

The sociolinguistic construction of character diversity  
in fictional television series

Susan Reichelt

Centre for Language and Communication Research

School of English, Communication and Philosophy

Cardiff University

A thesis submitted for the degree of

*Doctor of Philosophy*

Cardiff University, 2018





## Declarations

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed ..... 02/01/2018

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ..... 02/01/2018

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University's Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed ..... 02/01/2018

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ..... 02/01/2018

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loans after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.

Signed ..... 02/01/2018



## Abstract

This thesis investigates the language used in six fictional television series (1997 – 2014). The overall aim is to find out how linguistic patterns contribute to distinguishing features of characters and character groups.

Throughout the thesis, I answer three overarching questions:

1. How are individual linguistic variables used for purposes of characterization?
2. How do linguistic variables interact to create linguistic character styles?
3. Are characterization patterns used in similar ways across characters within individual series, as well as across series?

The thesis presents an interdisciplinary study of sociolinguistically meaningful stylization and produces a useful account of the underused fictolinguistic approach that links concepts of variationist sociolinguistics with stylistics.

Through quantitative analysis and informed by previous sociolinguistic findings on the uses of five pragmatic forms (pragmatic markers, hedges, general extenders, modal adverbs, and intensifiers), I trace how language variation and change ties in with the individualization of fictional characters.

Findings suggest that linguistic patterns that link to character qualities are consistent across a variety of investigated features. Further, some features (e.g. pragmatic markers) appear to be used with greater variance than others (e.g. general extenders), suggesting that there are distinctions in terms of saliency and availability of characterization cues.

Further findings show linguistic variation correlating to particular character types, series production and genre, and character background (in particular nationality). Linguistic change is investigated through apparent time analyses for all features, as well as a brief real time analysis for selected contexts. Throughout the thesis, I touch upon concepts of indexicality, saliency, and authenticity.

Finally, the thesis concludes that the present study of fictional television dialogue adds to our understanding of current language use and linguistic perception and that more studies of this kind might further enhance our knowledge of the intrinsic relation between language and identity.



## Acknowledgements

It is at this point, the very final step before handing the thesis over, that I wish I could get a glance into the minds of the many great writers whose work I examine in these pages. Being able to put into words how grateful I am for the support and encouragement I received over the past four years is no easy task.

First and foremost, I would like to thank my supervisor Dr Mercedes Durham. In another place, for another degree, she provided me with lectures and plenty of reading, the right encouragement and an infectious joy for research that ultimately led me here. A PhD can be quite scary, from the actual research to teaching, presenting, reading reviewer comments, and of course, the overwhelming blank page. Thank you for the challenges in all of this, for the time you gave me, the excitement, the feedback, the chocolates, the many conversations that put me at ease, in short, for being my Giles. I could not have asked for more.

I also thank Dr Tereza Spilioti and Dr Nick Wilson for their helpful comments and encouragement at various stages of this project. Cardiff has been the home of most of this work and it is the people I met in Cardiff that contributed greatly to its shape. Thank you to everyone who has shared the wonders and woes of research: Dorottya, Piotr, Jessi, Eimi, Jaspal, Argyro, Emily, Rowan, and Gareth. I miss the lunch breaks, the reading groups, the board games. Thank you, Emanuele, for being the most wonderful flatmate, for providing me with food and much-needed escapes, together with the *Barefoot Contessa*.

Throughout the PhD I was lucky to present at a great number of conferences and learn from inspiring scholars. I thank Prof Dr Jannis Androutsopoulos for helpful comments and the platform to present my work in Hamburg, a city I miss dearly. Thank you also to Dr Monika Bednarek for the lovely talks over lunch and for (unknowingly) providing me with the final push in pursuing this PhD. Her book on *The Language of Fictional Television* is as dog-eared as my copy of *The Prisoner of Azkaban* (a whole lot).

I am grateful to Dr Vaclav Brezina and Dr Dana Gablasova for welcoming me to Lancaster, trusting me with an exciting new project and allowing me the time to finish this thesis. Thank you also to the generous support of the Sidney Perry Foundation, the Allan & Nesta Ferguson Charitable Education Trust, the Sir Richard Stapley Trust, as well as Cardiff's School for English, Communication and Philosophy for enabling me to focus on my studies full-time.

The past four years have gone by very fast and very slow in equal measures. They've been filled with amazing experiences and daunting challenges, as well as a few curveballs. I want to thank Jenny for her constant support throughout the good and the not-so-good. I couldn't have done any of this without her. Thanks also to Marika, the other HB, for sharing the experience, the pizza, the fate of empty fridges and full bookshelves.

Last but not least, thank you to my family for their support. I promise I didn't just watch TV these past few years.

# Table of Contents

Declarations.....	i
Abstract.....	iii
Acknowledgements.....	v
Table of Contents.....	vii
List of Tables.....	x
List of Figures.....	xi
1 Introduction.....	1
2 Literature Review – linguistic approaches to variation in fiction.....	9
2.1 Sociolinguistic Theory – relevant background.....	9
2.1.1 <i>Age as a social variable</i> .....	11
2.1.2 <i>Gender as a social variable</i> .....	14
2.1.3 <i>Social Class as a social variable</i> .....	15
2.1.4 <i>Social networks and communities of practice as social variables</i> .....	16
2.2 Fiction and character as landscape and agent of style.....	19
2.3 The language of television characters and previous linguistic applications ...	27
2.3.1 <i>Television dialogue</i> .....	27
2.3.2 <i>Previous linguistic approaches to television dialogue</i> .....	31
2.3.3 <i>Fictolinguistics and cinematic indexicality</i> .....	40
2.4 Characterization as sociolinguistic styling.....	48
2.4.1 <i>Functions of dialogue</i> .....	50
2.4.2 <i>Culpeper’s textual cues in characterization</i> .....	57
2.4.3 <i>Bednarek’s expressive character cues</i> .....	62
2.4.4 <i>Characterization through stance-taking</i> .....	64
3 The Television Dialogue Corpus.....	67
3.1 Corpus set-up.....	67
3.2 Series included in the TV corpus.....	74
3.2.1 <i>Buffy the Vampire Slayer and Angel</i> .....	75
3.2.2 <i>Parenthood</i> .....	79
3.2.3 <i>Gilmore Girls</i> .....	81
3.2.4 <i>Torchwood</i> .....	84
3.2.5 <i>Sherlock</i> .....	86
3.3 Corpus description.....	88

4	Methods.....	91
4.1	Corpus use.....	91
4.2	Feature extraction.....	92
4.3	Coding.....	95
4.4	R & Rstudio .....	99
4.4.1	<i>Distributional analyses</i> .....	100
4.4.2	<i>Statistical testing</i> .....	101
4.5	Limitations .....	101
5	Variation and Characterization analysis .....	103
5.1	Pragmatic Markers .....	104
5.1.1	<i>Feature definition</i> .....	104
5.1.2	<i>Token inclusion &amp; coding</i> .....	107
5.1.3	<i>General distributions</i> .....	110
5.1.4	<i>Distributions across categories</i> .....	112
5.1.5	<i>Pragmatic markers as characterization devices</i> .....	117
5.2	Hedges.....	120
5.2.1	<i>Feature definition</i> .....	120
5.2.2	<i>Token inclusion &amp; coding</i> .....	122
5.2.3	<i>General distributions</i> .....	123
5.2.4	<i>Distributions across categories</i> .....	125
5.2.5	<i>Hedges as characterization devices</i> .....	132
5.3	General extenders.....	135
5.3.1	<i>Feature definition</i> .....	135
5.3.2	<i>Token inclusion &amp; coding</i> .....	137
5.3.3	<i>General distributions</i> .....	137
5.3.4	<i>Distributions across categories</i> .....	140
5.3.5	<i>General extenders as characterization devices</i> .....	146
5.4	Modal adverbs.....	149
5.4.1	<i>Feature definition</i> .....	149
5.4.2	<i>Token inclusion &amp; coding</i> .....	154
5.4.3	<i>General distributions</i> .....	155
5.4.4	<i>Distributions across categories</i> .....	159
5.4.5	<i>Character-based distributions</i> .....	166
5.4.6	<i>General extenders as characterization devices</i> .....	173
5.5	Intensifiers.....	176
5.5.1	<i>Feature definition</i> .....	176
5.5.2	<i>Token inclusion &amp; coding</i> .....	183
5.5.3	<i>General distributions</i> .....	186
5.5.4	<i>Distributions across categories</i> .....	192
5.5.5	<i>Multivariate Analysis</i> .....	205
5.5.6	<i>Character-based analysis</i> .....	210
5.5.7	<i>Intensifiers as characterization devices</i> .....	213

6	Change and Characterization analysis .....	216
6.1	The discourse value as an indicator for linguistic change.....	216
6.1.1	<i>D-Value of pragmatic markers</i> .....	218
6.1.2	<i>D-Value of hedges</i> .....	231
6.1.3	<i>D-Values as indicators of change</i> .....	241
6.2	Changing characters, changing patterns.....	241
6.2.1	<i>Character change: Cordelia, from Buffy to Angel</i> .....	242
6.2.2	<i>Character types: from Lorelai to Sarah</i> .....	245
7	Discussion .....	250
7.1	Summary .....	250
7.1.1	<i>Outliers</i> .....	252
7.1.2	<i>Common patterns</i> .....	253
7.1.3	<i>Change</i> .....	253
7.2	How are linguistic variables used for the purpose of characterization?.....	254
7.3	How do linguistic variables interact to create character styles? .....	257
7.4	Are patterns congruent across characters and series? .....	259
7.5	A fictolinguistic approach: an evaluation .....	260
8	Conclusion .....	262
	List of Media Sources .....	264
	References .....	266
	Appendices.....	276
	Appendix A: Distribution tables .....	277
	<i>Pragmatic markers</i> .....	277
	<i>Hedges</i> .....	278
	<i>General extenders</i> .....	279
	<i>Modal adverbs</i> .....	280
	<i>Intensifiers</i> .....	281
	Appendix B: Frequencies, incl. raw counts .....	282
	Appendix C: Significance tests .....	288
	Appendix D: Scripts for distribution plots.....	289
	Appendix E: Distribution plots .....	290
	<i>Pragmatic markers</i> .....	290
	<i>Hedges</i> .....	295
	<i>General extenders</i> .....	298
	<i>Modal adverbs</i> .....	304
	<i>Intensifiers</i> .....	309

## List of Tables

Table 1: Cues in characterization (Bednarek (2011a) after Culpeper (2001)).....	63
Table 2: Series overview.....	68
Table 3: Corpus summary.....	89
Table 4: Pragmatic markers distribution, per 10,000 words.....	111
Table 5: Use of <i>I mean</i> (per 10,000 words) main characters in <i>Buffy</i> and <i>Gilmore Girls</i> .....	114
Table 6: Overall distributions hedges, tokens per 10,000 words.....	124
Table 7: Hedge use per 10,000 words across gender and age.....	127
Table 8: General extenders distribution, per 10,000 words.....	138
Table 9: Comparison GE use: Television Dialogue Corpus and Quaglio (2009).....	139
Table 10: Each □ represents 100 tokens per million words, table adapted from Biber et al. (1999: 869).....	155
Table 11: Quaglio's <i>Friends</i> data and Television Dialogue Corpus, modal adverbs per 10,000 words.....	157
Table 12: Modal adverb use across speaker groups, per 10,000 words.....	158
Table 13: Cross tabulation of modal adverbs: Age & Gender, means per 10,000 words .....	163
Table 14: Modal adverb by nationality, means per 10,000 words.....	164
Table 15: Most frequent intensifiers in Tagliamonte and collaborators' studies.....	179
Table 16: Most frequent intensifiers in Barnfield and Buchstaller (2010).....	180
Table 17: Complete set of intensifiers in the Television Dialogue Corpus.....	187
Table 18: Intensifier use in the Television Dialogue Corpus (per 10,000 words).....	190
Table 19: Multivariate analyses of <i>very</i> by age group.....	206
Table 20: Multivariate analyses of <i>really</i> by age group.....	207
Table 21: Multivariate analyses of <i>so</i> by age group.....	209
Table 22: D-value (%) by genre, adapted from Beeching (2016:105; 191;135).....	220
Table 23: <i>Sort of / kind of</i> distribution in TV corpus.....	231
Table 24: D-value (in %) by genre, adapted from Beeching (2016:161).....	232
Table 25: Categorization of characterization cues.....	255

## List of Figures

Figure 1: <i>Gilmore Girls</i> scene composition-----	49
Figure 2: Script extract from <i>Gilmore Girls</i> revival (Netflix, 2016) as printed in (Nussbaum, 2016) -----	70
Figure 3: Subtitles in <i>Gilmore Girls</i> ("The Deer Hunters", season 1 episode 4)-----	71
Figure 4: Sound descriptor <i>Sherlock</i> ("The Hounds of Baskerville", season 2 episode 2) -----	71
Figure 5: Sociogram: <i>Buffy the Vampire Slayer &amp; Angel</i> -----	77
Figure 6: Sociogram: <i>Parenthood</i> -----	79
Figure 7: Sociogram: <i>Gilmore Girls</i> -----	82
Figure 8: Sociogram: <i>Torchwood</i> -----	85
Figure 9: Sociogram: <i>Sherlock</i> -----	87
Figure 10: Storage of series files-----	92
Figure 11: Loaded TV Corpus in AntConc-----	93
Figure 12: Search for <i>you know</i> -----	95
Figure 13: Coding <i>kind of</i> -----	97
Figure 14: Frequency table for hedges-----	99
Figure 15: Pragmatic markers across gender, per 10,000 words-----	113
Figure 16: Pragmatic markers across age, per 10,000 words-----	115
Figure 17: Pragmatic markers across nationality, per 10,000 words-----	116
Figure 18: Hedges across gender, per 10,000 words-----	125
Figure 19: Hedges across age groups, per 10,000 words-----	127
Figure 20: Stratification of hedges across gender and age, per 10,000 words-----	128
Figure 21: Hedges across nationality, per 10,000 words-----	128
Figure 22: Hedges compared: <i>sort of</i> relative to <i>kind of</i> -----	130
Figure 23: General extender use across gender, per 10,000 words-----	141
Figure 24: General extenders across age, per 10,000 words-----	142
Figure 25: General extenders across nationality, per 10,000 words-----	143
Figure 26: General extenders across characters, per 10,000 words-----	145
Figure 27: Modal adverb use across gender, per 10,000 words-----	161
Figure 28: Modal adverb use across age, per 10,000 words-----	162

Figure 29: Modal adverb use across nationality, per 10,000 words-----	165
Figure 30: Modal adverb use cross-tabulation: nationality & age (per 10,000 words)	166
Figure 31: Modal adverb <i>maybe</i> across series and characters, per 10,000 words-----	167
Figure 32: Modal adverb <i>perhaps</i> across series and characters, per 10,000 words----	169
Figure 33: Modal adverb <i>probably</i> across series & characters, per 10,000 words-----	170
Figure 34: Modal adverbs across characters, per 10,000 words-----	172
Figure 35: Intensification terminology-----	178
Figure 36: Main boosters across gender-----	194
Figure 37: Main maximizers across gender-----	196
Figure 38: Main boosters across age-----	199
Figure 39: Main maximizers across age-----	201
Figure 40: Main boosters across nationality-----	203
Figure 41: Main maximizers across nationality-----	204
Figure 42: Intensifier variation across all characters and series-----	211
Figure 43: D-value of pragmatic markers-----	218
Figure 44: D-Value pragmatic markers across gender-----	221
Figure 45: D-Value pragmatic markers across age-----	222
Figure 46: D-Value pragmatic markers across nationality-----	224
Figure 47: D-Value across characters: <i>Parenthood</i> -----	225
Figure 48: D-Value across characters: <i>Gilmore Girls</i> -----	226
Figure 49: D-Value across characters: <i>Torchwood</i> -----	227
Figure 50: D-Value across characters: <i>Sherlock</i> -----	228
Figure 51: D-Value across characters: <i>Angel</i> -----	229
Figure 52: D-Value across characters: <i>Buffy the Vampire Slayer</i> -----	229
Figure 53: D-Value hedges across gender-----	232
Figure 54: D-Value hedges across age-----	233
Figure 55: D-Value hedges across nationality-----	234
Figure 56: D-Values of hedges across characters, <i>Parenthood</i> -----	235
Figure 57: D-Values of hedges across characters: <i>Gilmore Girls</i> -----	237
Figure 58: D-Values of hedges across characters: <i>Angel</i> -----	238
Figure 59: D-Values of hedges across characters: <i>Buffy the Vampire Slayer</i> -----	239
Figure 60: D-Values of hedges across characters: <i>Torchwood</i> -----	240

Figure 61: D-Value of hedges across characters: <i>Sherlock</i> -----	240
Figure 62: Changing patterns in Cordelia's linguistic profile-----	243
Figure 63: D-value Lorelai Gilmore & Sarah Braverman across the series' running time (in %) -----	247



# 1 Introduction

Giles (VO)

Previously on *Buffy the Vampire Slayer*...

Giles

*You are the chosen one. You alone can stop them.*

Buffy

*Who?*

Giles

*The vampires.*

Montage of five years of Buffy fighting (and defeating) demons, vampires, and monsters. By her side always: her friends Willow and Xander, Anya, Tara, and Spike, as well as watcher Giles.

Giles (VO)

*Into each generation a Slayer is born...*

(Scene cont'd.) Graveyard, Sunnydale California. It's night-time. Xander and Anya are attacked by a vampire.

Anya

*Xander!*

Anya falls to the ground with the vampire approaching. Xander tries to defend himself, but is interrupted by Buffy entering the scene.

Buffy

*I got it!*

She high-kicks the vampire who falls to the ground. He gets up and both start a close up fight with Anya and Xander looking on from the side. The vampire lifts Buffy into the air but before he can do anything Spike enters the scene and joins the fight - releasing Buffy from the vampire's grip.

Spike is punched and falls just as Buffy gets up again and manages to stake the vampire who turns to dust. Buffy proudly looks onto the spot the vampire just moments ago occupied.

Buffy

*That'll put marzipan in your pie plate, bingo!*

Everyone looks around, confused by what Buffy just said.

(From opening sequence in *Buffy the Vampire Slayer*, season 6 premiere "Bargaining, pt.1")

*That'll put marzipan in your pie plate, bingo?*

The confusion that follows this sentence is not just felt by the bystanders (Spike: “What’s with the Dadaism?”), but likewise by audiences watching this fantasy television series at home. Avid watchers of the *Buffy the Vampire Slayer*’s previous five seasons have come to get to know the characters and the way they behave, move, and speak. Buffy’s triumphant exclamation is absurd and completely out of character: it signals that something is different.

Over the past five seasons, the audience has come to expect the characters’ creative language use, with Buffy usually leading in the use of new and innovative forms (see for instance Adams, 2003), silly puns and witty references<sup>1</sup>. But this particular witticism is different. In fact, it is this sentence that gives away that this Buffy is not *the real* Buffy at all, but rather a lookalike robot that is being used to hide the fact that Buffy actually died a few months prior<sup>2</sup>. The single line of dialogue here points to this unique background information and enables the audience to separate what is supposedly real and what is not. Her dialogue is able to authenticate Buffy’s fictive realness because the audience knows what her dialogue *should* sound like.

---

<sup>1</sup> From season 4, episode 12:

Professor Walsh: We thought you were a myth.

Buffy: Well, you were myth-taken.

From season 5, episode 1:

Buffy: So - let me get this straight. You're Dracula. The guy. The count.

Dracula: I am.

Buffy: This isn't just a fanboy thing, is it? 'Cause I've fought more than a couple of pimply, overweight vamps who called themselves "Lestat".

<sup>2</sup> She is revived shortly after and continues as titular character for another two seasons.

This thesis explores the notion of character-typical language and how language is used to create and maintain recognizable individuals on screen. By focusing on a particular set of expressions (detailed in chapter 5), I will show that language use adds to a number of different characterization processes. First and foremost, I am interested in how characters are individualized, i.e. why we perceive each character as having ‘their own’ voice (and if that is indeed the case). This also relates to the question of how characters are grouped.

Previous research has found that there are strong linguistic indicators of in- and outgroups (predominantly friendship groups, e.g. Babel (2006) or Mandala (2007)) within television series. A systematic approach that examines over-arching character groupings, i.e. characters sharing similar background information, across series is however currently still missing. Another process related to characterization is the mediated dissemination of linguistic markers and stereotypes. The thesis thus investigates features that index particular speaker qualities and whether they are

- a) used as characterization devices,
- b) further perpetuate common stereotypes, or
- c) reject or even reframe indexicality markers in accordance with characterization patterns.

Over the past decade, fictional television dialogue has received heightened academic recognition, particularly within linguistics (see chapter 2). This reflects a shift not only in terms of research focus, but also in terms of recognizing fictional television series as a possible data source. In the next few paragraphs, I will contextualize the study of fictional television dialogue within academia to highlight some reasoning behind previous neglect, as well as reasons for the now steady increasing interest in popular fictional media, starting with media studies and expanding into areas of linguistics.

