999. Building a network theory of social capital. *Connections*, 22(1), pp. 28–51.
- Lipski, J.M., 1976. Prejudice and Pronunciation. *American Speech*, 51(1/2), pp. 109–118.
- Llamas, C., 2007. “A place between places”: Language and identities in a border town. *Language in Society*, 36, pp. 579–604.
- Llamas, C., 2010. Convergence and Divergence Across a National Border. In C. Llamas & D. J. L. Watt, eds. *Language and Identities*. Edinburgh: Edinburgh University Press, pp. 227–246.
- Mendoza-Denton, N., 2008. *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*, Wiley-Blackwell.
- Millard, J., 2011. MetaFilter Corpus – Frequency Tables. Available at: <http://stuff.metafilter.com/corpus/freq/> [Accessed March 29, 2011].
- Miller, H. & Arnold, J., 2001. Self in web home pages: Gender, identity and power in cyberspace. *Towards cyberpsychology: Mind, cognition and society in the Internet age*, pp. 73–94.

References

- Milroy, J. & Milroy, L., 1985. Linguistic change, social network and speaker innovation. *Linguistics*, 21(2), pp. 339–384.
- Milroy, L., 1980. *Language and Social Networks*, Oxford: Blackwell Publishing Ltd.
- Milroy, L., 2002. *Social Networks*, Oxford: Blackwell Publishing Ltd.
- Milroy, L. & Milroy, J., 1992. Social network and social class: Toward an integrated sociolinguistic model. *Language in Society*, 21(1), pp. 1–26.
- Moore, E., 2006. “You tell all the stories”: Using narrative to explore hierarchy within a Community of Practice. *Journal of Sociolinguistics*, 10(5), pp. 611–640.
- Munson, B., 2010. Levels of Phonological Abstraction and Knowledge of Socially Motivated Speech-Sound Variation: A Review, a Proposal, and a Commentary on the Papers by Clopper, Pierrehumbert, and Tamati, Drager, Foulkes, Mack, and Smith, Hall, and Munson. *Laboratory Phonology*, 1(1), pp.157–177.
- OED, 2014. Oxford English Dictionary [OED] Online. “HiFi”, “WiFi.”
- Pullum, G., 2004. Snowclones: lexicographical dating to the second. *Language Log*. Available at: <http://itre.cis.upenn.edu/~myl/language-log/archives/000350.html> [Accessed July 29, 2014].
- Putnam, R.D., 2000. *Bowling alone: the collapse and revival of American community*, New York: Simon & Schuster.
- Read, A.W., 1933. Pronunciation of the Word ‘Missouri’. *American Speech*, 8(4), pp. 22–36.
- Reid, T.B.W., 1956. Linguistics, Structuralism and Philology. *Archivum Linguisticum*, 8(1), pp. 28–37.
- Rheingold, H., 1993. *The Virtual Community*, Available at: <http://www.rheingold.com/vc/book/>.
- Rogers, E.M. & Shoemaker, F.F., 1971. *Communication of Innovations* 2nd ed., New York: Free Press.
- Russ, B., 2012. Examining Large-Scale Regional Variation Through Online Geotagged Corpora. In *ADS Annual Meeting*. Portland, OR.
- Scargill, M.H., 1957. Sources of Canadian English. *Journal of English and Germanic Philology*, (56), pp. 610–614.
- Schneider, E.W., 2008. Introduction: varieties of English in the Americas and the Caribbean. In E. W. Schneider, ed. *Varieties of English: The Americas and the Caribbean*. Berlin and New York: Mouton de Gruyter, pp. 23–33.
- Selkirk, E.O., 1981. English compounding and the theory of word structure. In M. Moortgat, H. v.d. Hulst, & T. Hoekstra, eds. *The scope of lexical rules*. Dordrecht: Foris, pp. 229–277.
- Sessions, L.F., 2010. How Offline Gatherings Affect Online Communities. *Information, Communication & Society*, 13(3), pp. 375–395.
- Shapiro, M., 1997. Broad and Flat A in Marked Words. *American Speech*, 72(4), pp. 437–439.
- Silva, D.J., Peters, S.A., Duhaish, F.B., Kim, S., Koo, Y., Marji, L. & Park, J., 2011. Variation in the IRAQ vowels outside the public forum: the indexing of political persuasion reconsidered. *American Speech*, 86(2), pp. 179–191.
- Silva, L., Goel, L. & Mousavidin, E., 2008. Exploring the dynamics of blog communities: the case of MetaFilter. *Information Systems Journal*, 19(1), pp. 55–81.
- Silverstein, M., 2003. Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23(3-4), pp. 193–229.
- Sproat, R., Black, A.W., Chen, S., Kumar, S., Ostendorf, M., Richards, C., 2001. Normalization of non-standard words. *Computer Speech & Language*, 15(3), pp.287–333.
- Squires, L., 2010. Enregistering internet language. *Language in Society*, 39(04), pp. 457–492.
- Stanford, J., 2013. How to Uncover Social Variables: A Focus on Clans. In C. Mallinson, B. Childs, & G. Van Herk, eds. *Data Collection in Sociolinguistics: Methods and Applications*. New York, NY, USA: Routledge.
- Stommel, W., 2008. Conversation Analysis and Community of Practice as Approaches to Studying Online Community. *Language@ Internet*, 5(5), pp. 1–22. Available at: <http://www.languageatinternet.org/articles/2008/1537> [Accessed February 20, 2011].
- Stuart-Smith, J., 2008. Scottish English: Phonology. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin, New York: Mouton de Gruyter, pp. 48–70.
- Sullivan, D., 2014. On MetaFilter Being Penalized By Google: An Explainer. *SearchEngineLand.com*. Available at: <http://searchengineland.com/metafilter-penalized-google-192168> [Accessed May 22, 2014].
- Tagliamonte, S.A. & Denis, D., 2008. Linguistic Ruin? LOL! Instant Messaging And Teen Language. *American Speech*, 83, pp. 3–34.
- Thomas, E.R., 2008. Rural Southern white accents. In E. W. Schneider, ed. *Varieties of English: The Americas and the Caribbean*. Berlin and New York: Mouton de Gruyter, pp. 87–114.
- Thompson, C., 2008. I’m so totally, digitally close to you. *New York Times Online Magazine*. Available at: <http://www.nytimes.com/2008/09/07/magazine/07awareness-t.html> [Accessed January 15, 2012].