General academic reluctance to include fictional television is concerned with the scholarly approval of television as a research subject itself and the specific values that are

attached to different cultural artefacts. For years, popular culture was deemed as equal to the term ‘low culture’ – as standing opposed to ‘high culture’, creating a dichotomy between what is valued art and elite and what is trivial and “labelled as ‘trash’” (Schudson, 1987a:51). Adorno for instance goes as far as to say that low culture “impedes the development of autonomous, independent individuals who judge and decide consciously for themselves” (2001:89-99). The increasing popularity of television to societies in particular caused many people to not only decry it as a “parasite” (MacDonald 1957 in Tavin 2005:104) that threatens “the very existence of civilization” (Giroux & Simon 1989:6), but to warn over loss of morality and cognitive functions (Starker 1989:170). An emphasis on the inherent qualities (or lack thereof) of cultural artefacts lead to popular culture being dismissed as not aesthetic, too every-day and trivial (meaning without interest), too transparent (meaning boring), and generally impoverishing our culture (Tavin 2005). Over the past three decades, however, a shift in most academic disciplines was noticeable whereby value is not placed on the cultural artefact per se, but rather its contextualization through consumption: “the quality of *reading* (...) takes center stage” (Schudson 1987:59). Increasing attention towards popular culture in social sciences and humanities, including linguistics (ed. volumes for instance include Piazza et al., 2011b; Scannell, 1991; Ventola and Moya Guijarro, 2009)<sup>3</sup>, shows that associations of popular culture as a lesser valued part of culture are mostly unlearned.

---

<sup>3</sup> See also publications by Adams (2011; 2013); Androutopoulos (2012b); Bednarek (2010; 2011b; 2012a; 2012b); Beers Fägersten (2016); Bleichenbacher (2012); Bubel (2006); Djonov and Zhao (2014); Dose (2013); Dragojevic et al. (2016); Herbst (1994); Higgins (2012); Hodson (2014); Johnson and Milani (2010); Kozloff (2000); Lorenzo-Dus and Garcés-Conejos Blitvich (2013); Mandala (2007; 2008); Mittell (2015); Mittmann (2006); Moss (2001); Petrucci (2012); Piazza et al. (2011a; 2011b); Planchenault (2012); Queen (2004; 2015b); Rey (1996); Richardson (1999; 2010); Richardson and Queen (2012); Ruddell (2006); Thornham and Purvis (2005), as well as the extensive list by Bednarek and Zago (2017)

Media studies are, for obvious reasons, frontrunners in investigations of popular culture, including fictional film and television. However, as Kozloff (2000) illustrates, an increased amount of scholarly work in the field of scripted media does not necessarily also include work on its language. She critically reflects that in media studies it is the *visual* in audio-visual material that was always the preferable option to study: “Films must tell their stories visually – editing, deep focus, lighting, camera movement, and nifty special effects are what really counts. Dialogue, on the other hand, is just something we have to put up with” (2000: 4). When audio was included, she continues, it would most often investigate “sound technology, film music, and sound theory” (2000:6), disregarding dialogue yet again. This neglect is also apparent in screenwriting advice books that, when discussing film or television dialogue, do so in critical ways that aim at pointing out “bad” or unrealistic dialogue, a concept that will be discussed later on, rather than investigating what functions fictional dialogue has and what role it plays in constructing fictional settings.

Turning to linguistics, another theoretical notion stood in the way of including the language of fictional media in scholarship. As briefly mentioned in the preface of Queen’s book on language in popular media (2015: viii), the fact that “linguists tend to see the media as somehow not ‘real’” meant that for many researchers, fictional dialogue was deemed not worthy of much consideration. As illustrated in chapter 2, this approach shifted somewhat with interdisciplinary studies combining stylistics with linguistic studies. The question of “realness” of language however remains a difficult one, in particular for sociolinguistics where investigations would traditionally aim for the most natural (‘vernacular’) language. Over the past decade or so however, frameworks were established that opened linguistic studies up to new avenues. As Stamou (2014:118-119) summarizes, notions of, for instance, ‘crossing’ (Rampton 1995) or ‘stylization’

(Coupland 2001) helped the field with the idea that scripted language is indeed ‘real’ within its own context and thus very much worthy of close analysis.

The present study puts fictional dialogue of six television series at the centre of investigation and explores how the language we hear links to the characters we see on screen. The thesis attempts to tease apart why we, as audiences of television series, come to expect dialogue that is specific to each character. The notion of knowing when a character is using their own words and how they differ from other characters’ words is grounded in the establishment of patterned dialogue variation: the fact that every character exhibits what McKee calls “a singular voice” (2016:156).

The thesis investigates how characterization in fictional television series is achieved through linguistic variation and change. Embedded in theoretical frameworks of sociolinguistics and stylistics, it will analyse fictional television language and show how character-specific dialogue supports social categorizations of age and gender, regional backgrounds, and social class and how these categories support the establishment of fictional characters in the first place. Its aim is to analyse how certain features of language are used to construct and maintain characters, as well as indicate development over time. Alongside these themes, I attempt to answer the following overarching research questions:

- (1) How are individual linguistic variables used for purposes of characterization?
- (2) How do linguistic variables interact to create linguistic character styles?
- (3) Are characterization patterns used in similar ways across characters within individual series, as well as across series?

To answer these questions, the thesis is divided into two parts: theoretical disposition and practical application. In chapter 2, I provide an overview of relevant sociolinguistic

studies, frameworks and how sociolinguistic notions of style interlock with theories found in stylistics. Specifically, I explore the conceptualization of fictionality, character, and characterization. The framework of fictolinguistics ties approaches to naturally occurring language and sociolinguistics together with those detailing styling of the fictional authentic. Finally, I present the basis for my methodology: a focus on linguistic features of stance as a means of characterization.

In the second part, starting with chapter 3, I introduce the corpus of sourced dialogue from six television series, spanning two decades and close to forty different characters that is used for the quantitative sociolinguistic analyses. By investigating the character-specific use of discourse features that are related to the notion of stance (e.g. hedges, pragmatic markers, intensifiers), I am able to point to variation patterns that are linked to concepts of indexicality, generational change, and register. The analyses (chapters 5 and 6) present the findings for each of the features and a first interpretation of how characterization processes are, at least partly, guided by linguistic choices. Finally, the discussion chapter (chapter 7) illustrates what the findings mean with reference to the immediate context of the corpus and how the research questions posed here can be answered. I then return to a more general application of my findings in my conclusion (chapter 8), where I frame the outcome of this study within linguistic research, and sociolinguistics in particular.

Going back to the beginning of this chapter, I highlighted the difficult journey popular culture had in being recognized within academia. The past decades have seen a shift in academic attitudes and more theoretical acceptance of scripted fictional language within linguistics in particular. The thesis thus continues this work and further establishes fictional television language as an important cultural artefact that deserves academic attention as much as other language data.

Kozloff, in her conclusion in *Overhearing Film Dialogue*, writes about “prejudices against film dialogue (...) lingering like the undead” (2000:269) and likens the eventual acceptance in academia to “vanquishing a vampire”. Following from the opening scene of *Buffy the Vampire Slayer* at the beginning of this chapter then, this thesis explores characters and their unmistakable language and, just as the robot-type Buffy in that scene, it will seek to highlight the important role language plays in identity construction, real or not.

## **2 Literature Review – linguistic approaches to variation in fiction**

### **2.1 Sociolinguistic Theory – relevant background**

The study is predominantly grounded in sociolinguistics: language, society, and the interrelationship of both. The main departure from traditional sociolinguistic theory is the contextualization of society as fictional through the framework of ‘fictolinguistics’. As a starting point, I will highlight some of the relevant studies that are part of sociolinguistic research. This selective summary teases apart the applicability of sociolinguistic theory to non-traditional sociolinguistic data, that is, data that is not spontaneous and naturally occurring but planned and scripted. With that, I return to the issue of ‘real’ language, authenticity, and how linguistics approaches performed media language and the notion of style in building identities.

Previous literature presented here is kept purposefully broad in scope. More detailed and focused reference studies, specifically on the linguistic features that are part of the analysis, will be presented in chapters 5 and 6.

Sociolinguistics, broadly, observes linguistic variation and its correlation with social circumstances, how it changes over time, and how it influences the way we identify ourselves and others with and through language. Language variation and language change are anchor points within sociolinguistics, following Labov’s initial claim (1963)<sup>4</sup> of

---

<sup>4</sup> Labov, in Martha’s Vineyard, found that variation of certain sound structures correlated with speaker backgrounds: Fishermen that identified closely with the island would use features that were associated with the island, making their language more conservative and traditional. This also made them stand apart from visitors to the island.

variation in speech not as random or by chance, but systematic and patterned; this is also referred to as “structured heterogeneity” (Labov, 1972a; 2001; Weinreich et al., 1968).

The beginnings of the field aimed at exploring the correlation of language patterns with macrosocial categories, such as age, gender, or social class. Chambers, in discussing these categories, said that we, as speakers, “embody in our speech, as in our dress, manners, and material possessions, the hallmarks of our social background” (1995:7). The same of course can be applied to fictional speakers, if on a much more conscious level. The social background of the character, essentially who they are, is thus reflected not only in the visual, something media studies have investigated for some time now, but also in the auditory – the dialogues, monologues, voice overs, and off-screens, as Richardson (2010:3) describes below:

Television consumers are characteristically referred to as its *viewers* and described as *watching* programs on TV. But the experience of television is seriously incomplete unless viewers are also *listeners*, who engage with the various mixes of sound, speech and music that the medium has to offer, in combination with its visual images.

The “various mixes” that Richardson mentions above describe language variation quite aptly. Speakers, fictional or real, exhibit a specific profile of language depending on their social background, who they are speaking to, and what they are saying. Sociolinguistics investigates the linguistic variable (the piece of language that can vary within the speech of one person as well as across speakers) and what factors may influence its use. The different choices speakers have in using a variable are called variants.

A well-researched linguistic variable is, for instance, the quotative (D’Arcy (2007); Rickford, Wasow, et al. (2007); or Buchstaller and D’Arcy (2009)). The following

extracts exemplify different ways in which characters on television may introduce direct speech or thought:

- (1) Lorelai: But, still, when it snows, something inside me *says*, 'Hey, that's your present'. (*GG*)
- (2) Lorelai: But then I asked myself, 'W.W.T.B.F.C.D.'? And it came to me in a flash, Ø 'I'm gonna make waffles'. (*GG*)
- (3) Buffy: So then Kathy's *like*, 'It's share time'. And I'm *like*, 'Oh yeah'? (*BVS*)

Variants for reporting direct speech include, among others, *say*, *go*, *all*, *like*, and Ø. Which variant is chosen by the speaker depends on a number of factors, including internal (linguistic) contexts, as well as social contexts (such as age and gender, but also type of conversation).

In the following I want to illustrate the notion of social categories and how, according to sociolinguistic research, they may correlate with language. I will do this by briefly outlining the findings related to some of the key social factors studied in sociolinguistics.

### **2.1.1 Age as a social variable**

A speaker's age is a social factor that can be used to investigate a number of characteristics of language use, in particular how language use is changing over time. According to Chambers (1995:147), observing language change as it happens is "the most striking single accomplishment of contemporary linguistics". Speaker age can be used as a predictor for generational language change, essentially allowing the researcher to track change as it happens rather than after the fact, which, for a long time, was assumed the only manner of diachronic study. Linguistic change can be observed in apparent time studies (Bailey, 2008) that investigate language use at one point in time, suggesting that

older speakers reflect language use of the past and younger speakers exhibit new and innovative linguistic forms. Real time studies, in contrast, investigate language at two different points in time to find whether change has occurred (cf. Bailey, 2008) (Murphy, 2010b:5-8). Linguistic change, it is worth noting, is not limited to community-wide shifts that can be pinned down through speakers of different generations. Age grading, for instance, is the idea that speakers themselves will change their use of language within their lifetime. This might correlate with social circumstances (e.g. speakers change their style as they move from school to professional contexts) and rather than indicating a diachronic change within the linguistic system, speakers exhibit a shift that repeats itself throughout generations. This somewhat complicates interpretations drawn from apparent and real time studies, as the individual linguistic change that is due to reasons outside of generational shifts is usually difficult to detect (cf. Wagner, 2012 for a detailed discussion).

Additionally, there are other factors that need to be considered when talking about age as a social variable. Eckert (1997:154-155) says, on the notion of chronological age, that “as social and biological development do not move in lockstep with chronological age, it can only provide an approximate measure of speakers’ age-related place in society”. Instead of a possibly decontextualized chronological age, some studies thus turned to age measures through life events and experiences, detailed further in Murphy (2010b:3-4). The measure that is of importance here is that of ‘social age’ which describes age as represented through life events within society. This follows an emic approach, which “groups speakers [...] according to some shared experience of time which can be related to life stage or history” (Murphy, 2010b:3-4), an approach previously suggested by Eckert (1984). Life stages follow societal norms and expectations, which have an effect on a person’s behaviour, including their language. Differences are commonly drawn between

childhood, adolescence, middle, and old age<sup>5</sup>, which correlate with life events: first day at school, individualization and puberty, graduation, first job, retirement, etc. It is assumed that these life events mark a possible change in how a person identifies and, consequently, how they use language to portray themselves.

A considerable amount of research has been conducted for individual age groups and their linguistic particularities. Childhood language research in variationist sociolinguistics is predominantly concerned with acquisition of language and variation itself (Labov, 2009; Smith et al., 2007), as well as sociolinguistic competencies (Bayley and Regan, 2004). Following that, the language use of teenagers and young adults is viewed as a period of exhibited language change, individualization and innovation (Bauer and Bauer, 2002; Eckert, 2000; Stenström, 1999; Stenström et al., 2002; Tagliamonte, 2005). According to Coupland, adulthood is “an unmarked demographic condition” (Coupland, 2004:69) and Bailey (2008:324) concludes that “vernaculars generally remain stable during the adult years”, albeit how these adult years can be defined exactly remains unclear.

Age as a social variable, as indicated above, is a versatile factor when it comes to observing language development over time (be that the life time of an individual, or time across generations). As a possible factor in the present study, the saliency of perceived variation between generations will be of particular interest. Three main age groups at strikingly different points in their lives (high school students, adults who have entered the work force, and adults who are close to or at retiring age) will account for variation according to different life stages.

---

<sup>5</sup> Depending on the study these can be more fine-grained.

### 2.1.2 *Gender as a social variable*

Gender, as a social construct (but usually based on the physical binary sex), is one of the most well-researched social categories in (socio-)linguistics with work focusing mainly on gender as a social variable (Cameron, 1996; Cheshire, 2002; Coates, 2015; Eckert, 2014; Eckert and McConnell-Ginet, 1999; Ehrlich et al., 2017; Murphy, 2010a) and many more including gender as one of various variables in their investigation of linguistic variation. (A simple search for sociolinguistic journal articles exploring gender as a possible relevant factor in language variation from 2016 to mid-2017 offers well over 1000 results.)

Early sociolinguistic studies found that women, in various settings, show a greater use of overtly prestigious and/or standard forms (Wolfram 1969; Trudgill 1974; Macaulay 1977). Further to that, Labov (1990:205-215) formulated principles of how gender interacts with language variation, whereby women not only preferred standard forms, but were often found to be more innovative in their language. Possible explanations sought for these patterns usually included some contextualization within social and cultural settings. Trudgill (1972), for instance, claims that women face more pressure to conform to set social norms and are more upwardly mobile, thus they orient themselves more so than men to standard language forms. Eckert (2012:90) writes that “gender dynamics were seen as resulting from the effects of these categories [socioeconomic status] on speakers’ orientation to their assigned place in that hierarchy”. Over time however, other studies revealed that Labov’s initial principles did not apply across-the-board and that variation cannot always be accounted and reasoned for by limiting the influencing factors to broad social categories. I will elaborate on this particular point further down.

Popular views and anecdotal commentary on gender differences in language use are concerned with, for example, an apparent overuse of vague language by women (cf. Lakoff 1975) associated with their weaker positioning in society. However, as other studies have pointed out, there is no consistent evidence given for this distinctive use in vernacular speech (thoroughly investigated in, for instance, Holmes, 2008). For the present study it is useful to regard earlier studies on dichotomous language use as possible indicators for stereotypical stylization in television contexts. Gender, as a salient social marker, is likely to be used for characterization and I suggest that certain indexes are used for character type recognition despite not being a true representation of naturally occurring language. Possibly one of the most global phenomena here is that of the ‘Valley Girl’, or ‘California white girl speech’ (Eckert, 2002:3), linguistically marked through increased use of discourse markers and ‘uptalk’. This particular stereotype, also investigated by Podesva (2011) and Bucholtz et al. (2007) will be apparent in the current study as well. A particular focus of the analysis is to see whether language variation across characters can be attributed to gender and/or specific gender stereotypes, and in how far this characterization is consistent (scripted contexts follow similar patterns of gendered dialogue throughout) and stable (if stereotypes are implied, are they maintained throughout?).

### **2.1.3 *Social Class as a social variable***

Social class, socioeconomic status, and educational background can all be said to contribute to the notion of social class as a possible linguistic constraint. With the inaugural sociolinguistic studies (e.g. Labov, 1966; Trudgill 1974) putting social class front and centre of likely factors influencing linguistic variation, the effect of social class was often the assumed starting point of investigation, as indicated in Eckert’s quote (2012:90) above.

General findings propose that speakers of a higher social class would use more standard or prestigious language forms while speakers in a lower social class would use more non-standard or local forms. Associations between lower social class and salient non-standard linguistic forms led to stereotypical representations of lesser educated people speaking in stigmatized ways and non-standard forms being reflective of ‘improper’ speakers (cf. Lippi-Green 1997). Similarly to age as a social factor, social class is a possible constraint that needs careful definition before any findings are to be interpreted. With changing social systems across cultures as well as time for instance, it becomes clear that social class is not universally or indefinitely applicable. Further, linguists have not yet agreed on how to appropriately assign social class. Some studies include information about the educational background of the individual, the economic standing of parents, the socioeconomic context of area of residence or any one combination of the aforementioned with the social status of the speaker.

For this study, determining a comparable social class category was deemed impossible.<sup>6</sup> Nevertheless, characters that are introduced with overt acknowledgements regarding their socioeconomic position in society are possibly also characterized linguistically as distinguished in that regard. The analysis, while not focusing on social class per se, investigates in what ways characters are distinct from each other and I suggest that any correlation with social class can be detected as such.

#### **2.1.4 *Social networks and communities of practice as social variables***

As mentioned above, early studies in language variation and change would often confine social constraints into seemingly static categories. This ultimately suggests that the

---

<sup>6</sup> Partly because information was not available for each character and partly because fictional contexts are not comparable.

speaker is more or less restricted within set groups and varying styles are “conceived purely as the output of varying attention to speech” (Eckert, 2012:89). This approach was somewhat altered with Milroy’s study of network structure in different suburbs of Belfast (1980). She found that it is not social categories alone that have an influence on language variation, but how people organize themselves within social networks is equally important to consider. Language is seen as a meaningful attribute of identity and linguistic variation can be seen as correlating to how speakers identify themselves with reference to their belonging. Not only can linguistic likeness indicate group belonging, it also shows clear distinctions between two or more groups (Bucholtz, 1999; Eckert, 2000). On a playground in Reading, Cheshire (1982) found that the use of non-standard language corresponds to attitudes different peer groups have and similar work is continued by Moore (2003; 2004; 2006; 2010) in a school in Bolton. Exploring social networks and communities of practice (CoP, after Eckert and McConnell-Ginet (1992)) as possible factors of language variation is less about the individual speaker and more focused on the interactional aspect of language use and how we communicate in order to form rapport. It is important to note that none of these categories occur in isolation. Rather, a speaker’s language is very much influenced by all of these factors at various points in their linguistic identity construction. Eckert (2012:93) writes that “linguistic variables do not index categories, but characteristics”, meaning that social meaning in language variation might be informed by the above summarized social factors, but that speakers are not static and set within clearly demarcated groups: speakers are negotiating their identity by varying in their linguistic choices. Elsewhere, she argues that “different ways of saying things are intended to signal different ways of being, which includes different potential things to say” (2018:159). Thus, a variationist approach is ideally suited in investigating the linguistic construction of identities. The following paragraphs highlight the new focus on style in sociolinguistics and how it links into realms of stylistics.

Eckert summarizes previous sociolinguistic theory (2012) as the three waves of variation study. These go from an initial conceptualization of rather static social categories and speakers as “stable carriers of dialect” (Eckert, 2012:97) to “stylistic agents, tailoring linguistic styles in ongoing and lifelong projects of self-construction and differentiation” (2012:97-98). For the present study, all of the three waves are highly relevant when it comes to interpreting language variation and change with regards to characterization.

It would seem self-evident to focus on the third wave in particular, which finds “variation as a reflection of social identities and categories to the linguistic practice in which speakers place themselves in the social landscape through stylistic practice” (Eckert, 2012:94). Speakers in the context of fictional television series are characters that place themselves within the social landscape of a fictional world. It is the task in this study to find the social identities and categories that inform the characters and their stylistic practice in that setting. In that, the study fully commits to the third wave and its claim of variation as an essential effort of identification. The caveat here however is the application of agency. In fictional television series, it is through the cooperation of multiple people that a character (who, in its non-actual being, has no agency to speak of) is created. The construction of that character is very much aligned with the first and second wave sociolinguistic theory, as it reinforces categorical assignment that is more or less static. As will be explored further in following sections, characterization builds upon recognizable patterns that the audience can use in order to get to know a fictional being. These recognizable patterns are likely informed by stereotypes and what the writers perceive as linguistic forms that are appropriate and authentically styled after social categories of their characters (e.g. what a social and outgoing woman supposedly sounds like, what a non-conforming vampire might sound like). It is this point where the

applicability of scripted dialogue to sociolinguistic theory needs to be evaluated and negotiated and under which the proposed research questions will be addressed.

One of the possibly conflicting issues raised is the agency of linguistic choice and how we approach the elusive authentic speaker. As embedded in sociolinguistic study, I am reiterating a point made by Eckert (2018: 161): “The big question for the study of meaning in variation is how linguistic styles are constructed: what kinds of meanings can variables have, and how do they combine to yield the larger meanings of styles?”. Encompassed here are the three waves of sociolinguistic studies and how the field as such has evolved around its interest in socially meaningful variation of language. As is tradition within sociolinguistics, any and all research questions can only be answered by turning to the agents of variation themselves. As a metaphorical access to the speech community then, the next section focuses on fictionality, what and who we understand characters to be, and how frameworks from stylistics can help disentangle the notion of authenticity.

## **2.2 Fiction and character as landscape and agent of style**

Searle (1975:324) positions fiction as a form of pretence, which he defines as “a performance which is *as if* one were doing or being the thing and without any intent to deceive”<sup>7</sup>. He further claims that a performed play (which by definition is closest resembling scripted television series) is a “pretended state of affairs” (1975:328). The linguistic state of affairs, the inspiration for the language found in fiction so to speak, I argue, is a multi-layered collection of naturally occurring language, linguistic stereotypes, and more or less consciously available perceptions of language patterns. In that, fictional

---

<sup>7</sup> Searle defines his take on performance through a non-deceiving pretense of illocutionary acts. He states that while fictional propositions of illocutionary acts are non-existent, the context of agreed-upon pretense between writer and reader (or consumer more broadly)

dialogue is inevitably tied to a sense of non-fictional. Searle writes that “fictional genres are defined by the nonfictional commitments involved in the work of fiction” (1975:331). These nonfictional commitments are described as “general facts about what is possible for people to do and what the world is like” (1975:331). The operational key word here is ‘possible’ as opposed to what we know as ‘real’ or ‘actual’. The notion of possible worlds, initially proposed by Leibniz (1710/ 2000) and framed within the scholarly work of philosophy and logic, describes the coexistence of our actual world (that is tied to indisputable true or false propositions) and possible worlds (which are tied to either true or false propositions that may or may not be true in our actual world). Doležel (1988) summarizes three fundamental theses for the possible worlds theory as applied to fictional texts:

1. “Fictional worlds are sets of possible states of affairs” (1988:482)

This first point refers to the idea that fictional worlds are legitimized as distinct from the real, actual world and links to Searle’s notion of pretence of states of affairs, as discussed above. Importantly, fictional entities are consistently independent of the actual world (principle of ontological homogeneity), which is “a necessary condition of the coexistence and compossibility of fictional particulars” (1988:483). This leads into the second point Doležel makes:

2. “The set of fictional worlds is unlimited and maximally varied“ (1988:483)

Here, Doležel highlights the important notion that any possible fictional world is indeed fictional to the same degree. He says that “there is no justification for a double semantics of fictionality, one for fictions of the ‘realistic’ type and another one for ‘fantastic’ fictions” (1988:483). The fictional world of *Sherlock Holmes* for instance has high degrees of contextual recognizability for the consumer of the text, be that literary or audio-visual, as the city of London is well known in our actual world. This stands

seemingly opposed to fictional places such as Hogwarts (*Harry Potter* book series and films), Winterfell (*Song of Ice and Fire* series, *Game of Thrones* series), or Sunnydale (*Buffy the Vampire Slayer*), which exist with little or no concrete reference to the actual world.<sup>8</sup> Nevertheless, all fictional worlds, independent of the degree of likeness to the actual world, are to be interpreted as equally fictional. This somewhat touches upon Mandala's (2010) exposition on how literary critics, in the past, have favoured fiction that was set within worlds modelled more closely after actuality:

When the 'respectable' literary world did take an interest in science fiction and fantasy, it insisted, as Parker (1956:601) noted, on reclassifying it. Tallis (1984), for example, accepted *Nineteen-Eighty-four* and *Gulliver's Travels* as great works of literature by arguing that they were more mimetic than fantastic in their 'scrupulously realistic realization' (Tallis, 1984:193). (Mandala, 2010:3-4)

Following Doležel's principle then, analysis of fictionality should not be evaluated by its probability to be real, as, "for the reader, (...) the imaginative leap into the novel's world of time and space must be made in both cases" (Hutcheon, 1980:78, in Doležel, 1988:483). Rather, a fictional settings is defined "by its own global, macrostructural organization" (1988:484). This means that possible worlds are possible because they make sense within the contexts they create.<sup>9</sup>

---

<sup>8</sup> Although they are often ambiguously compared to areas the audience might recognize (*Hogwarts* as a place in Scotland, *Winterfell* in the north of a continental island that is similar to the British Isles, and *Sunnydale* as a typical city in California in the United States).

<sup>9</sup> This is, of course, somewhat complex in fictional story telling that relies on its audience not knowing the whole fictional truth. In some cases, sense-making is actively challenged whereby the possible world that is established as fictional canon is called into question (see for instance the finale of *Roseanne* which established the possible world of the previous seasons as embedded in another possible world in which the main character dreams the story line, or episode 'Normal Again' in *Buffy the Vampire Slayer*, in which the audience is switching between two possible worlds, one describing the previously known plot, the other describing a possible world that is aligned more closely with our actual world (i.e. no vampires)).

### 3. “Fictional worlds are accessible from the actual world“ (1988:484)

Doležel’s final principle of possible worlds theory in application to fictional texts describes what Searle calls nonfictional commitment (see above): “The actual world participates in the formation of fictional worlds by providing models of its structure (...)” (1988:485), or, as Semino writes: “we use our knowledge of reality to fill any relevant gaps in the content of fictional worlds” (2014:64). The state of affairs that constitutes pretence in fictional contexts thus has no claim to being real or authentic, but rather encourages its audience to believe in the possibility of the fictional through recognition of familiar patterns.

To put this theory into concrete context, I will highlight the defining principles of fictionality via *Torchwood*, a British science fiction series and part of the corpus for this thesis (for plot and character summary see chapter 3). *Torchwood* is set in a fictional version of Cardiff and tells the story of the people working for the Torchwood institute, a secret agency that deals with alien life forms coming to Earth. The first episode begins as main character Gwen witnesses Torchwood’s operations and as a result learns about the existence of aliens and supernatural powers in Cardiff. The beginning is very much contextualized in a world familiar to the audience; “actual world Cardiff” contributes to the construction of “*Torchwood* Cardiff” by providing recognizable landmarks, such as the Millennium Centre, or Gwen’s Welsh English accent. Additionally, the structure of the fictional world of *Torchwood* is made possible through consistent world-building: the story unfolds in, for the fictional context, logical ways and the audience recognizes the fictional world as consistent. Supernatural elements that are part of the series are as much indisputably impossible in “actual world Cardiff”, as they are possible in “*Torchwood* Cardiff”. The audience is able to appreciate that distinction while following along with the episodes, something that Semino alludes to by stating that “the construction of

alternative realities by verbal or visual means is a universal cultural phenomenon, and ordinary people have no difficulties in extending their everyday notions of truth and falsity to the products of such activity” (Semino, 2014:61).

Eder (2008:65) provides a related notion of fictional worlds and says that

eine fiktive Welt wird dort verstanden als System nicht-wirklicher, möglicher Sachverhalte bzw. als Zusammenhang von Gegenständen, Individuen, Raum und Zeit, Ereignissen, Gesetzmäßigkeiten usw., der durch einen fiktionalen Text konstruiert wird.

(a fictional world is understood as a system of non-real, possible circumstances, or rather, a connection of things, beings, space and time, actions, norms, etc., which are all contextualized inside a fictional text)<sup>10</sup>

Summarizing these points, fiction can be described as a pretended state of affairs that proposes possible circumstances on the basis of our real-world understanding and recognition of what is, or can be, true and false. The idea that fiction is somehow unreal and not of ‘this’ world (whatever the interpretation of that may be) is herewith irrelevant and the notion of what is authentic in terms of fictionality is re-defined away from authentic “real” to authentic “possible”:

The genesis of fictional worlds can be seen as an extreme case of world-change, a change from nonexistence into (fictional) existence. The special illocutionary force of literary speech acts that produces this change is called the force of authentication. A non-actualized possible state of affairs becomes a fictional existent by being authenticated in a felicitously uttered literary speech act. To exist fictionally means to exist as a textually authenticated possible.” (Doležel, 1988:490)

---

<sup>10</sup> Translation mine.

This is also acknowledged by Searle, who states that “Holmes and Watson [from Doyle’s *Sherlock Holmes* series] never existed at all, which is not of course to deny that they exist in fiction and can be talked about as such” (1975:329). Possibly one of the most thorough works that does just that is Eder’s analysis of characters as fictional entities. He defines characters, in their simplest form, as “wiedererkennbare[s] fiktive[s] Wesen mit einem Innenleben – genauer: mit der Fähigkeit zu mentaler Intentionalität” (recognizable fictional beings with an inner life – more specifically: with the ability of mental intentionality)<sup>11</sup> (2008:64). Intentionality here, he says further, is not to be interpreted as aim or purpose, but rather an object-directed reference: characters can think, feel, wish, and hope (2008:63-4). This intentionality, as an intrinsically human quality, creates the parasocial relationships consumers of fictional content, from literature to film and television, experience. Giles (2010:443) illustrates this, saying that “Harry Potter, Prince Harry, Harry Hill; in the imagination of the reader or viewer these distinctions blur. There are just Harries [sic], as real as Harry who lives next door; perhaps more real, since they barely know Harry next door except to see him occasionally putting out his rubbish”. In that, the recognizability of characters relates to the previously mentioned non-fictional commitments of fictional stories. The audience models their perception of characters after what they know from the actual world; we perceive evil characters as evil because their actions are deemed as such within our (actual) world; we are empathetic to characters because we experience empathy in our daily lives.

---

<sup>11</sup> Translation mine.

When Walter White from AMC's *Breaking Bad* is driven to illegal activities by his inability to pay for his cancer treatment, the audience feels conflicted in recognizing his actions as illegal but his reasons for the crime to be ultimately understandable.

Further, the repetition of action by the characters evokes the notion of familiarity: the audience expects and, in some cases, can predict how characters behave.

The Canadian science fiction series *Orphan Black* has more than 9 characters performed by the same actor (Tatiana Maslany). Through consistent presentation of each character, the audience is able to tell the characters apart, despite them looking similar. This goes as far as being able to tell when one character imitates another, explored also in Queen (2015a).<sup>12</sup> The extract from *Buffy the Vampire Slayer*, reproduced in the introduction, shows a similar play on audiences' familiarity with a character's 'usual' self. When characters change over time, the audience takes note because they have gotten to know these characters in a specific role.

In some cases, this change can also be seen as a breach in authenticity in so-called out-of-character moments. An instance highlighting this clearly is given at the end of season five of *Game of Thrones* (HBO 2011-), when Stannis sacrifices his daughter with whom he has developed a close relationship in just the previous episode. Fans of the series debated for weeks how this choice defined Stannis, how it went against previous character exposition and how it was not authentic and standing opposed to the 'real', because, as one article puts it: "The *real* Stannis would rather die making an honest effort at getting the throne [sic] than live with killing his own daughter trying to get it [emphasis mine]"

---

<sup>12</sup> In *Orphan Black* these moments occur frequently, e.g. Tatiana Maslany as Sarah (with a London-based working class accent) switches into the persona of Beth (a Canadian character) and the audience is cued in because of accent slip-ups where the Canadian accent is deliberately interrupted with British accent features.

(Brayson 2015)<sup>13</sup>. What happens here is that fictional content creates its own level of authenticity, in storytelling and plot development as well as through dialogue and “typical” voices.

Despite being an arguably essential part of empathetic consumption of a fictional story, characters have not always been recognized as particularly meaningful. As previously mentioned, fictional characters have only recently found recognition within sociolinguistic research due to the inherent focus on naturally occurring language (further explored in the following section). Stylistics, as the study of style in (literary) fiction, has sparingly engaged with the conceptualization of characters. Leech and Short, in the second edition of *Style in Fiction* (2007: 296), acknowledge this fact and say

At the time we wrote the original book, characterisation was not very high on the agenda of prose fiction study in stylistics, or criticism more generally. Perhaps because of the dominance of the critical discussion of viewpoint in prose fiction, the study of characterisation in twentieth-century criticism was, by and large, more evident in the criticism of drama than that of prose fiction, though it was not very plentiful even there.

They continue by drafting the development of character within the field, saying that characters, in the early 1900s were considered ‘real’ people and any interpretation was thus highly based on real world analysis. This was then met with the argument of fictional characters as being not just ‘unreal’, but also irrelevant to the textual analysis of any fictive text (cf. Weinsheimer (1979)). Only at the turn of the century did stylistic studies take up the concept of fictional characters again, with Culpeper (2001) providing one of

---

<sup>13</sup> The ‘real’ here might refer to the character established over the previous episodes of the series, or possibly the character from the source material, Martin’s *A Song of Ice and Fire* series. In both cases, it is worth noting that the audience perceives there to be realness to a fictional entity that was built up through cumulative story telling.

the most thorough investigation of characters, their language, and behaviour. His work, summarized further on in this chapter, links the character to the text level and enables a linguistic analysis of the in-text fictional being, rather than the author behind each character.

While stylistics is predominantly concerned with the artistic effect of written and/or performed content, other branches of linguistics have approached mediated and fictional language of characters in diverse ways. The next section presents a summary of these different approaches before turning to the framework employed within this study.

## **2.3 The language of television characters and previous linguistic applications**

### **2.3.1 *Television dialogue***

Television language in general is staged performance: rehearsed, self-aware, stagey, and at times hyperbolic (Bell and Gibson 2011:558 after Coupland 2007). Performance studies include radio broadcasts, live performance speeches and music videos, which has increased in focus in linguistic research as well as interdisciplinary studies (see for instance Bell, 1992; Bell and Gibson, 2011; Coupland, 2009; 2011; Dumas, 2016; Stamou, 2011; 2012; Stamou et al., 2012). Within sociolinguistics, working with media texts has developed into a well-established research area and it is possible to find a wide array of studies on news coverage (print and broadcast), talk shows and radio presentations, music, as well as online data such as social media or blogs (see for instance edited volumes by Bednarek and Martin 2010; Johnson and Milani 2010).

When it comes to fictional audio-visual language, uptake in linguistics is on a steady increase (Androutsopoulos, 2012a; b; Queen, 2004; 2015a; b; Richardson, 1999; 2010; Richardson and Queen, 2012; Tagliamonte and Roberts, 2005), despite some initial

hesitance. As indicated in the introduction, fictional television in particular is still seen by many as too trivial for serious academic engagement. The disregard for including fictional television in linguistic research is partly due to the very fact that it is scripted speech, which seems somewhat antithetical to the purpose of (socio)linguistic studies:

In those branches of language studies in which *conversation* is of interest, naturally occurring unmediated talk takes precedence over other kinds because of its greater claim to authenticity [emphasis in original] (Kozloff 2000:14)

As such, one might argue that any observation of language as spoken by a fictive person is unreliable and of little use to the study of language within *our* [read: *real life*] society.

If the purpose is to investigate real life language variation and change in the traditional sociolinguistic sense, looking at fictional television or film data is indeed futile, because, as Kozloff (2000) puts it: “[...] dialogue may strive mightily to imitate natural conversation, but is always imitation”. She also notes that “[l]inguists who use film dialogue as accurate case studies of everyday conversation are operating on mistaken assumptions” (2000:19).

Consequently, the focus of sociolinguistic studies on fictional dialogue should not be to investigate naturally occurring language but rather related themes including notions of authenticity, replicability of variation, indications of change, perpetuation of indexicality, etc.

While people may argue that television has lost its impact as “the dominant mass medium of the second half of the twentieth century” (Richardson 2010:3), streaming services such as Netflix or Amazon Instant Video give reason to believe that audiences have merely shifted the way they consume, not their consumption overall. In fact, the most recent consumer report carried out through Ericsson (2016) says that while linear television

viewing times declined by 2.5 hours per week from 2012 to 2016, hours spent watching content on mobile devices actually increased by 4 hours. Weekly hours spent watching fictional content, such as the here relevant fictional television series, are averaged at a total of 7.5 hours across devices (based on reports including representative data from more than 40 countries)<sup>14</sup>.

Furthermore, the new multimodal availability of television series through not only viewing the content itself allows for an ubiquitousness that goes far beyond the approximately 42 minutes of weekly entertainment per series episode<sup>15</sup>: audiences partake in fan discussions online or panels at conventions, read news articles and interviews, as well as consume spin-offs in novel, audio, and video game formats. Despite its overwhelming presence, influence of this type of media on its audiences has always been a contested issue in social studies, including linguistics. A focus article by Sayers (2014) illustrates current discussions within the field, something he calls “one of the hottest potatoes in early twenty-first century sociolinguistics” (2014:187). Laypeople often blame the media for so-called language decay<sup>16</sup>, but linguists are hesitant to make media responsible for language changes. While some scholars dispute any impact television might have on language use or change, others are open to the idea, claiming

---

<sup>14</sup> Based on Ericsson ConsumerLabs (see reference list). Concrete viewerships for individual series is increasingly difficult to obtain as cross-platform consumption spreads across audiences.

<sup>15</sup> *Buffy the Vampire Slayer*, as one of the shows included in the analysis of this study, was on air for seven seasons until 2003, was exported worldwide and has reruns even ten years after officially wrapping. Numerous novels and comics were published, a successful spin-off created (*Angel*, also included in the analysis), as well as merchandise ranging from DVDs, video games to action figurines and jewellery.

<sup>16</sup> In an article in *The Guardian* on the changing semantic meaning of *literally* within the Oxford Dictionary (meaning not only the opposite of *figuratively*, but also used as emphazier), the author also complains about the increased use of discourse marker *like* thusly: “What’s this little linguistic slimeball doing? It fills cracks. In an ugly way” (Sutherland 2013) Elsewhere exactly this usage is brought forward when complaining about influence television has on language, asking “when the hell did articulation become a television novelty?” (Rafidi 2012).

that new linguistic features were detected on a surprisingly global level that suggests equally global causes.

In response to Sayer's article on possible influence the media might carry with regards to language change, Trudgill (2014) noted that lexical change (or "catchphrases" which he claims are oftentimes short-lived "fads" (2014:216ff)) could indeed be taken from exposure on television programmes or written media as it is change "above the level of conscious awareness" (2014:215). At the same time, he disregarded any influence when it comes to change below the level of consciousness, such as phonological shifts, claiming that if influence was indeed the case, a spread of American pronunciation for instance would be detectable on the British Isles (2014:216). In a follow-up commentary Androutsopoulos (2014) attests to proposed media influence on standardization of local dialects in continental Europe in various studies (2014:242). Simultaneously, he points out the importance of being wary of generalizations with data that is difficult to obtain and processes that are "too complex and unpredictable by [a] framework's notion of 'media texts'" (2014:246). Finally, Stuart-Smith (2014) summarized findings from the Glasgow Media Project concerning 'Cockney' features such as TH-fronting in the Glaswegian vernacular. She said "media factors might be observed as distinct factors alongside other social factors" (2014:255) when it comes to language change, though suggested that speakers were not acquiring new features as much as "enabling existing or latent variation" (2014:257). Tracing possible influence of media language on naturally occurring language is, as can be seen from the many different viewpoints, not easily done. Similarly, in what ways our language is affected by television and film will possibly remain a disputed issue among linguists. What is not disputed however, is the impact outside of just language change. Television series do not only imitate language; they use language to portray attitudes, introduce and reinforce characters and, in the most basic

function of the medium, tell a story. In that, most scholars are agreed: exposure to media such as television has become remarkable and suggests that it might intrude in the audiences' lives in a variety of ways. As stated earlier in this section, media is becoming more and more ubiquitous and while research into its exact traces within language change is debated, linguistic research is nonetheless diverse and ever-growing.

### ***2.3.2 Previous linguistic approaches to television dialogue***

In the following I will map out different directions research has taken in investigating media, specifically fictional language as found in television series and film, and where the present study can be grounded.

Within linguistic research of fictional language there are two main strands of studies that divide scholarship: those that use media language as a means to investigate our society, and those that use it to learn about language within the medium itself.

#### *Linguistic studies using fictional media*

Part of that first group are a number of attitudinal studies of fictional dialogue in both television and film that have focussed, among other things, on multilingualism (Mandala 2008; Bleichenbacher 2012), non-standard varieties (Lippi-Green 2012; Planchenault 2012) and gender issues (Rey 1996). They observe language patterns that are indicative of attitudes to refer back to societal issues: for instance, how representation of language diversity in animated movies might raise negative associations of non-standard varieties with children, as discussed in Lippi-Green's study (2012).