References

- Thurlow, C., 2006. From Statistical Panic to Moral Panic: The Metadiscursive Construction and Popular Exaggeration of New Media Language in the Print Media. *Journal of Computer-Mediated Communication*, 11(3), pp. 667–701.
- Thurlow, C. & Brown, A., 2003. Generation Txt? The sociolinguistics of young people's text-messaging. *Discourse analysis online*, 1(1), pp.1–28.
- Trudgill, P. & Hannah, J., 2008. *International English: A guide to the varieties of Standard English* 5th ed., London: Hodder Education.
- Tucker, G.R., 2011. Re-Naming Texas: Competing Mexican and Anglo Placenames in Texas, 1821–1836. *Names: A Journal of Onomastics*, 59(3), pp. 139–151.
- Upton, C., 2008. Synopsis: phonological variation in the British Isles. In B. Kortmann & C. Upton, eds. *Varieties of English: The British Isles*. Berlin and New York: Mouton de Gruyter, pp. 269–282.
- Warnick, Q., 2010. *What We Talk about when We Talk about Talking: Ethos at Work in an Online Community*. Diss. Ames, Iowa: Iowa State University.
- Watt, D. & Ingham, C., 2000. Durational Evidence of the Scottish Vowel Length Rule in Berwick English. *Leeds Working Papers in Linguistics*, 8, pp. 205–228.
- Weinreich, U., Labov, W. & Herzog, M.I., 1968. Empirical Foundations for a Theory of Language Change. In *Directions for historical linguistics*. pp. 95–188.
- Wells, J.C., 1982. *Accents of English*, Cambridge University Press.
- Wells, J.C., 1990. Syllabification and Allophony. In S. Ramsaran, ed. *Studies in the Pronunciation of English, a Commemorative Volume in Honour of AC Gimson*. London, New York: Routledge, pp. 76–86.
- Wenger, E., 1998. *Communities of practice: learning, meaning, and identity*, Cambridge, UK; New York, NY: Cambridge University Press.
- Wenger, E., 2011. Slide: Levels of Participation. Available at: <http://wenger-trayner.com/resources/slide-forms-of-participation/>.
- Wilson, S.M. & Peterson, L.C., 2002. The Anthropology of Online Communities. *Annual Review of Anthropology*, 31(1), pp. 449–467.
- Witten, K., 2012. Sociophonetic Variation in an Internet Place Name. *Names: A Journal of Onomastics*, 60(4), pp. 220–230.