In one of the earliest sociolinguistic studies on fictional television, Harwood and Giles (1992) investigated attitudes towards the elderly in the sitcom *Golden Girls* (NBC, 1985-

1992<sup>17</sup>). They illustrated how the conscious use of language in specific situations, as well as by certain characters, can create attitudinal dynamics of this particular demographic group. While they stated that visibility of the elderly on television itself is a prime achievement of the series, they also observed how humour and particular external comedic strategies (i.e. laughing track and *mise-en-scène*) are managed in order to paint a rather negative picture of the main characters, all elderly women. The overwhelming combination of dialogue containing age marking with humorous effect (1992:426) and prescribed laughter creates an almost Pavlovian reaction to ridicule, “sustaining, and contributing to, an agist [sic] culture” (1992:405). Furthermore, many of the age-marked features spoken by the main characters focus on negative effects of aging (1992:427). This is quite a surprising finding, as previous accounts of the series (by media critics and audiences) actually praised its positive representation of that demographic (1992:404-5). Their study illustrates the ways in which fictional television can be used to observe attitudes and attitude changes, as well as the importance of a close linguistic reading of the series’ language. While superficial observations of the series’ content remarked positively on the depiction of the elderly, a detailed analysis such as the one conducted by Harwood and Giles can shed new light on the many ways language is using prevalent attitudes that are incorporated (yet not necessarily always apparent) into a show’s dialogue and is able to uncover the possible impact it may have on the audience. It becomes clear here that an analysis of scripted language can aim at many more foci than simply using a “copied” version of our everyday spontaneous language to investigate sociolinguistic entities.

---

<sup>17</sup> Information of television series includes the original network, as well as first broadcast times on American/British television, depending on the show’s origin

Other linguistic studies that are part of the first strand of fictional language research, as mentioned above, investigated areas such as language teaching (Rodgers and Webb 2011; Dose 2013) matters of translation (Herbst 1994; Queen 2004; Mittmann 2006), or power discourse (Ruddell 2006; Herrmann 2013).

All of the aforementioned studies observe fictional dialogue from a more or less outside perspective in that they research ways in which language from this particular medium can give insights into phenomena within our world. They turn towards media data to see how it can be used in non-fictional areas or how it might mirror or trigger attitudes affecting society: what fictional language does to and for us, the outside viewer.

#### *Linguistic studies within fictional media*

The second strand of linguistic research of fictional television is more reflective in that it focuses on the language within the medium itself. The central point here is how language is appropriated inside the programme, rather than how language can be used for societal efforts such as teaching or raising of attitude awareness.

Following Adolph's distinction between inter- and intra-textual analysis (2006:65-9), I differentiate two main sub-groups of studies here: inter-source and intra-source.

Firstly, inter-source, or comparative studies investigate fictional television (or film) language in order to compare it to naturally occurring language. It is studies like these that will focus on authenticity and realization of realities of our language within fictional contexts. And secondly, the subgroup where I mainly locate my own research: intra-source studies focus on language within fictional contexts and how it is used to portray speakers or characters within their fictional setting and independent of authentic representations of what we call reality.

Both subgroups inform each other as it is vital to acknowledge the fictional context and performative aims when comparing spontaneous language to scripted language just as much as one needs the context of how speaker types are conceptualized through language within naturally occurring contexts to infer characterization within scripted language. I will introduce studies from both subfields in more detail in the following, highlighting what they contribute to the general area of study and how they relate to my own study.

#### *Comparative Studies: Inter-source*

In a corpus study of a range of fictional television series from various genres, Bednarek (2012) set out to “explore the question to what extent television dialogue is a language variety in its own right and what features [...] characterise such a variety” (2012:36). According to her, while frequencies of particular features are different to naturally occurring language, overall patterns are comparable and that, much like the languages we find in real life, language in fictional television series is highly dependent on the speakers’ (or characters’) “settings, relationships, actions, [and] events” (2012:60). Furthermore, recent studies have found that *telecinematic discourse* (Piazza et al. 2011), i.e. the language found in films and television, exhibits language patterns congruent to sociolinguistic findings and that methods of linguistic research are indeed applicable. This means that scripted language can be analysed similarly to spontaneous language, a fact that led many researchers to conduct comparative studies: studies that look into the series’ and their use of language, always with reference to naturally occurring language.

In a comparative study investigating the use of intensifiers, Tagliamonte and Roberts (2005) were able to compare the language of fictional television series *Friends* (NBC 1994-2004) to two studies that were conducted in similar methodological frameworks (Ito and Tagliamonte 2003; Tagliamonte 2008). Intensifiers, which are prone to quick

change in use in a relatively short amount of time, were found to be used in patterns similar to what was previously found in spontaneous speech. Further, the change from preferred use of one intensifier to another was not only traceable on the show: it appeared in sociolinguistic patterns that mirrored expectations set by naturally occurring language change. Tagliamonte and Roberts highlighted that female speakers were more likely to use new incoming features and even suggested that television language might indicate ongoing language change where data from unscripted language was not as readily available for analysis (2005:297).

Notable in this regard is also Quaglio's work (2009). He compared a number of pragmatic features (including intensifiers) from sitcom *Friends* to a corpus of conversational American English (from the *Longman Grammar Corpus*) and found that while certain differences are observable, scripted and spontaneous speech are similar in many ways. He proposed that further analyses of language features through television dialogue corpora "seems perfectly appropriate" (2009). He offers a range of possible research avenues, such as diachronic analyses of features seldom found in spontaneous talk, conversation studies, and realizations of humour (2009:149), albeit never acknowledging television language's value outside its relation to naturally occurring language. His, as well as Tagliamonte and Roberts' (2005) research are highly relevant to the present study, even though the approach is rather limited to a comparative analysis. By investigating a number of individual features and frequencies of occurrence in both corpora, they are able to account for a detailed linguistic profile that will be used as a starting point for the present analysis.

Sociolinguistic patterns are apparent in fictional television data, indicating that common theories of gender or age variability could be further investigated and attributed to creation and support of characterization.

Bednarek, in a number of further studies, adopts Quaglio's findings and investigates fictional television language on a comparative level and highlights methodological issues that arise from studying scripted discourse. Her research finds that it is particularly promising to focus on expressive features (2008; 2011), features that appear comparatively frequent, as well as significantly for characterization purposes within scripted language. Sub-chapter 2.4 details this study's aim to expand Bednarek's findings of how expressivity can be used as an indicator for characterization in fictional contexts.

While Bednarek treats television language as a variety distinct from naturally occurring language, her main analyses are still primarily focused on comparing scriptedness with spontaneous discourse. Although this offers invaluable insights into describing language as presented on television screens, it goes little beyond that. Her detailed study on characterization, which introduces a first thorough investigation of the linguistic means of character creation, will be discussed below.

#### *Stylistic studies: Intra-source*

Studies in this second subgroup of linguistic studies that investigate the language within fictional settings such as television or film are only slowly emerging, particularly from a sociolinguistic background. Many studies on performed language have a strong focus on the question of authenticity which inevitably calls for comparative studies first and foremost. I will highlight below how authenticity in itself is an issue in fictional contexts that is often ambiguous.

The interest of many studies lies with the aim to explore how a medium can represent language and make the audience believe in what is happening on screen. Within a field that once set out to explore language that is at its core real, authentic and spontaneous, it becomes a prerequisite in sociolinguistics to refer back to exactly that. Allowing research

into what is essentially 'fake' language then comes with the promise to relate this scriptedness to 'real' language. In the following I will present studies that focus on the language in these fictional settings without a comparative approach.

Linguistic studies that investigate how language is used in scripted television and film language are usually grounded in stylistics, such as Queen's study (2015), which investigates language variation and characterization in film and television. Her study examines "what language and language variation add to the media (**rather than what the media reveals about language variation**) [emphasis mine]" (2015:154). As mentioned previously, these studies are more reflective and inward-aimed when it comes to scripted language and only take linguistic background theories as a starting point for investigation.

In a detailed analysis of one of the main characters in the television series *Gilmore Girls* (WB 2000-2007), Bednarek (2010) focuses on character stability and uniqueness. In comparing Lorelai's speech patterns to each of the other characters, she is able to define the various and more importantly varying semantic fields individual characters would talk in. Further, she finds that Lorelai, throughout the included seasons of analysis, is relatively stable in her use of language and specific usage changes only plot-dependent (more frequency of work-related language in season where her work-life became central theme for instance).

Bubel (2006) also conducted a study on interactional language and investigates the four main characters' use of language in two seasons of *Sex and the City* (HBO 1998-2004) and how friendship between the characters can be inferred by the audience through linguistic means. Her overall findings are that linguistic patterns of alignment and affiliation not only indicate *that* the four main characters form a close-knit group of friends, but that two of the four women form a core group with the other two being

marginal members (2006:257-260). This is a particularly important finding for the present study, as it shows that linguistic means are not only used to give an individual voice to the characters, but also to their relationships with each other.

Mandala (2007), in a study that examines friendship bonds within a fictional television series, focuses on the characters' usage of *y*-suffixes in the creation of marked adjectives in three seasons of *Buffy the Vampire Slayer* (WB/UPN 1997-2003). Instances of the phenomenon are, among others, *Heart-of-Darkness-y*, *cute-y*, *non-slay-y*, *out-of-the-loop-y*, or *kabloo-y* (2007:56-57). Her study relates to a finding by Adams (2003), who includes this feature in his lexicon on *Slayer Slang* and says that it "has very quickly come to characterize the whole Scooby Gang [the central group of friends on the show], serving as an adhesive that binds them together" (2003:42). His account however is not supported by detailed linguistic analysis, but intuition. Mandala (2007), in her study, expands on Adams' findings by systematically looking at frequencies and how the *y*-suffixed adjectives are consciously employed to highlight core members of a group. She finds that this marked feature characterizes the group and that the frequencies in which they are used indicate in-group membership statuses of the characters.

While this study only accounts for marked features, it also indicates the varied roles linguistic choices can play in the creation of character. Mandala refers back to previous research on social networks by Milroy (1980), who says that close-knit networks are oftentimes characterized by shared vernacular features (cf. Mandala, 2007:58). The deliberate use of linguistic forms that cater towards individual characters and character groupings is especially remarkable. We find here one of the first quantitative discussions on how fictional television language is not simply written-to-sound-like-spoken language, it is language that carries consciously created linguistic choices for the sake of characterization.

In recognizing what is done on a linguistic level on television, we can focus on the conscious command writers and actors have over linguistic means, that is, in what ways they know of the impact linguistic choices have on creating characters. The question of how much awareness goes into creating fictive content and how language works in a medium that demands constant attention from its audience.

The following excerpt from the show *Buffy the Vampire Slayer* (WB, UPN 1997-2003) for instance, if repeated in “our” world, would not make much sense. In the show however, the viewer is not startled by mentions of vampires and demons – they are willing to believe in these creatures for the sake of the story and focus on the emotional aspect of this scene.

- (4) Buffy: Impulsive? Do you remember my ex-boyfriend, the vampire? I slept with him, he lost his soul, now my boyfriend's gone forever, and the demon that wears his face is killing my friends. The next impulsive decision I make will involve my choice of dentures.  
(*BVS*)

What the audience is watching here is Buffy’s emotional reaction to what has happened previously in the season; her getting to terms with a break-up. The particularities of the story-telling, namely, that this break-up was due to mythical happenings and emergence of the literal evil is beside the point; credibility is earned through the evocation of a possible world, not necessarily realistic settings. Through the possible world and recognizable patterns audiences get invested in the characters’ fates and hang on their every word, even discussing at length what certain parts of dialogue meant and what role the chosen words play in the overall plot. Cordia, a blogger, summarizes her opinion on a later episode from *Buffy the Vampire Slayer* in the following:

Anya's dialogue was very interesting to me. It starts off with her typical blunt questioning and then degenerates into her true feelings of loss, confusion, and questions about morality. [...] And I really liked how her personality was used at the hospital when she says exactly what needs to be said about Joyce to Buffy. (Cordia 2013)

These sorts of responses by audiences underline the impact a television series can have outside of network ratings, but also how important dialogue writing is for the show and its appeal for the audience. Even years after television series have been aired, discussions such as the one above are kept up, making these fictional places and characters relevant to our own lives. The fact that the extract above highlights what the blogger thinks is 'typical' for Anya suggests that she is so familiar with this character that she knows how she typically sounds. The writers thus managed to create a speaker that fuels linguistic anticipation in the audience, making this fictional character authentic within the fictional setting.

In this study, I investigate these voices within the fictional setting, not comparing the fictional to the real but taking the fictional *as* real within its context. The present study thus aims at looking into the stories and the language used, in particular how the characters are created to be unique speakers that pose as subjects for this kind of audience reaction.

### **2.3.3 *Fictolinguistics and cinematic indexicality***

As this chapter has highlighted, in recent years new theoretical approaches have widened the means by which telecinematic discourse can be analysed. By borrowing from stylistic research of literature (prose), we are able to observe sociolinguistic goings-on within the fictional world by reframing authenticity away from the notion of 'real' and towards the

notion of ‘fictionally appropriate’. The following theoretical positioning forms the backbone of my research and informs both methodology and interpretation of results.

Ferguson (1998), in a study examining seemingly inconsistent dialects in Victorian novels, introduces the term *fictolinguistics*, i.e. “the systems of language that appear in novels and *both* deviate from accepted or expected socio-linguistic patterns *and* indicate identifiable alternative patterns congruent to other aspects of the fictional world [emphasis in original]” (1998:3).

Using three Victorian novels (*Wuthering Heights* (1847), *Bleak House* (1853), and *Tess of the D’Urbervilles* (1891)), Ferguson illustrates the advantages of observing dialect in novels not only by their consistency and authenticity regarding expectations the reader has according to real-world social knowledge, but also through the fictional lines of the narrative (1998:7). In her analysis of Dickens’ *Bleak House* for instance, she finds that dialect features are used by the author to “construct and reinforce boundaries between characters” and “suggests that the function of dialect in *Bleak House* is not simply to present realistically the various sounds of different ways of speaking, but to establish characters in a metanarrative system” (1998: 8). The dialect features found in these works of fiction mark not only a representation of the language we know to exist in our world, but they are also used to present characters within their own world; characters’ speech has to convince the audience/reader (it has to sound ‘real’), the character must seem likely to exist in the fictional possible world (dialogue has to be appropriate for the setting), and it has to individualize the character as distinct from those around them.

The concept of *fictolinguistics* refers to what Androutsopoulos (2012) later calls *cinematic indexicality*, which he defines as “a cover term for the various layers of indexical meaning that sociolinguistic difference can articulate in film” (2012:302) and

with which he questions the “notion of authenticity in sociolinguistics, which problematizes the epistemological and methodological conditions as well as implications of how authentic dialect speakers have been theorised” (2012:302). Going back to the issue of traditional sociolinguistic research and its quest for authenticity, this supports how authenticity is ambiguous in contexts of fictional narratives. Characters within fictional stories are as authentic as they are inauthentic to the non-fictional outside, as discussed previously in this chapter.

For the sake of continuity and ease of understanding, I will present my analysis under the term of fictolinguistics, which includes the notion of cinematic indexicality within this framework. Both perspectives frame sociolinguistic means of naturally occurring language as categories constraining language use, but accept societal factors influencing language variation to be “contained within” the work of fiction:

To understand how dialect works in the novel, we must understand how it fits within the socio-linguistic system *constructed* by the novel [...], as well as how it responds to the socio-linguistic patterns accepted and expected by the world outside the novel [emphasis by me]. (Ferguson 1998:3)

This addresses the issue that literary critics oftentimes remark on the fact that a dialect is not accurately depicted within a fictional world and thus seemingly represents a character and their background wrongly. The concept of fictolinguistics mitigates the need to explain “deviations from the expected norm [what we know from the real world]” and rather seeks explanations of language variation within the world created in the text (1998:3).

Fictolinguistics, although applied mainly within literary studies by Ferguson and others (Hodson, 2014; Ilhem, 2013; Reich, 2013) can be applied to any work of fiction, including television and film. Although it is not a new concept by far, studies referring to

fictolinguistics are surprisingly sparse. In her volume on dialect in film and literature, Hodson (2014) picks up this concept and gives further credence to its appropriateness in this particular field of research:

The term thus moves us beyond analysing language varieties in literary texts in order to rate them in terms of their real-world accuracy or consistency [...] and instead enables us to see that they form an integral part of the fictional world within which they appear. (2014:14)

Hodson illustrates a variety of examples from film and literature in which the dialect presented did not accurately display “real world” dialects. In written form, any kind of phonetic representation seems inauthentic without the use of official phonetic alphabets, which of course is not ideal in a novel not exclusively read by linguists. In film, the performance of dialects is made easier through the element of audio recording. However, actors are not always native users of the dialect that needs to be depicted and sometimes an authentic representation will even make the film unintelligible for audiences who do not know that particular variety (see Hodson (2014: 63) for a detailed account of dialect representation in the film *Trainspotting* (1996)). In both cases the audience (or reader) is confronted with a version of dialect that needs to account for readability, intelligibility, performance and of course awareness – a version, I argue, which within the fictional setting still exhibits uniqueness of characters, comprehensible variation and stability.

Talking about linguistic characterization for film (and for the sake of the present study, small screen media such as television), Hodson investigates the role of stereotyping within character creation. As script writers try to establish characters on screen in a quick and recognizable fashion, dialects that are used to aid a character’s identity will oftentimes become a representation of cliché (2014:60). Hodson hereby distinguishes between *character stereotyping*, as the representation of character groups, and *linguistic*

*stereotyping*, the portrayal of non-standard dialects in film (2014:61). Within this study those two concepts are co-dependent, as any dialect will inevitably inform any kind of character, which in turn is established through perceptions the audience has of particular personas. Successful writing on television will usually break away from this first set of stereotyping to further develop a character, ultimately shifting the audience's recognition of the character from a stereotype into a unique character (minor characters are oftentimes limited in that regard and remain one-dimensional clichés).

Stereotyping is an integral part of storytelling in film and television, as it is a tool through which writers can quickly create easily recognizable characters to ensure that the audiences engage quickly with the content. An example can be seen in the science fiction show *Firefly* (FOX, 2002-2003) and the accompanying movie *Serenity* (2005), in which main character Mal is portrayed as a loveable scoundrel. Many instances show how he is portrayed as a rash con artist, eager to do harm before thinking, typically dressed in scruffy clothing and carrying a gun. Linguistically, he is heard using non-standard forms such as in (5) or expletives as in (6), facilitating to the prescriptivist notion of non-standard language equating lower or even antagonistic status, a pattern previously mentioned in Lippi-Green (1997).

(5) Mal: You're welcome on my boat. God ain't. (*Firefly*)

(6) Mal: Once, just once, I want things to go according to the gorram plan! (*Serenity*)

Throughout the series he engages with this stereotype and turns it around into making the audience trust him and his actions, showing that his apparent wrong-doings are quite the opposite. His non-standard features are combined with a new stereotype, shifting the linguistic stereotype of a dialect from a formerly negative index to a positive one.

Linguistic features are used to underline characters' traits, but they can also be representative of an ensemble of characters. In sociolinguistic approaches, there are few studies that have examined fictional television dialogue with reference to theories of social networks and communities of practice.

One of many instances in which this awareness of linguistic means to build characters becomes clear is taken from *Buffy the Vampire Slayer*'s final season, in example (7) below.

- (7) Buffy: It's like all the Hellmouth's energy's trying to escape from  
that one little spot, and it's getting all...  
Robin: Focus-y.  
Buffy: Careful - starting to speak like me now. (*BVS*)

Robin Wood, a character who first appeared only a few episodes prior to this scene, is talking to Buffy, who is part of the core group of characters and title character of the show itself. He uses the marked *y*-suffix and is instantly marked as an unusual user of this code, indicating an awareness of the feature and what it stands for by the character (and thus the writers), but also an expected awareness of the audience in recognizing this to be a linguistic reference to a particular part of *Slayer Slang*, thus making this a humorous remark.

This goes back to the concept of *fictolinguistics* and demonstrates how linguistic theories (in this case theory on linguistic indicators of friendship networks) can be applied to fictional texts, despite the fact that the linguistic patterning is outside of naturally occurring language's actuality. Furthermore, departures from these theories can point to very specific choices about the characterization process, (much in the same way Ferguson (1998) found in her analysis of literary texts of *Tess of the D'Urberville* and others).

Mandala's previously introduced study (2007) notes that in some instances linguistic theory cannot account for all observed frequencies of the marked *y*-suffix feature. In her discussion on the ambiguous character Spike, she remarks that his high use of the feature is surprising as he is introduced as a villain and never fully accepted into the friendship group. Throughout the show however, he is portrayed as a complex character, a vampire who falls in love with the slayer and goes from being The Big Bad<sup>18</sup> to being the redemption-seeking man who eventually saves the world.

Seeking linguistic regularity in variation thus appears to be far more complex and in ways related to the distinctive waves of variationist theory that were introduced in the previous chapter. It seems easiest to focus on clearly defined character types (the hero, the villain, the comic relief, etc.) and to trace linguistic markers that betray relevant group belongings. Television dialogue, as seen here and further explored throughout this thesis, is much more nuanced than that and breaks with the audience's expectations to establish a character's evolving identity: "Where the results are surprising with respect to social network theory, these deviations see strategic, thematically motivated departures from expectation that can also be meaningful to the audience" (Mandala, 2007:66). Even though Spike is not fully part of the core group of friends (the *Scooby Gang*), his usage of *y*-suffixes indicates that he is a character who changes his allegiances and becomes somebody the characters and audience accepts to, eventually, be a trustworthy character (2007: 63-65).

---

<sup>18</sup> A term used in *Buffy the Vampire Slayer* to refer to a common season arch ending with the main characters battling evil.

This highlights not only the linguistic creativity through which characters are created, as applicable within the fictolinguistic framework, but also that character development and linguistic change is traceable through linguistic analysis.

Alternatively, it might be argued that incongruity of characters within a series could imply randomness in scriptwriting or a lack of construction of social background of the characters. The fact however that unexpected character choices (including linguistic choices) are notable suggests that characters are indeed written in a stable way and create a certain feeling of familiarity with the audience. It is this kind of familiarity that is sometimes even acknowledged within the writing, as can be seen in an episode of season four of *Buffy the Vampire Slayer*, in which Faith switches bodies with Buffy. The switch itself is not known to the other characters and the audience is only let in on the secret through a telling line uttered by Faith/Buffy (8).

(8) Joyce:            You sure you're okay?  
    Faith/Buffy:    Five-by-five. (*BVS*)

“Five-by-five” is a line that has been repeatedly used by Faith in previous seasons, marking it as uniquely hers. In the same episode this is highlighted by Willow:

(9) Willow:        Don't worry, we're sure to spot Faith first. She's  
                      like this cleavagy slut-bomb walking around “ooh, check me  
                      out, I'm wicked cool, I'm five-by-five”.  
    Tara:            Five-by-five? Five what by five what?  
    Willow:         See, that's the thing. No one knows. (*BVS*)

Linguistic uniqueness here is consciously used as a plot device and indicates how important linguistic means are for the distinction of characters. Language choices, such as the use of a phrase like “five-by-five”, marked y-suffix-adjectives, or certain dialectal

markers by characters, create boundaries by which they can be individualized and/or grouped. These linguistic means help locate a character in reference to the other characters as well as within the narrative itself.

## **2.4 Characterization as sociolinguistic styling**

For the concept of creating characters, I follow Queen's definition of characterization as

linked to achieving individual distinctiveness, on the one hand, and to engaging general ideas of social similarity and social difference on the other. These two facets of characterization highlight why a given character might be simultaneously stereotypical and uniquely individual (2015b:155).

In addition, McKee (2016:40) defines characterization as “a character's total appearance, the sum of all surface traits and behaviors”. This last point is important to add in order to highlight the fact that the characterization process is constant; characters are evolving, and the very nature of television is the continuous revelation of information, not just in terms of story, but also in terms of what is known about the characters.

Audio-visual characterization is a combination of various modes, at the most basic level distinguished between audio (soundtrack, diegetic sound effect, narration, dialogue) and visual (mise-en-scene, scene selection and transition). Before focusing on the linguistic components of characterization, namely the dialogue, I will briefly exemplify how the variety of modes that are part of audio-visual media cooperate in creating characters' personas.

The following Figure shows a still from *Gilmore Girls*' first episode (for a series summary and character introduction see chapter 3). The scene opens to the kitchen of a hotel (with previous scenes including the lobby, as well as exterior shots). A continuous shot of Sookie shows how she navigates through her kitchen with narrowly averting a variety of

accidents (with kitchen staff jumping routinely around her to avoid catching fires, burnt hands, and tumbling pots). Lorelai, who was previously seen attending the welcome desk of the hotel, joins Sookie in the kitchen with the news of her daughter’s admission to a private school (see dialogue in Figure 1).

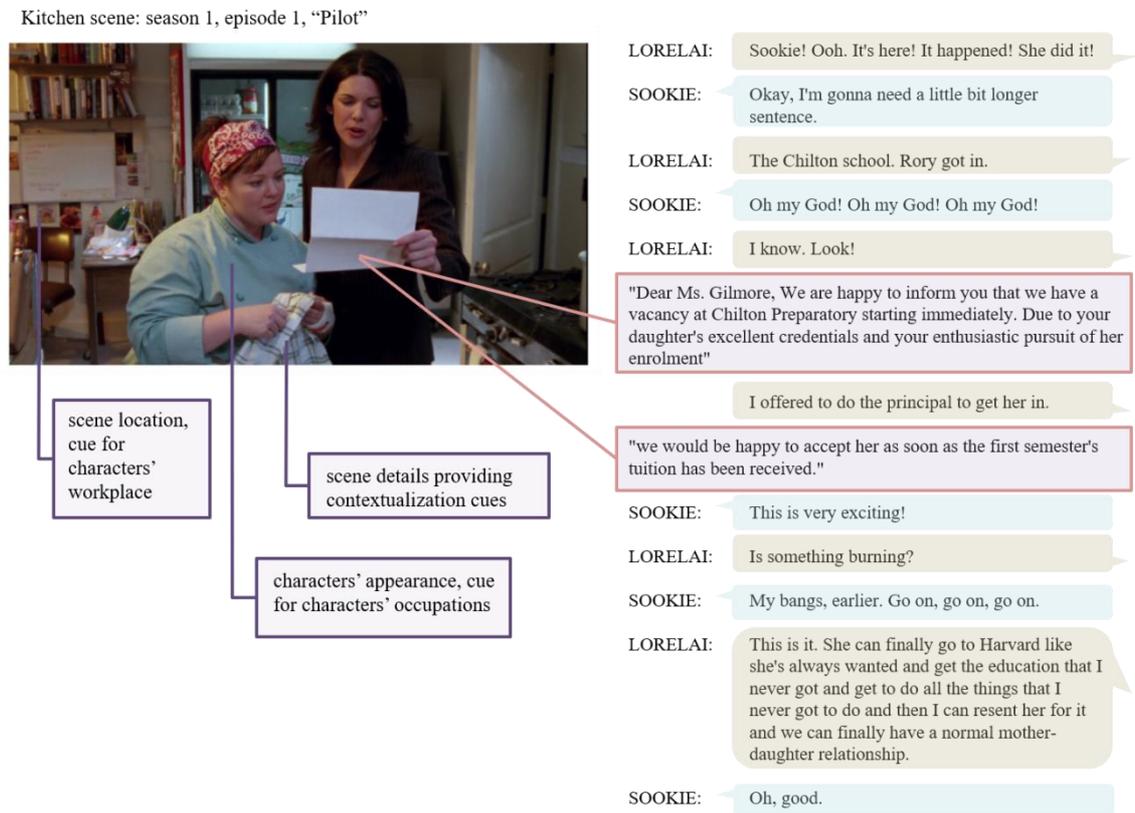


Figure 1: *Gilmore Girls* scene composition

In this scene, the audience is presented with characterizing information of both Sookie and Lorelai. The setting of the hotel kitchen, as previously mentioned and pictured in Figure 1, indicates one of the main locations of the series, as well as the workplace links for the characters. Lorelai is working in managing capacity (further underlined by her business suit), while Sookie is working as a chef in the kitchen (her outfit, the chef jacket, also marks her as “on the job”). Further, Sookie and Lorelai’s friendship that goes beyond work acquaintance is highlighted by the topic of the conversation, the form of the

conversation (informal), as well as their close proximity to each other while they talk, and finally, their hug halfway through the scene.

Other characterization aspects here are for instance Sookie's multiple band aids as well as burnt bangs that point to her comical clumsiness (previously shown through the accident-prone cooking techniques and actual hitting of her sous chef with a pan), Lorelai's wittiness by joking about influencing the headmaster's decision by immoral means, as well as a brief indication of her past by comparing her daughter's opportunities to her own lack thereof.

The construction of the individual characters, and the scene as a whole, clearly comes together through the combinational effort of visual and auditive means. This study focuses on the dialogue specifically as a character-building asset, but it is worth keeping in mind that other modes of course also have an impact on how characters are introduced and maintained. From here on however, characterization will be discussed with dialogue as its main reference.

#### **2.4.1 *Functions of dialogue***

Dialogue, I argue, is one of, if not the most important feature aiding characterization in fictional television series. With it, characters evolve and react to one another, they expose themselves, their actions and emotions, as well as the overall narrative through their speech. Kozloff (2000:33-34) assigns nine functions of dialogue to film, which can be equally applied to television, which are:

- anchorage of the diegesis and characters
- character revelation
- communication of narrative causality
- viewer evaluation and emotions
- exploitations of the resources of language (poetic, humour, irony)
- thematic messages

- authorial commentary
- allegory
- opportunities for ‘star turns’

Of these, I want to pay special attention to character revelation, although it is worth noting that they are all linked and co-dependent. Kozloff’s functions can, to some degree, also be found in McKee’s summary of functions of dialogue, which he categorizes into three main parts: exposition, action, and characterization (2016:22).

Exposition is the level of information necessary for the viewer to follow the actual story and is highly dependent on timing within the fictional setting. The most obvious case of exposition is the backstory provided through the telling by the characters. An example is given in the above discussed scene from the pilot episode of *Gilmore Girls*, in which Lorelai reads out a letter sent to her about her daughter’s approval to start at an elite private school:

(10) Lorelai: I know. Look: ‘Dear Ms. Gilmore, We are happy to inform you that we have a vacancy at Chilton Preparatory starting immediately. Due to your daughter's excellent credentials and your enthusiastic pursuit of her enrolment.’  
(GG)

Through reading the text aloud, the audience instantly learns about the overall story arch (the new school), Lorelai’s daughter’s academic abilities, as well as Lorelai’s own stance towards this change. Using an external source is a convenient device by the writers, as they can present previously unknown information without having the main characters to explain their own background in a seemingly awkward way.

The following example (11) is the opening scene of *Sherlock*’s very first episode and illustrates how an outside source (here, a therapist) inserts information about the main

character in order to provide the necessary backstory that the audience needs in order to understand the overall story.

- (11) Ella: How's your blog going?  
John: Yeah, good. Very good.  
Ella: You haven't written a word, have you?  
John: You just wrote "Still has trust issues".  
Ella: And you read my writing upside down. D'you see what I mean? John, you're a soldier, and it's gonna take you a while to adjust to civilian life, and writing a blog about everything that happens to you will honestly help you. (*SH*)

Ella mentions that John is a soldier who possibly recently returned to 'normal' life and is still struggling with that change. Exposition of this type is particularly noticeable in the beginning of a series where the audience would not know yet how to interpret characters and settings without guidance. Further, too much exposition is seen as against the dogma of audio-visual media: "The axiom 'show, don't tell' warns against dialogue that substitutes passive explanations for dynamic dramatization" (McKee, 2016:24). In *Sherlock's* case for instance, reminders of John's past are rarely again mentioned as pointedly; rather, the series employs flashback scenes that show John while deployed, or in conversation with the fact being subtly indicated. Action is a function of dialogue that is a direct result from exposition and, simply put, dialogue action is what moves the story forward. In the above scene it is Sookie's excited exclamation of "Go on, go on, go on" that calls for more exposition, which in turn leads to a furthering of the story itself. Other examples of action in dialogue are exclamations, arguments, directives, etc.

The final function of dialogue is characterization (or character revelation after Kozloff (2000:33-34)), which in itself is further separated into three functions:

- to intrigue
- to convince
- to individualize

For the present study, I want to focus mainly on the last function. Before going into more detail on the individualization aspect however, I will provide a brief definition for the other two functions here.

The first function refers to the excitement a character can spark within their respective fictional context. Characters ideally provide some sort of fascination within the audience to encourage continued watching. In terms of characterization that means that the audience will want to figure out who a specific character really is (and another reason why exposition is often kept to a minimum). Intriguing characters are oftentimes characters that subvert the audience's expectations, almost challenging viewers to solve a puzzle. In *Buffy the Vampire Slayer* for instance, the main character is introduced as a young woman who, following common stereotypes of the genre, should be the "damsel in distress" type. Instead, she teases the monsters before defeating them without any help, prompting the audience to want to know why that is, who she is.

An intriguing character however can only be as intriguing as they are convincing. The audience has to believe that the character they are watching is real and consistent within the possible world they inhabit.

McKee (2016:41), in talking to scriptwriters, acknowledges the importance of making the audience believe that the character is real, even just for the moment of watching:

If your reader/audience thinks the thought, ‘I don’t believe a word she says’ because they sense your character is a liar, that could be a revelation of true character. But if they think the same thought because they don’t simply believe in your character, then it’s time for a rewrite. (McKee, 2016:41)

Convincing characters are likely characters, meaning that the audience can imagine that a person in that situation would or could talk like that. Convincingness of a character through their dialogue is somewhat relevant for this study. If a character uses language that is unexpected for them to use, it will ultimately also mean that they are less convincing in their persona. The question here is what language features are deemed expected or unexpected and for which characters.

This leads into the final function of characterization in dialogue: individualization. In order for dialogue to contribute to the creation of characters, these characters have to have distinct voices. Field (2003:58) claims dialogue in audio-visual media is “the heart and soul and nervous system” of a story, emphasizing the importance of a character’s particular voice for the success in storytelling on screen. It is surprising then, that very little is said of *how* dialogue is constructed, or how it can meaningfully add complexity to a character’s individuality. Indeed, Field’s guide on screenwriting provides little actual guidance, despite acknowledging that “dialogue is one of the most striking qualities about your [the reader’s] character” (2003:81). This appears to be a general trend with scriptwriters and those teaching the writing of characters. There is a consensus about what dialogue should be like in order to create individualized characters, but when prompted, writers are curiously ambiguous about the *how*, as seen in the excerpts below.

(12) On the distinct dialogue in *Buffy the Vampire Slayer*:

Jane: Each of our characters has their own distinct voice (...) and so you got a feel for what Xander could say versus what Faith could say.

Douglas: They're very specific and they all have their very specific world view.

Jane: It's really delightful, it's really part of making a full and textured world -is when you *really* start getting an ear for who would say exactly what.

Douglas: I once wrote a line for Xander and at the last minute said 'Well, Willow doesn't have much to do this scene, I'll give this line to Willow'... and Joss [showrunner] immediately said 'That's a Xander line, that's not a Willow line, you switched it, didn't you', I said 'Yes, I did'.

(Jane Espenson and Douglas Petrie, DVD *Buffy* season 3 special features)

(13) In response to the question of how distinct character voices are created on *Gilmore Girls*<sup>19</sup>:

I'm a writer, man, that's what I gotta do. (...) It's, you know, I don't know. You gotta hear them in your head. You gotta, like, hear who the person is and know, oh that person's gotta say it that way and then if they sound too much alike then you're like 'Well, that's not gonna work'. If anything could be given to somebody else, if any joke of Lorelai I could've handed to Rory or to Emily – it's not a good joke. Because it's gotta only work for Lorelai.

(Amy Sherman-Palladino, creator and showrunner for *Gilmore Girls* at the ATX *Gilmore Girls* reunion, 2016)

---

<sup>19</sup> The question was provided by me and presented to the two show-runners and the main cast members of *Gilmore Girls* at the reunion event. The above quote is an excerpt from a longer answer.

(14) On the creation of character for script writers:

The problem of creating real people in real situations can be so varied and so challenging that trying to define them is like trying to capture infinity in a glass. (...) But if you know your character well enough, if you feel comfortable inside his or her skin, the dialogue will be individual and appropriate and capture the 'essence' of that character. (Field, 2003: 64-81)

Examples (12) and (13) describe situations where the scriptwriters highlight that any dialogue line is styled for a particular character. The writer has to have 'a feel', 'an ear', 'hear them in their heads' to construct a well-rounded and distinct character. Field (2003), in his guide to screenwriting, alludes to this vague knowledge in example (14) by implying that if the writer is familiar enough with the character they want to create, the linguistic individualization will happen automatically.

Field's analogy of 'essence' of character (2003:81) mirrors McKee's idea of individualization:

A well-imagined, well-researched characterization creates a unique combination of biology, upbringing, physicality, mentality, emotionality, education, experience, attitudes, values, tastes, and every possible nuance of cultural influence that has given the character her individuality (2016: 41).

In terms of linguistic contribution to individuality in characterization, McKee (2016:153) attempts to attach characterizing functions to the syntax of dialogue:

Character-specific locutions depend on both sides of a sentence, subject and predicate. Subject (what or whom the sentence is about) and predicate (something about the subject) combine to create a line of dialogue that helps express two primary dimensions of characterization: knowledge and personality.

(...) A character's knowledge tends to be expressed in the names of things, nouns and verbs, while a character's personality tends to be expressed in the modifiers that color those nouns and verbs. (McKee, 2016:153-154)

This rather simplified guideline does not fully capture the versatility of the dialogue we find in fictional television series and, again, is kept rather vague in any actual explanation of how individualized dialogue attaches to particular characters. The following two subsections introduce approaches to fictional dialogue (the first on literary texts, the second on audio-visual texts) in more detail. Culpeper (2001) and Bednarek (2011a) discuss linguistic patterns that can be observed in characterization processes. In doing that, both studies inform my own analysis significantly.

#### **2.4.2 *Culpeper's textual cues in characterization***

Culpeper proposes a model of how characters are linguistically constructed within a fictional environment and discusses the so-called "surface structure" which "include(s) the particular linguistic choices attributed to characters" (2001:38). Culpeper says that reading the text of a play entails three crucial steps of forming characterizations.

- a) Constructing representation (of character-contexts) for all relevant characters
- b) Constructing a representation of the situation the characters appear in
- c) Constructing a representation of what the writer of the text intends us to understand by the character discourse (2001: 38)

Characterization, according to Culpeper, arises through three types of cues, each consisting of various individual functional means: explicit cues, implicit cues and authorial cues. With the help of examples from television dialogue I will introduce each of these cues, as well as illustrate their applicability to the medium of scripted language for television series.

*Explicit cues*

Explicit cues are cues that inform the reader (or audience) about a character through statements directly from the characters themselves, as in example (15) from *Gilmore Girls*:

*Girls*:

- (15) Emily: It's like a canoe.  
Lorelai: What's like a canoe?  
Emily: Life.  
Lorelai: Okay.  
Emily: You're just paddling along in a canoe.  
Lorelai: Mother, have you ever been in a canoe?  
Emily: Lorelai.  
Lorelai: Well I just can't picture you in a canoe.  
Emily: Your father and I have been paddling a canoe together for years. Only now, he's dropped the paddle.  
Lorelai: Ahh!  
Emily: He just dropped it. Not only that, but now the canoe is going in circles.  
Lorelai: Ah!  
Emily: Without your father there, I'm paddling on my side and the canoe is spinning in circles, and the harder I paddle, the faster it spins, and it's hard work, and I'm getting tired.  
Lorelai: Dizzy, I would think.  
Emily: You are in a kayak. You know how to do all of this.  
Lorelai: How does that put me in a kayak?  
Emily: Kayaks have paddles with things on both ends. You steer it by yourself.  
Lorelai: Mom, you know how to do things by yourself. You are totally capable.  
Emily: Sure, I went to Smith, and I was a history major, but I never had any plans to be an historian. I was always going to be a wife. I mean, the way I saw it, a woman's

job was to run a home, organize the social life of a family, and bolster her husband while he earned a living. It was a good system, and it was working very well all these years. Only when your husband isn't there because he's watching television in a dressing gown, you realize how dependent you are. I didn't even know I owned windmills.

Lorelai: Mom, now you know, and you know how to right-click.

Emily: But you. You provide for yourself. You're not dependent on anyone.

Lorelai: Hmm.

Emily: You're independent.

Lorelai: I am kayak, hear me roar.

(GG)

Here, Lorelai is giving computer lessons to her mother, who is overwhelmed by the responsibilities of taking over her husband's duties after he suffered a heart attack. She uses metaphors of boats to describe the different life styles she and her daughter have, and explicitly talks about her helplessness and her past ambitions that led her to this moment. The audience follows this exchange and receives a clear picture of both characters through what they say. What we can see in this particular extract are self-presentation, which "occurs when a character [...] provides explicit information about him or herself" (Culpeper, 2001:167) within the canoe metaphor Emily uses, as well as other-presentation, where "a character [...] provides explicit information about someone else" within the kayak metaphor Emily uses to describe her daughter Lorelai's life (Culpeper, 2001:167). Explicit cues are, in function, related to what McKee calls exposition of character (see above).

### *Implicit cues*

Implicit cues, according to Culpeper (2001:172) are verbal or non-verbal pieces of “character information which has to be derived by inference”. He includes a number of linguistic features, as can be seen in the following section below. I will elaborate on implicit cues and their contribution to character-creation later on in this chapter which means a brief outlook of how Culpeper sees these features will suffice at this point.

Example (16) shows that implicit cues (italicised) are varied and can indicate diverse characteristics of the participants in the discourse.

- (16) Rory: *So*, what do you think I should do?  
Lorelai: Oh, *honey*, I think it’s your decision.  
Rory: I know but tell me what you think I should do.  
Lorelai: *Well*, I think you should take a few days, *you know*, let the shock wear off.  
Rory: You're really not gonna give me your opinion?  
Lorelai: Only you know what you want.  
Rory: Yeah. I love him. I do. *I mean*, things have been really amazing lately. But, on the other hand, we are *so* young. I'm *only* 22. On the other hand, what does age matter when you're in love? On the other hand, what is the rush?  
Lorelai: *Well*, you're like a circus freak with all the hands.  
Rory: Won't you just tell me what to do?  
Lorelai: Honey, I'm sorry.  
(GG)

In this scene Lorelai talks about the marriage proposal her daughter Rory received in a previous scene. Rory is not certain of how she should react and asks her mother for help. Some noticeable implicit cues for their emotional involvement here are Lorelai’s repeated use of *honey* to address her daughter, showing empathy, her use of hedging devices *well*

and *you know* that show her reluctance to commit to an opinion on her daughter's decision, or Rory's uncertainty in this matter with an overuse of *on the other hand*, questioning her reasons (which she tries to justify with intensification and emphatics) repeatedly.

### *Authorial cues*

Finally, authorial cues are features that “do not arise directly from the character concerned” (Culpeper, 2001:229), but from the author of the text or, in the present study, the scriptwriters. Naming characters gives (script-) writers opportunity to attach personality traits and associations to a character even before any dialogue is used. Janet Brennan Croft, discussing naming in *Buffy the Vampire Slayer*, notes that “a personal name is a nexus for many deeply important concepts and feelings about being a person and having a place in the world in relation to other people and to a family” (2014:1).

The strong bond between characters and their names can be seen in *Buffy the Vampire Slayer* and its spin-off *Angel*, with Angel representing a rueful vampire after regaining his soul and human consciousness. When he loses his soul again in a moment of pure joy, his personality then changes and he becomes Angelus, a demon without any remorse or qualms. Henceforth this distinction of names for the same character is deliberately used as a cue for the audience to differentiate the character's alliances and adjust their evaluation while viewing. Names can further indicate a character's national background or heritage, as in *The Big Bang Theory*, where the Indian character is clearly identified as such by his name Rajesh Ramayan "Raj" Koothrappali or again, in *Buffy the Vampire Slayer*, Willow carries a surname associated with being Jewish: Rosenberg.

As well as naming, Culpeper (2001) also mentions stage directions. In plays, these are of course included in the actual text, while in television series the audience can only infer stage directions from what is being acted out on screen.

Culpeper's cues are useful anchor points with which dialogue can be analysed. His study focuses on the character's impression "in the reader's head" (2001:1) and is more concerned with the cognitive approach to characterization, "rather than with character – the output of that process" (2001:1). The latter is seized by Bednarek (2010, 2011a), who uses Culpeper's model to investigate the fictional character itself with what she calls expressive character identity. Her approach is most relevant to the present study.

### **2.4.3 *Bednarek's expressive character cues***

Taking Culpeper's theory on characterization cues in literary plays and applying it to fictional television, Bednarek (2010) outlines a list of expressive features that are likely to indicate character identities. Furthermore, she proposes a general set of characteristics for scripted language within fictional television (2010:65-66):

Television discourse needs to be comprehensible to the audience (avoiding unintelligible and vague language); entertain the audience (including emotional and aesthetic language: avoiding repetition, long monologues or narratives); create characters that the audience finds realistic (featuring informal language); and attract a large audience.

This general list repeats some points made earlier, particularly with reference to two functions of characterization: intrigue and convince. The third function, individualization, is addressed in the following.

Table (1) by Bednarek (2011a) summarizes Culpeper's characterization cues in her work that adapts his theory to fictional television series:

Table 1: Cues in characterization (Bednarek (2011a) after Culpeper (2001))

<b>Explicit cues</b>	<b>Self-presentation (character gives explicit information about self)</b>
	Other-presentation (character gives explicit information about other characters)
<b>Implicit cues</b>	Conversational structure (e.g. turn-length, turn-taking, turn, allocation, topic shift, topic control, incomplete turns/hesitations, interruptions)
	(Non)adherence to conversational maxims, conversational implicature
	Lexis (Germanic vs. Latinate, lexical richness/diversity, surge features/affective language, terms of address, keywords)
	Syntactic Structure
	Accent and dialect
	Verse and prose
	Paralinguistic features (e.g. tempo, pitch range/variation, loudness, voice quality)
	Visual features: kinesic features and appearance (e.g. stature, clothing, facial expression, posture)
	Context: a character's company and setting
	(Im)politeness strategies
<b>Authorial cues</b>	Proper names
	Stage directions

Bednarek combines Culpeper's theory with the medium of fictional television and compiles a list of "expressive textual cues" (2011a:8). She claims that expressivity within scripted dialogue carries "aspects such as ideology, emotionality, values and evaluations" (2011a:8); in other words, aspects that can contribute to the construction of a character.

In the present study, I will analyse a set of expressive linguistic features that can indicate (individual or group) identity patterns. This will allow me to analyse certain features that are more typical with some (or just one) characters, and those indicating group-

membership through shared usage. Furthermore, I will analyse the interaction of features, as well as expressive styles that form distinct character identities as they have been created by scriptwriters (from a semiotic perspective) and are perceived by the audience (from a cognitive perspective) (Bednarek 2010: 124–125).

#### **2.4.4 *Characterization through stance-taking***

Throughout this chapter, I have presented varying approaches to scripted characterization. The turn towards styling and stylization in identity-building within sociolinguistics, as well as the increased focus on characters as analysable entities within fiction in stylistics, come together in the present study. Both developments are interlocked with the broad concept of variation itself and any social effects of language variation. My research questions address these foci and aim at finding variation that correlates with characters and character types.

As mentioned above, linguistic means of characterization are versatile: textual cues (following Culpeper, 2001) include linguistic markers of identification in explicit, implicit, and meta or authorial ways. Bednarek categorizes these functional features as parts of expressive character identity which encompass “‘emotional identity’, ‘attitudinal identity’, ‘ideological identity’, etc.” (2010:118) and are “used in a broad sense to include emotions, evaluations, attitudes, values and ideologies” (2010:119).

Put differently, linguistic features that serve as identifying functions in dialogue, are often markers of stance. Stance taking, writes Bucholtz (2009:147), “may come to be ideologically tied to larger social categories” and stance in general, she says, is linked closely to notions of style and identity. Because this study is an investigation of identity-construction (characterization) and styling (scripting dialogue), the analysis strategically focuses on markers of stance.

Johnstone defines stance as “methods, linguistic or other, by which interactants create and signal relationships with the propositions they utter and with the people they interact with” (2009:31). Keeping the agency of stance taking in mind, it is likely that speakers link particular linguistic expressions that serve their stance to more general social statuses. Ochs (1992) in particular investigates how stance taking intertwines with indexicality and the emergence of indexed speaker styles.

In the context of fictional television dialogue, these indexed speaker styles are comparable to character types and it can thus be assumed that stance might play a key role in creating fictional identities. Johnstone writes that “repeated sets of stance taking moves can emerge as relatively stabilized repertoires, sometimes called ‘styles’, associated with situations or social identities” (2009:33), which aligns with functions of dialogue discussed earlier in this chapter. Characters provide repeated linguistic patterns so as to create a sense of familiarity and recognizable individuality for the audience. Sociolinguistic studies have explored the notion of stance taking and style most prominently in Eckert’s study in a Detroit high school (1989, 2000). Social identities (here, identities of ‘jocks’ and ‘burnouts’) are formed through speakers’ adherence to particular patterns in style taking. The framing (and continuous reframing) of stance as style creates indexical meanings that contribute not only to the speakers’ agentive partaking within social groups, but also the listeners’ recognition of these groups. In the context of television language, that means that stance taking by the characters creates social identities that are recognizable to the audience. Landert (2017:489) highlights the importance of stance for fictional dialogue and points to stance expressions’ ability to “provide a resource for characterisation and character alignment”, both in implicit and explicit ways.

Studies generally distinguish between three different types of stance: evaluative stance describes linguistic forms of assessing value, affective stance includes emotionality, and epistemic stance “deals with the degree of certainty or reliability of information” (Landert, 2017:490).

This study analyses epistemic stance as linguistic expressions that are used to indicate a character’s commitment (to themselves, to what they say, to others). In order to find out how stance taking links to identity construction (cf. Bucholtz and Hall 2005), I will focus on five different expression groups.

- pragmatic markers (*you know, I mean, like*)
- hedges (*sort of, kind of*)
- modal adverbs (*maybe, probably, perhaps, possibly*)
- general extenders (*and something, or anything, and stuff*)
- intensifiers (*very, really, so, etc.*)

The first four features are more or less associated with lessened commitment in discourse, be that through actual linguistic function or public assumption. Intensifiers, as the final group, are features that increase propositional (and emotional) commitment.

All feature groups will be introduced and contextualized in current research in the respective sub sections of chapter 5. Before turning to the analysis, the following two chapters will introduce the data (the Television Dialogue Corpus), how it was collected and what methods inform the analysis.

## **3 The Television Dialogue Corpus**

### **3.1 Corpus set-up**

The corpus used for this study was constructed with the research questions in mind. A detailed character-based analysis necessitated that enough data was provided for each speaker, but also that enough speakers were included so that they could be compared with each other and across social categories.

The series I included had a running time of at least three years so as to provide enough context in terms of individual character word counts, as well as characterization background. The analysis focuses on established linguistic characterization patterns, so I refrained from including all characters of each series. Instead, only characters in reoccurring roles and enough screen time as well as word counts were included. What ‘enough’ means here is dependent on whether the respective series is based on an ensemble or core set of characters and will be further explored in the individual series subsections.

I chose six relatively recent television programmes, incorporating a time-span of close to two decades with broadcast dates from 1997 up to 2014 (some series continued after that cut-off point). The series can be separated into three overarching genre descriptions, with two series per genre, as seen in Table 2 below. Further links across the series, as well as plot summaries and character introductions will be given in the individual series sections in chapter 3.

Table 2: Series overview

Genre	Series	Total broadcast	seasons	episodes
fantasy drama	<i>Buffy the Vampire Slayer</i>	WB 1997-2001, UPN 2001-2003	7	144
	<i>Angel</i>	WB 1999-2004	5	111
family dramedy	<i>Gilmore Girls</i>	WB 2000-2007	7	154
	<i>Parenthood</i>	NBC 2010-2015	4	68
British drama	<i>Sherlock</i>	BBC 2010- 2017	3	9 <sup>20</sup>
	<i>Torchwood</i>	BBC 2006-2011	3 <sup>21</sup>	31

Practical restrictions on the choice of series for the corpus included the availability as well as the quality of dialogue transcripts. It was too time consuming to do the transcriptions myself, given that the aim of the study was to allow for a broad-stroke quantitative analysis of a number of characters. Options for the linguistic investigation of dialogue range from obtaining the original scripts, to using subtitling or closed captions, and finally fan-made transcriptions. For this project I decided on the latter and will highlight the reasoning behind this choice, as well as some arguments against using scripts or subtitles, below.

Many popular television series (and films) enjoy a lively fan community that is engaged with the content on screen and eager to discuss past, future or non-canonical storylines. A fortuitous side effect of that engagement is the transcription of whole episodes by fans and subsequent sharing of these transcriptions on dedicated online websites. Because the

---

<sup>20</sup> Leading up to season 3, BBC released a 7 minute teaser trailer. This is not included in the episode count.

<sup>21</sup> Does not include season 4 ('Miracle Day'), which is set in the US and underwent a change in production both in terms of production company and writers. While it would be undoubtedly insightful to trace in how far linguistic characterization might have changed due to such drastic changes, the amount of dialogue was not enough to provide any meaningful results. By excluding this season, I can account for a consistent production and writing background for the series.

personal investment into the series and the characters is relatively high, fan communities provide a level of quality control. Transcripts as found online do not follow official script conventions as such but are often structured in comparable ways. This method of data gathering, comparable to crowdsourcing, is relatively common for research into television dialogue, the reasons being high accuracy of dialogue representation, easy access, and wealth of data. Studies that used fan transcriptions include Bednarek (2010; 2011b); Piazza et al. (2011a); Quaglio (2009), Tagliamonte and Roberts (2005), and Reichelt and Durham (2017).

Other options for the collection of dialogue from the series are briefly considered here.

### **Original scripts:**

The scripts by the writers and/or producers of a series are arguably what comes closest to the intended construction of a character. However, there are several caveats with this: scripts are difficult to obtain, even years after broadcast. It is near impossible to receive the rights for the original scripts of a complete series (which can be thousands of pages) as the copyright usually remains with the production studios. Furthermore, scripts do not consistently reflect what the audience sees or rather, hears, on screen. Ad-hoc changes to the dialogue while filming, scenes that are cut in post-production, as well as improvisation by the actors remain unaccounted for in the original scripts and would have to be adjusted manually. With some series totalling a hundred and more hour-long episodes, this is a sheer impossibility.

In addition to that, the formatting conventions of scripts (see below) make any manipulation of the dialogue into corpus-able data inconvenient in terms of time and very complex in terms of methods.

We zip through a rush of people and land on LORELAI GILMORE sitting on the steps of the gazebo, two giant cups of take-out coffee in her hands and a paper bag on her lap. She hears a familiar voice:

RORY (O.C.)

Hey!

Lorelai looks and we PAN UP to find RORY GILMORE standing there, smiling and gorgeous.

LORELAI

(annoyed)

That's how you look when you get off a plane?

RORY

That's how you say hello?

LORELAI

(getting up)

You've been stuffed in a glorified tin can for seven hours surrounded by people with consumption and diphtheria, scabies, hummus dips, rabid dogs and drugged up children attacking your seat and stealing your change.

RORY

What airline are you flying?

LORELAI

You should look drawn and blotchy. You should be singing "I Dreamed a Dream" with a bad haircut while selling yourself to French dockworkers. Instead, you look perfect. Admit it. You've been Goop'd.

RORY

I have not been Goop'd.

LORELAI

You do yoga in the aisles wearing cashmere sweatpants while your comfort dog watches "Zoolander" on his watch.

Figure 2: Script extract from *Gilmore Girls* revival (Netflix, 2016) as printed in (Nussbaum, 2016)

A reason against using the original scripts is also the idea that the characters are constructed through the combined efforts of a production team, not just the writer. Omitting any input by actors, directors, or producers would thus not be a true reflection of the characters the audience gets to know on screen.

### Subtitles/ closed captions (CC):

Subtitles or closed captions are optional audience aids that often accompany series episodes on DVD releases, as well as on online streaming services such as Netflix. They can represent spoken words (see Figure 3), as well as a more detailed representation of other sounds included in the screening (see Figure 4).



Figure 3: Subtitles in Gilmore Girls ("The Deer Hunters", season 1 episode 4)



Figure 4: Sound descriptor *Sherlock* ("The Hounds of Baskerville", season 2 episode 2)

While subtitles seem to be a more convenient option than scripts and offer higher chances of consistent availability, conventions within the genre are restrictive. Figure 3 shows a scene from *Gilmore Girls* in which Luke and Rory are talking to each other. The way the subtitles are formatted on the screen makes it easy for the audience to gauge who is saying what and when. The problem here is that because the screen can only contain a certain number of words and lines, and the subtitle frame has to stay up for long enough so that the audience can read, scenes that contain a higher amount of dialogue will be impossible to fully caption. The dialogue of the scene above actually reads thus:

(17)            Luke:        You look like you need pie.  
                  Rory:        I do? (*GG*)

Luke's line is shortened to accommodate the subtitling conventions. For the purposes of capturing what is said on screen this is a perfectly reasonable adjustment, but for a study investigating the detailed dialogue and how it is creating linguistic patterns, this is obviously less than ideal.

Automatic subtitling, as increasingly explored in natural language processing, is a possible option for future research. However, transcriptions are still inconsistent and not representative for any sort of detailed linguistic analysis. The above scene, as captioned by YouTube, retains Luke's whole line, but omits Rory's response altogether. As the dialogue continues the subtitles mistake 'towel' for 'trout', 'he' for 'you', and 'addy' for 'a D', all within a scene less than three minutes long.

As mentioned above, I decided to use fan transcriptions for the corpus construction. With the data compiled by unknown sources online, I followed a strict protocol before including a series' transcription to the corpus:

1. No lone or isolated transcription

As I included a series in the main corpus, I paid attention to the source website and only included an episode if **all** episodes from that series can be found on that same site. I found that website hosts that offer whole transcript catalogues will also be more careful about the quality of the transcripts and ensure that the format is consistent throughout. This is important for the data manipulation, as well as accuracy tests.

2. As few transcribers as possible

There are no consistent formatting frameworks available for each fan who is transcribing a television series and certain conventions will differ across people. For that reason, I tried to keep the number of transcribers for each individual series as low as possible.

3. Spelling errors exclude the transcription

During the search for transcriptions, as well as during data mining and compiling of the corpus, I read through the transcripts to see whether correct orthography is used. Most transcriptions are proof-read and were in a form that is arguably similar to what I could have provided through transcription myself.<sup>22</sup>

4. Accuracy errors exclude the transcription

Similar to the spelling, I also double checked for accuracy of what is said on screen. Because I had to read the transcripts while listening to the episodes, this was the most time-intensive step.

---

<sup>22</sup> For one of the series however I had to start a new search for transcriptions and discard all collected transcripts as there were too many errors in the scripts.

For each included series I decided on a number of episodes (at least three for the shorter series, more depending on how many episodes there are in total per series) that I watched while reading through the script. This was not only to test the validity of the transcription itself, but also to note important conventions by the individual transcribers. This included the handling of filled pauses (if they were included, how were they represented), emphatics (vowel lengthening might be indicated by repetition of a letter which can differ), or shortening (was *you know* also represented as *y'know* for example).

Only if all episodes of a series were accurately transcribed and the transcription was consistent throughout, was a series included in the television corpus.

### **3.2 Series included in the TV corpus**

This section introduces each of the series included in the Television Dialogue Corpus. The analysis focuses on dialogue patterns of individual characters, which meant that for each of the series I also had to decide which of the characters it was best to include. The main requirement was that the character had to be meaningful and established for the fictional setting, as it can be assumed that minor characters will not be styled with the same scrutiny (or will be limited to rather stereotypical one-dimensional portrayals).

For ensemble cast series (where characters are similarly important to the overall story), I decided to include all the main characters that contributed sufficient word counts overall (the minimum was set at 5000 words). In series that had clear core characters with overwhelmingly more speaking time than the rest of the cast, I made sure that the characters included had at least ten percent of the main character's total word count to allow for comparability. All characters that make up the final corpus are vital to the individual stories and can be assumed to be characterized as unique personas.

### 3.2.1 *Buffy the Vampire Slayer and Angel*

*Buffy the Vampire Slayer* (henceforth *Buffy*) was a popular US fantasy series predominantly aimed at younger audiences. IMDb provides the following synopsis for the series:

A young woman is forced to fulfill her destiny of fighting vampires and demons with the help of her friends all the while struggling to live a normal teenage life of heart break and drama. (Buffy, 2017)

*Angel* is a spin-off series that was added to the *Buffy* universe after the latter's third season. Targeting a slightly older audience, the themes of the series moved from issues of teenage life to topics including addiction, corporate power struggles, and the vague duality of good and evil:

The vampire Angel, cursed with a soul, moves to Los Angeles and aids people with supernatural-related problems while questing for his own redemption. (Angel, 2017)

The inclusion of this series in the study is due to the reappearance of three of the characters that were introduced in *Buffy* (Angel, Cordelia and Spike, see information below), which enabled me to analyse their characteristic speech patterns over a prolonged period of time. As both television series are written by the same team, as well as set in the same universe (*Buffyverse*, as it is called by creators and fans), it is possible to analyse these characters over the course of two series.

*Buffy* is set in the fictional Californian city of Sunnydale and revolves around Buffy Summers and her destiny to fight vampires. Starting when she is 16 years old in season one, the audience is introduced to an outgoing young woman who, in the first episode,

appears to fulfil stereotypes of the supposedly clueless Valley Girl that needs to be rescued. Defying expectation, she is portrayed as quick-witted and independent, fighting her own battles both in the metaphorical and literal sense. After graduating high school, Buffy attends the local college for just over a year (season four and beginning of season five), before dropping out in order to care for her sick mother and sister. Season five introduces darker themes with the death of Buffy's mother, Buffy's own sacrificial death, and depression after being resurrected against her will. Season seven ends with a final battle that destroys Sunnydale.

Xander and Willow become her friends early on in the series and, aware of the fantastic elements in Buffy's life, remain by her side for the rest of the series' seven seasons. Willow's character arc throughout the series revolves around introspection and finding herself. Early portrayals (up until season four) focus on her insecurity around other people and within unknown situations. The later seasons introduce her romantic relationship with fellow witch Tara (not included in the corpus), as well as addiction and issues of loss, first with Buffy's death, then when Tara is shot. Her story ends in the final episode where she uses the magic that she previously struggled with and aids in Buffy's final battle by becoming a goddess.

Xander is introduced as Willow's counterpart: where she is introvert and nervous, he is shown to be outspoken and used as comic relief. After graduating high school, he works as a contractor and is often seen as a needed human influence on the otherwise supernatural ensemble.

Giles, the high school's librarian and watcher<sup>23</sup> of Buffy, is older than the core characters and often provides a voice of reason and maturity. When the high school is destroyed at Buffy's graduation ceremony, he takes up work as the owner of a magic shop and much of his continuing storyline is motivated by his conflicting feelings towards his responsibilities in Buffy's life as her watcher and father figure.

These four characters can be described as the core group of the series, acknowledged within the fictional world as the Scooby Gang, a reference to the four friends solving mysteries in *Scooby Doo*. The plot below (Figure 5<sup>24</sup>) illustrates how the core characters relate to other characters from the series.

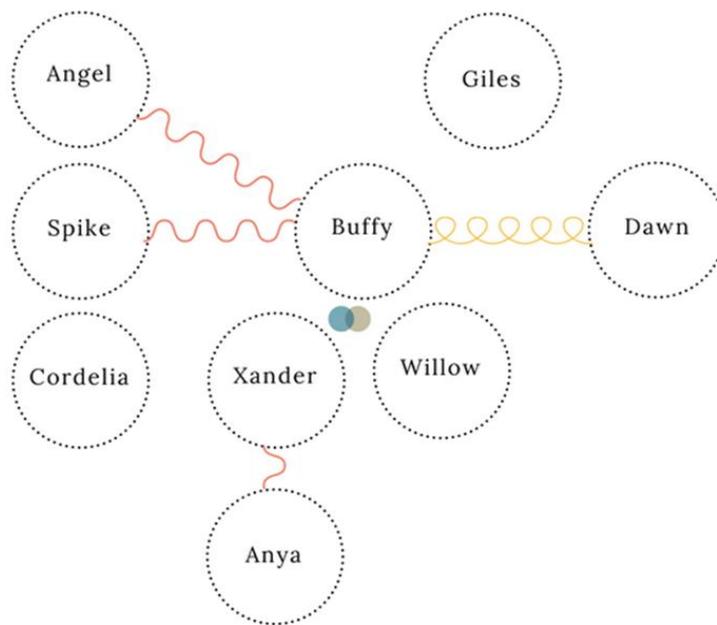


Figure 5: Sociogram: *Buffy the Vampire Slayer & Angel*

<sup>23</sup> A sort of supervisor for her slayer abilities and progression

<sup>24</sup> Red lines between characters indicate (previous) romantic relationships, yellow lines indicate siblings. A small Venn diagram between characters (here, Buffy, Xander, and Willow) indicates a closed-knit friendship.

Anya, who for most of the series is Xander's main love interest, is a former demon who joins the Scooby Gang in season three. Dawn joins the ensemble in season five as Buffy's 'new' 14-year-old sister<sup>25</sup>. Both Anya and Dawn rely on one major link to the core group of *Buffy* (Xander and Buffy respectively) which makes their status within the ensemble somewhat less stable. Linguistically, this was found to be portrayed through their lessened use of core-group-specific markers (adjective suffix -y) in Mandala (2007).

Cordelia and Angel were *Buffy*-regulars for the first three seasons before both moving to the spin-off *Angel*. Cordelia attended high-school with Buffy, Willow, and Xander and was initially portrayed as Buffy's antagonist and a stereotype of the popular cheerleader. In *Angel* she was part of the core character group and appeared for the whole five seasons as part of Angel's supernatural detective agency. Angel is a vampire with a soul and Buffy's main love interest in the first three seasons. After leaving Sunnydale he continues to fight against evil forces while seeking redemption for his wrong-doings during his soul-less life. Lastly, Spike first appeared in *Buffy* in season two and joined the ensemble of *Angel* after *Buffy* officially ended (season five for *Angel*). For most of the series his status is ambiguous between antagonist, reluctant friend, and love interest of Buffy. As a vampire, much of his characterization is tied to stereotypes of rebellion and British-inspired punk culture, catering to his disregard of the status quo which is linked to Buffy and Giles.

Academically, the show has received considerable attention. With publications in gender studies (e.g. Helford, 2002), sociology (e.g. Erickson, 2002), or philosophy, their own

---

<sup>25</sup> She is not part of the first four seasons which is explained as her actually being energy that was transformed into human form. Fake memories established her as a rounded character despite the oddity of the late inclusion of such a core character to the group.

journal (Slayage – the Joss Whedon Studies Association), and a biennial conference, the series is regarded as highly influential and seminal in many ways.

In regards to linguistic research, previous work relevant to the study at hand is focusing on in-group membership (Mandala 2007), rhetorical strategies (Masson 2006), linguistic markedness (Kirchner 2006 and Adams 2003), and code switching (Ruddell 2006), as mentioned in previous sections. The study on characterization patterns through variation of intensifiers (Reichelt and Durham, 2017) can be regarded as the origin for the present research.

### 3.2.2 *Parenthood*

*Parenthood* is a dramedy that predominantly revolves around family dynamics.

*Parenthood* had a total running time of six seasons with the first four included in the TV dialogue corpus (due to transcript collection overlapping with the final broadcasts, the final two seasons could unfortunately not be incorporated). The below plot (Figure 6) summarizes the relationships of the ten characters that I included in the final corpus:

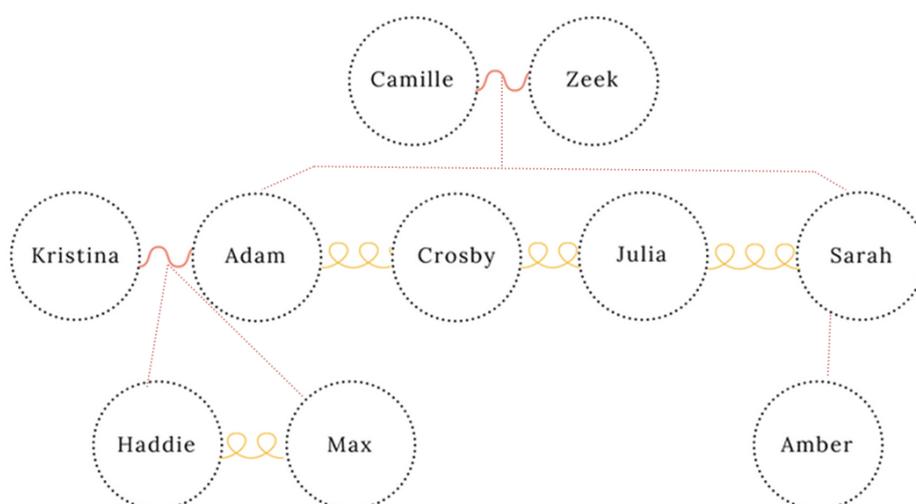


Figure 6: Sociogram: *Parenthood*

Zeek and Camille are the heads of the Braverman clan. They are first introduced as the central point to which their children and grandchildren gravitate, e.g. for family dinners and family disputes. They have four children who are each characterized in very distinct ways:

Adam is the most conservative of the Bravermans and he, alongside his wife Kristina and children Haddie and Max, are initially portrayed as the picture perfect American family. Over the course of the seasons, the family struggles to overcome several hardships including economic uncertainties caused by Adam's job loss and Kristina's fight against breast cancer. A continuous storyline for these four, as well as the extended family, is Max's diagnosis with Asperger's syndrome. The eight-year-old boy portrays many stereotypes of this particular type of autism spectrum disorder, including diminished competency in social interactions, understanding of non-literal contexts (metaphors, jokes, etc.) and an inability to control his moods (e.g. temper tantrums).

Julia is Camille and Zeek's youngest daughter. A successful lawyer at the start of the series, she is introduced as struggling to manage her career and her family (husband Joel and daughter Sydney; not part of the corpus) at the same time. In comparison to her siblings, Julia is shown to be more reserved and less emotional, particularly compared to Sarah, her older sister.

Sarah joins the other Bravermans at the start of season one when she moves back into her parents' house, children Amber and Drew (not part of the corpus) in tow. At 38 years old, she struggles with finding her career path, going from bar tending, writing, designing, to photography. Her oldest child, Amber, is Haddie's age and attends high-school at the start of the series.

Finally, Crosby is portrayed as Adam's counterpart: where Adam has control and maturity (steady job at the start of the series, house and family), Crosby lacks stability in life and is often seen as a goof. In the course of the series this juxtaposition shifts slightly as Adam's life unhinges (as described above) and Crosby grows into the role of responsible father, husband, and business owner (his girlfriend/late wife, as well as son Jabbar are not part of the corpus).

### **3.2.3 *Gilmore Girls***

*Gilmore Girls* is a dramedy that depicts the unusual mother-daughter relationship between Lorelai and Rory Gilmore.

Lorelai, 32 at the beginning of the first season, is a strong-willed single mother who left her family's home when Rory was still a baby. Much of her story line is based on self-fulfilment and professional success. In season four she opens her own inn with her best friend, and chef of a hotel, Sookie.

Rory is 16 at the beginning of the series. The first three seasons follow her as she tries to fit in with the student body at a prestigious prep school. Set on her goal to become a foreign correspondent, much of her character plot is focused on her academic achievements which include her attending Yale University in season four and graduating with her first journalism job lined up in the final episode of season seven. While attending school, Rory bonds (although not very closely at the beginning) with Paris, a fellow student. They become roommates in Yale and throughout the series Paris often acts as the overbearing and over-the-top comic relief.

Emily and Richard Gilmore are part of Connecticut's high society and return to Lorelai and Rory's life when the former asks them for financial support. In return, the

grandparents demand weekly dinners with both their daughter and granddaughter. The ensuing recurring conflicts between the generations are a driving point of the series.

Also part of the corpus is Lorelai's main romantic interest, Luke. As the owner of the diner he participates greatly in the lives of both Lorelai and Rory, who seldom cook for themselves.

Fulfilling the popular television trope of the prolonged suggested but unfulfilled relationship<sup>26</sup>, Luke and Lorelai are one of the main romantic relationships in the series which otherwise focuses on family values and how three generations get along with each other.

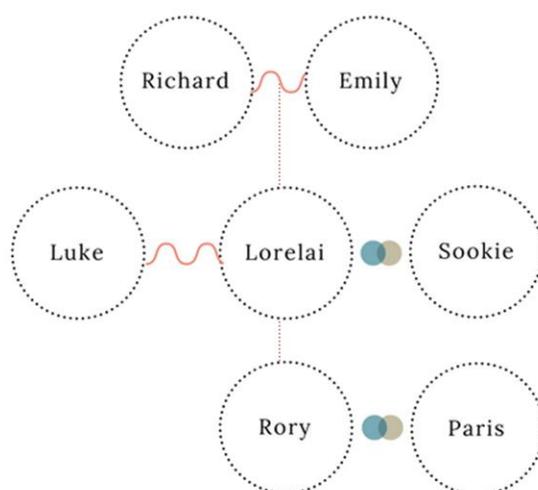


Figure 7: Sociogram: *Gilmore Girls*

Linguistically, *Gilmore Girls* received increased attention for its fast-paced dialogue which is mentioned in *The Gilmore Girls companion* (2014:163) and something lead actor Lauren Graham acknowledges in interviews as well as in the title of her 2016

---

<sup>26</sup> Often referred to as ‘Will they, won’t they’ and known as a popular plot device in television. Other such examples are Ross and Rachel from *Friends*, Mulder and Scully from *The X-Files*, Dawson and Joey (and Pacey and Joey) from *Dawson’s Creek*, Jim and Pam from *The Office* (US), JD and Eliot from *Scrubs*, Max and Fran from *The Nanny*, and many more.

memoir “Talking as Fast as I can” (2016). In fact, while other transcripts of 42-minute episodes included in the corpus consisted of approximately 50 to 60 pages, *Gilmore Girls*’ transcripts amounted to up to 80 pages per episode. This also affected the word counts per character, which I will discuss further down.

Further to the amount of dialogue, the characters’ language was also filled with intertextuality (c.f. Rawlins, 2010). References often reflect individual characters’ backgrounds (musician Lane mostly references obscure band trivia, while Rory and Lorelai have back and forth conversations including movies and literary facts) but also more generally point to the series as “culturally literate” (Ellcessor, 2012:60).

(18) Scene: Rory talks about Paris’ new relationship with a professor

Rory: Yes, what about that? This guy’s risking everything. His job, his reputation

Lorelai: Yes, well, he’ll always have Paris.

(referencing a famous line from the movie *Casablanca*)

(19)

Lorelai: Solar Systems 50% off!

Rory: That’s not just a post-Christmas sale, that’s a post-Pluto sale!

(referencing the revocation of Pluto’s status as a planet in 2006)

(20) Scene: everyone is waiting for Richard to join Friday night dinner

Lorelai: Godot was just here, he said: ‘I ain’t waiting for Richard’, grabbed a roll and left. It’s been forever!

(referencing the play *Waiting for Godot* by Samuel Beckett)

The series has also been the focus of a number of studies by Bednarek, who investigated the *Gilmore Girls*' dialogue in terms of stance on vegetarianism (2010), character stability (2011b), or multimodal creation of a scene (2010).

The series was renewed for an eighth season in 2016 and four more episodes (each at 90 minutes length) were shot under Netflix's production. Because of the late addition however, I was not able to include these episodes into the corpus. Possible future research covering the original next to the new content would undoubtedly shed additional light on questions of character development and (linguistic) change over time.

### **3.2.4 *Torchwood***

*Torchwood* is a science fantasy series and thus thematically similar to *Buffy* and *Angel*, although, as with *Angel*, it is aimed primarily at an adult audience. The corpus includes the first three seasons of the series.

Set in Cardiff, this *Doctor Who* spin-off (and anagram) revolves around a small group of people who investigate supernatural occurrences in the Welsh capital and further afield.

The first season's opening included the following monologue:

Torchwood: outside the government, beyond the police. Tracking down alien life on Earth, arming the human race against the future. The twenty-first century is when everything changes. And you've got to be ready.

The central character is an American, Captain Jack Harkness, who appeared in a number of episodes of *Doctor Who* as a fellow time traveller and thus serves as a point of familiarity for audiences switching between the two programmes.

Each member of the original *Torchwood* team has a specific purpose: Toshiko is the computer expert, Owen the medic, and Ianto a sort of odd-job man. The first season starts

with Gwen, who previously worked for the police in Cardiff, joining the Torchwood Institute team. Her partner, and later husband, Rhys, is unaware of what Torchwood is during the beginning of the series but, by the third season, has joined the group and is, similarly to *Buffy*'s Xander, often used for comic relief.

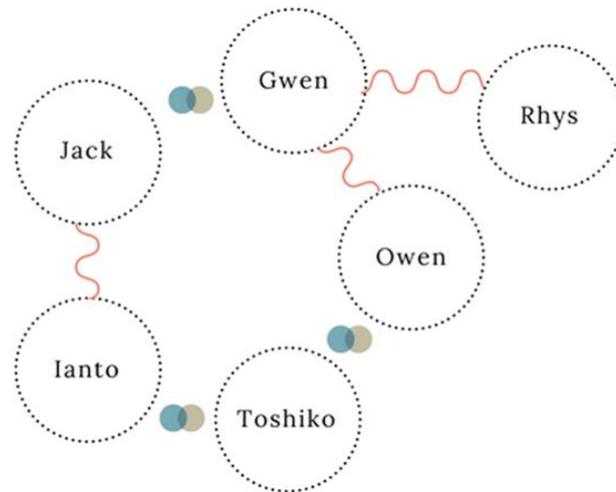


Figure 8: Sociogram: *Torchwood*

The first two seasons mainly consist of self-contained episodes (called ‘Monster of the Week’ in *Buffy* and *Angel*'s case, but is equally applicable here). During the second season Owen dies but remains part of the cast in a resurrected state for the remainder of the season. In the finale he eventually leaves the ensemble, along with Toshiko who also dies.

The third season, based in London, also follows one main story arch where the remaining members of Torchwood are caught in the middle of negotiations between aliens and the government. By the end of it, Ianto is killed and the group disbands.<sup>27</sup>

---

<sup>27</sup> A fourth season is set in the US (in co-production of BBC and Starz) and introduces a largely new cast. As mentioned elsewhere, a change in production location, team, and writers were the main reason not to include this final season in the corpus for the present study.

The series explored a number of relationships across the ensemble cast (as indicated in the sociogram above). In particular its treatment of bisexuality and the relationship between Jack and Ianto sparked attention not just among fans of the series, but likewise in scholarly work (see for instance Dhaenens, 2013; Powers, 2016; Wilde, 2015).

### 3.2.5 *Sherlock*

*Sherlock* is one of the more recent adaptations of Doyle's famous detective stories. Similarly to *Torchwood*, *Sherlock* is produced by the BBC and is set in present-day London. The corpus includes the first three seasons<sup>28</sup> and is restricted to the two main characters. The initial corpus collections (and part of the analysis) also included more peripheral characters, such as Mrs. Hudson, the landlady, Molly, a scientist, and John's wife Mary, but the word count for each of these characters was considerably lower than that of other characters included in the corpus and eventually deemed unreliable when it came to interpretation of results.

Included in the corpus is title character Sherlock who works as a private detective and is, in his own words, "a high-functioning sociopath" (*A Study in Pink*, 2010). The series explores his remarkable ability to solve cases, his attention to detail (which he attributes to his 'mind palace'), as well as his worsening addiction to drugs. In the first episode of the first season he befriends John Watson, a medical doctor who has just returned from Afghanistan.

---

<sup>28</sup> A Victorian-era Christmas special episode (2016) as well as season four (2017) containing three episodes were broadcast after the corpus collection was completed. It is unclear at this point whether a fifth season is planned.

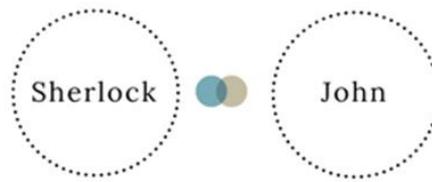


Figure 9: Sociogram: *Sherlock*

John, who is still adjusting to his life back in the United Kingdom, moves in with Sherlock and henceforth accompanies Sherlock on his cases. By the end of the first season they are partners in crime (prevention) and reliant on each other. John is instrumental in Sherlock's success as a private detective by publishing case stories on an online blog.

By the end of season two Sherlock simulates his own death by jumping off a high-rise roof as part of his effort to defeat Moriarty, his nemesis. He vanishes for two years all the while making John believe that he is still dead. The third season also features John's wedding to Mary at which Sherlock expresses his deep gratitude for their friendship.

The series has received attention because of Sherlock's portrayed mental state. While Sherlock diagnoses himself (see quote above), there is some discussion as to whether the character might be on the autism spectrum. While there is no official confirmation (or denial)<sup>29</sup>, there are some cues given that are comparable to what can be seen with Max from *Parenthood* (see above discussion). His straight-forward manner and directness to the point of being rude, selected attention and intense focus on details all point to a possible othering that can be seen in *Parenthood*. Even without a clear commitment to a

---

<sup>29</sup> This is similar to Sheldon from *The Big Bang Theory*, who is portrayed with some overlapping qualities. This might be a case of producers/actors not wanting to commit to a specific situation as that will demand the audience's request for authentic representation of the disorder.

diagnosis, my analysis might be able to detect some overlap with Max's linguistic profile in terms of pragmatic competency.

### **3.3 Corpus description**

The corpus comprised approximately 1.7 million words of spoken dialogue. Table 3 below summarizes the distribution of word counts for each character, as well as series and social factors.

It is clear from the table that word counts for the individual characters differ greatly, from the lowest word count just over 5,000 words (Rhys), and the highest at more than sixty times as much (Lorelai with 317,313 words). This considerably complicates any wider comparison of frequencies of linguistic features, which is why all results will be normalized to per 10,000 words.

Table 3: Corpus summary

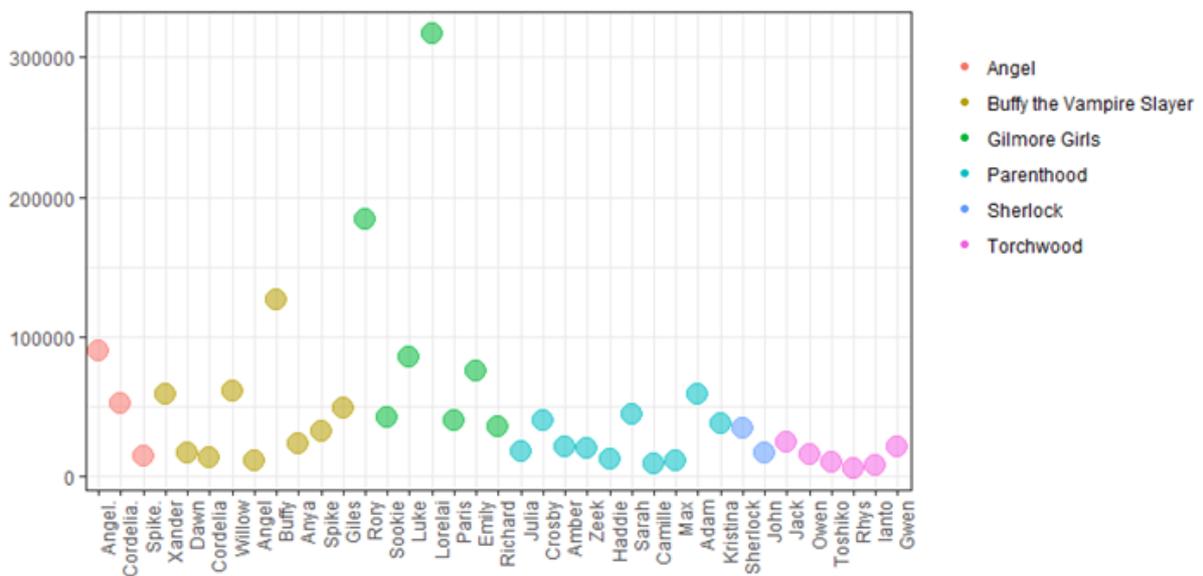
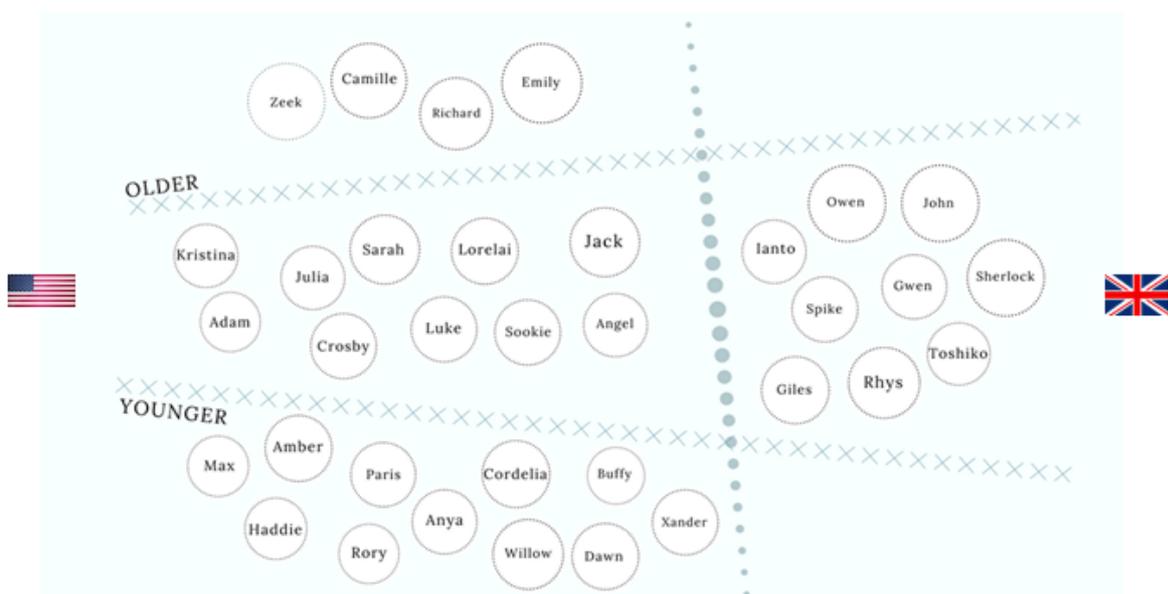
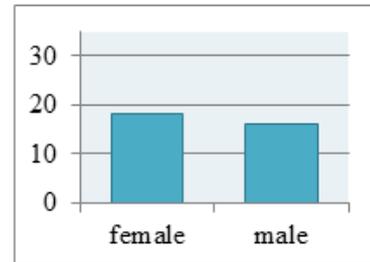
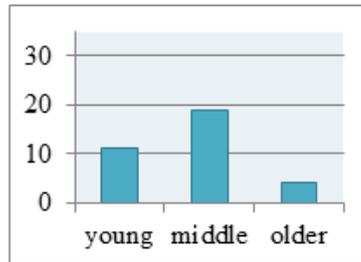
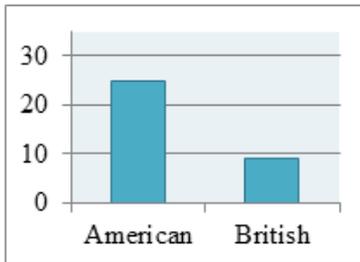
<b>Series</b>	<b>character</b>	<b>word counts</b>	<b>gender</b>	<b>age group</b>	<b>nationality</b>
<i>Angel</i>	Angel	89905	male	middle	American
	Cordelia	51857	female	young	American
	Spike	13660	male	middle	English
<i>Buffy the Vampire Slayer</i>	Xander	58509	male	young	American
	Dawn	16215	female	young	American
	Cordelia	13335	female	young	American
	Willow	60492	female	young	American
	Angel	11022	male	middle	American
	Buffy	126682	female	young	American
	Anya	22716	female	middle	American
	Spike	32228	male	middle	English
	Giles	48505	male	middle	English
	<i>Gilmore Girls</i>	Rory	183463	female	young
Sookie		41820	female	middle	American
Luke		84922	male	middle	American
Lorelai		317313	female	middle	American
Paris		40036	female	young	American
Emily		75047	female	older	American
Richard		35703	male	older	American
<i>Parenthood</i>	Julia	17726	female	middle	American
	Crosby	40039	male	middle	American
	Amber	21094	female	young	American
	Zeek	19603	male	older	American
	Haddie	12483	female	young	American
	Sarah	44579	female	middle	American
	Camille	8933	female	older	American
	Max	11208	male	young	American
	Adam	58206	male	middle	American
	Kristina	38037	female	middle	American
<i>Sherlock</i>	Sherlock	33742	male	middle	English
	John	16328	male	middle	English
<i>Torchwood</i>	Jack	23682	male	middle	American
	Owen	15285	male	middle	English
	Toshiko	9401	female	middle	English
	Rhys	5027	male	middle	Welsh
	Ianto	7000	male	middle	Welsh
	Gwen	20590	female	middle	Welsh
<b>Word count overall</b>		1726393			
<b>Word counts by series</b>	<i>Angel</i>	155422			
	<i>Buffy the Vampire Slayer</i>	389704			
	<i>Gilmore Girls</i>	778304			
	<i>Parenthood</i>	271908			
	<i>Sherlock</i>	50070			
	<i>Torchwood</i>	80985			
<b>Word counts by gender</b>	female	1121819			
	male	604574			
<b>Word counts by age</b>	young	618090			
	middle	969017			
	older	139286			
<b>Word counts by nationality</b>	American	1524627			
	British	201766			

## The television dialogue corpus

Dialogue subsets:

*Buffy the Vampire Slayer (BVS)*, *Angel (A)*, *Gilmore Girls (GG)*,  
*Parenthood (PH)*, *Torchwood (TW)*, *Sherlock (SH)*

total word count: 1726393



## **4 Methods**

The analytical part of this thesis is based on empirical observations of language feature distributions across the dialogue of individual characters as well as series that have been introduced in the previous section. This chapter will introduce the extraction of the chosen language features as well as give an overview of the quantitative methods used in the analysis that will follow.

### **4.1 Corpus use**

As previously mentioned, dialogue for the individual characters (for each season) was stored in separate files, ordered by series (Figure 10), and totalling 175 files. This enabled me to get familiar with each character's language on its own and to perform initial word list searches. This was particularly important in deciding which features to investigate, as not all features appeared in near-enough frequencies to allow for quantitative evidence for character-specific patterns. Further, it provided a first indication of differences that could be later tested for significance on a larger scale.

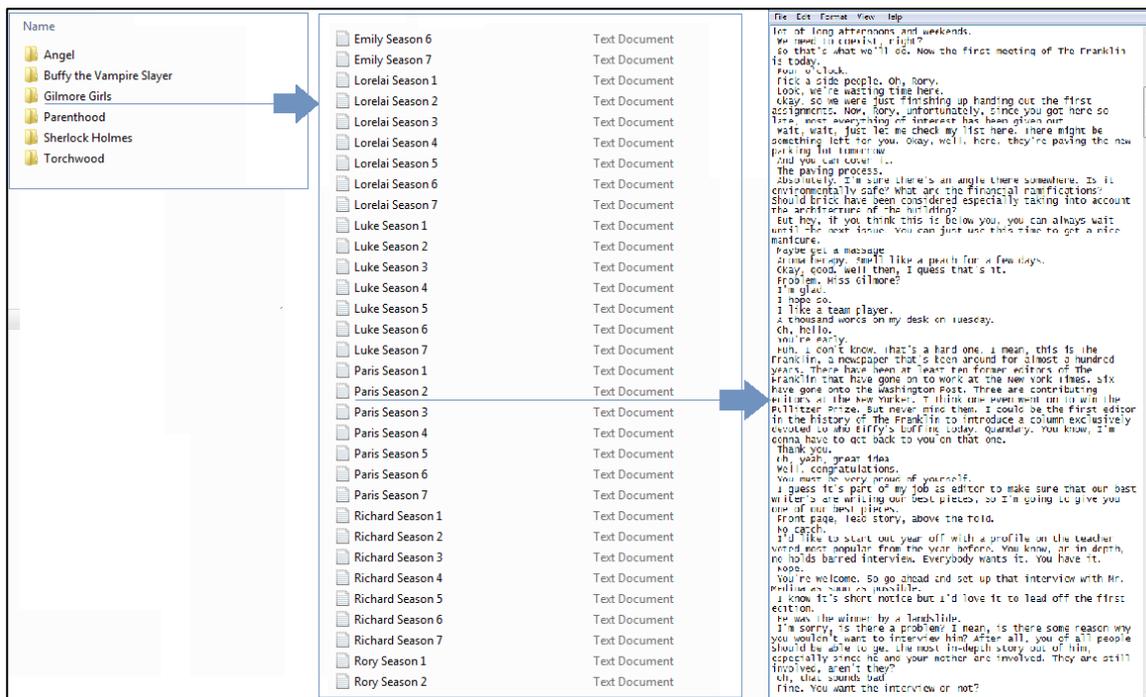


Figure 10: Storage of series files

## 4.2 Feature extraction

For the feature extraction, each linguistic variable had to be clearly defined in the contexts they would appear in and the functions they would perform. I will present the process of feature extraction as well as the preparation for analysis (i.e. coding the tokens that were extracted) with example variables used in the analysis.

Example for feature extraction: Pragmatic markers

The pragmatic markers included in the analysis are *you know*, *I mean*, and *like*. However, because *you know* has varying phonetic realizations, forms *y'know* and *ya know* were also included.

The extraction of features from the sub-corpora was done with AntConc (Anthony 2014)<sup>30</sup>. AntConc is free corpus software with basic functions for creating word lists, concordance lines, and individual and n-gram searches using actual word and regex forms. For the feature extraction the whole directory (175 files) was loaded into the programme (Figure 11).

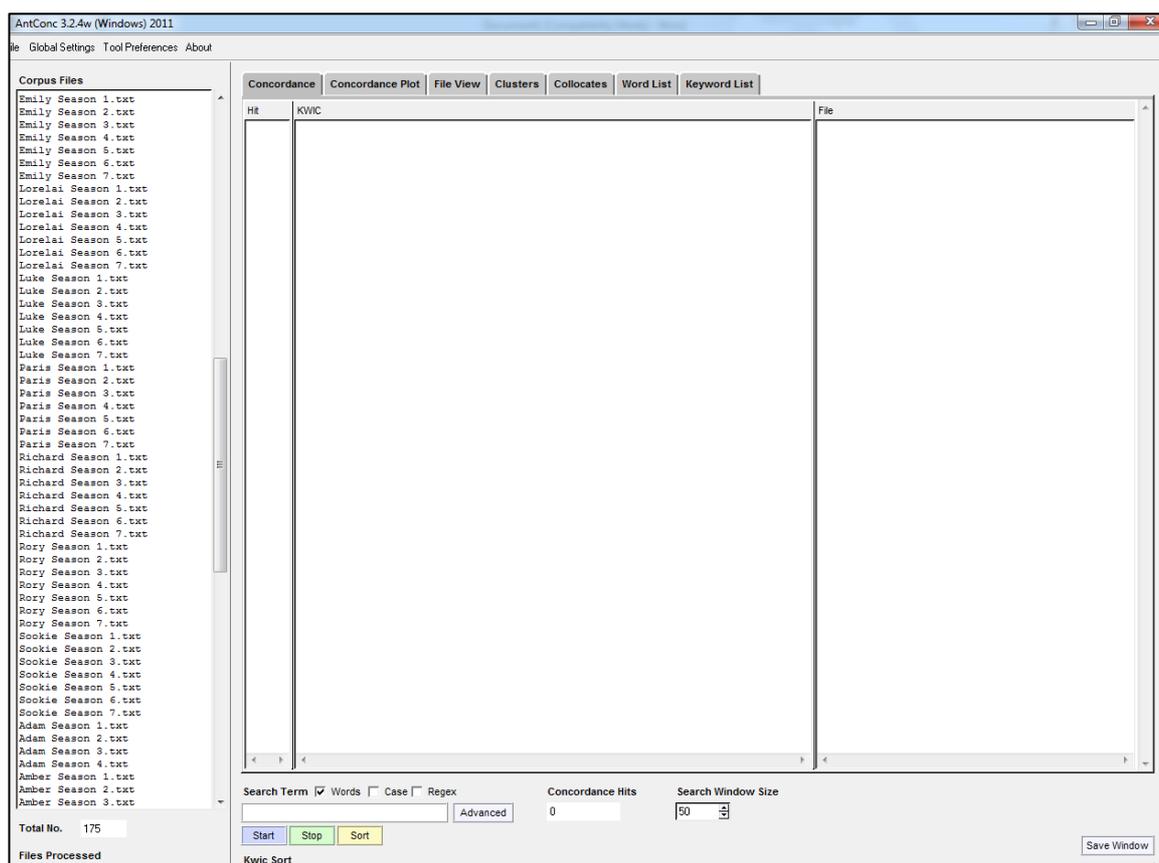


Figure 11: Loaded TV Corpus in AntConc

The individual variants for pragmatic markers were then searched for. The programme treats simple word (and multi-word) searches in lower case (as per setting), enabling me to include all forms independent of their syntactic position. Some transcripts would pre-

---

<sup>30</sup> For the present study versions 3.2.4 (2011), and 3.4.4 (2014) were used (depending on the machine used at the time). Both versions yield identical search outputs.

define pragmatic marker functions by inserting the form between commas indicating pauses (see example 21).

- (21) Xander: But, uh, what makes you think that's, *like*, a good idea?  
(*BVS*)

However, because transcription conventions were not consistent across all series and contributors, and some pragmatic marker functions were not recognized in between pauses, I included every possible representation (see examples (22) and (23) for different representations of pragmatic marker *like*).

- (22) Willow: You mean *like*, some hell-beastie rode in with Buffy?  
Like...we're responsible for this? (*BVS*)
- (23) Amber: I watched my mom, *like*... Get completely dragged down  
by somebody. (*PH*)

The output in AntConc for a single search unit (Figure 12) provides concordance lines with 50 characters before and after the search term (middle panel marked 'A') which is needed to disambiguate the function of each token.<sup>31</sup> The search indicates a first account of how many tokens are included in the search at this point (in the example below 5920 tokens were included in the initial search for *you know*).

---

<sup>31</sup> The left panel marked 'B' provides the token line as a numbered item and the right panel marked 'C' indicates information of the sub-corpus (which character and season) for each token.

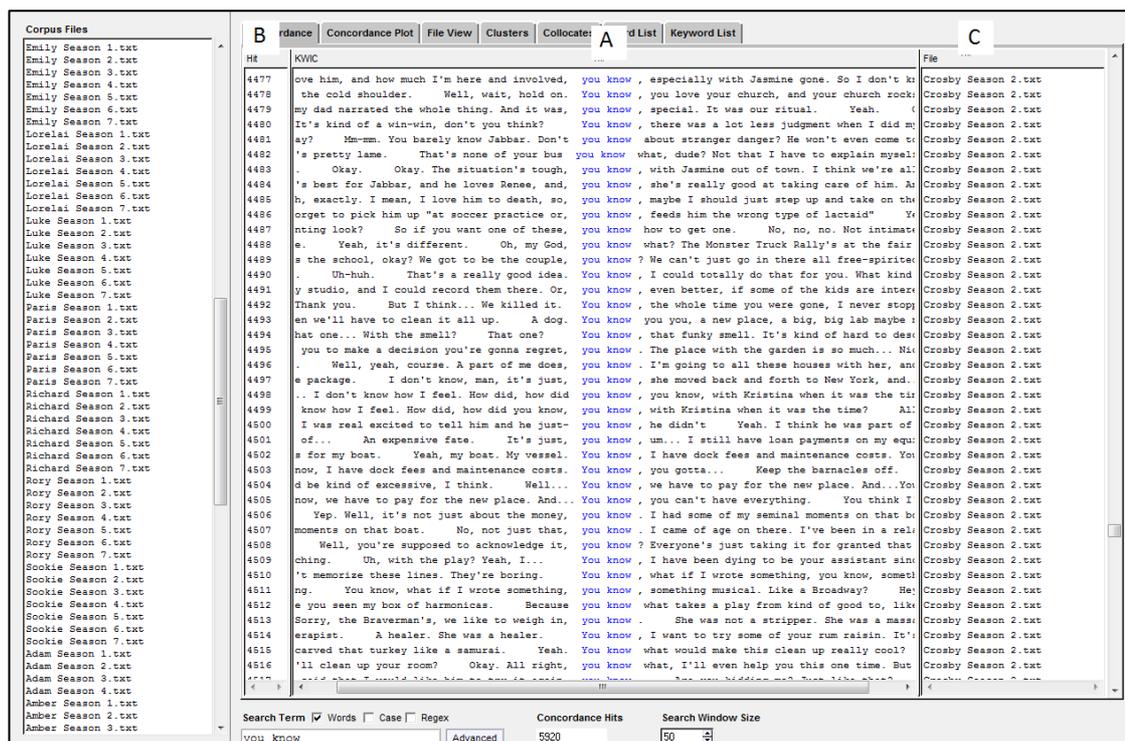


Figure 12: Search for *you know*

All searches were individually exported into text files (.txt) for further coding. These files include each token line as well as the meta-data of character, season, and token number. For *you know* a total of 5997 token lines were extracted, with just over 1% of tokens in contracted or colloquial forms (*y'know* and *ya know*).

### 4.3 Coding

Coding of extracted data describes the process of categorizing and excluding token lines from the analysis. For most sociolinguistic studies this is a manual and time-intensive process, as discourse functions cannot yet be exclusively determined by automatic machine-coding.

Examples for coding: Hedges (*sort of, kind of*)

In the case of hedges for instance, extracting all forms that are described as *sort/kind + of* does not automatically follow similar function within discourse. Example (24) shows the

above described form in a context that is not hedging. If we take surrounding structures into consideration we might delimit the output to forms that are not preceded by articles (as these forms would usually follow typification functions – detailed further in the analysis section on hedges). However, as example (25) shows, not all token lines follow clear and definable syntactic structures.

(24) Cordelia: I bet this sort of thing happens all the time. (*BVS*)

(25) Lorelai: No, it was a - kind of a ‘hello’ kind of thing. (*GG*)

Software that enables automatic semantic tagging and tagging of parts of speech were considered for this part of data preparation, but too many inconsistencies were found that ultimately led to manual coding<sup>32</sup>.

An Excel file was prepared for each variant to record coding as well as ensure meta-data was kept for all tokens that were applicable for the analysis. Each token line from the variable text files was copied into the excel files and manually examined for function and context. Forms that were not applicable were coded for their exclusion reason rather than deleted from the file so that the process could be documented and reproducible at a later point.

---

<sup>32</sup> For instance, automatic tagging of pragmatic marker *like* is not 100% accurate in distinguishing all five functions I have included in the analysis. Similarly, only through manual coding of hedges *sort of* and *kind of* was I able to determine syntactic patterns as well as exceptions to these patterns that would otherwise have gone unnoticed.

Not included in the analysis itself but nonetheless noteworthy in this context are forms such as vocatives that are multifunctional (semantic field of ‘food’ for instance was marked for the *Gilmore Girls* corpus, which is likely for one of the main characters being a chef – but on closer inspection it turned out that in fact the characters used high frequencies of *honey* as a vocative. The highly marked semantic field of ‘geography’ for the same series turned out to be almost exclusively based on the fact that one of the characters is named Paris.) Themes of the series also changed the tagged wordlists. Intensifier/expletive *bloody* for instance shows up at higher frequencies in crime series (that deal with *bloody* murder) due to higher adjective/adverbial use and at disproportionate levels in series that have vampires in them.

Figure 13 shows the detailed coding of *kind of* with examples from *Gilmore Girls*.

	B	C	D	E	F	G	H	I	J	K
	TEXT	CORPUS	SERIES	YEAR	CHARACTER	FUNCTION	DET + 'T'	'T' + DET	MOD	SPECIAL
83	ner. Some mosaic tiles, some colored beads. Still <b>kind of</b> kids' stuff, but Terrence showed me how working w	Paris Season 4.txt	<i>Gilmore Girls</i>	2003	Paris	hedge			n	
84	f yours aren't quite as charming. It means these <b>kind of</b> clubs look good on your resumé. No one studies m	Paris Season 4.txt	<i>Gilmore Girls</i>	2003	Paris	type	these		n	
85	ly Nancy Spungening it, chasing the dragon, and I <b>kind of</b> have to be here for support. The chick's bouncing	Paris Season 4.txt	<i>Gilmore Girls</i>	2003	Paris	hedge			v	
86	aking. At least it's warm in Burma. This is the <b>kind of</b> cold you read about in a Dickens novel. We should	Paris Season 4.txt	<i>Gilmore Girls</i>	2003	Paris	type	the		n	
87	a good idea to hang out with people my own age. I <b>kind of</b> hoped that he would invite me to go to Denver wit	Paris Season 4.txt	<i>Gilmore Girls</i>	2003	Paris	hedge			v	
88	died right at the height of my passion for him. I <b>kind of</b> wonder what would have happened if he had lived.	Paris Season 5.txt	<i>Gilmore Girls</i>	2004	Paris	hedge			v	
89	And putting me on hold is not one of them. I am <b>kind of</b> hungry. And C-span can kiss my ass. In a minute.	Paris Season 5.txt	<i>Gilmore Girls</i>	2004	Paris	hedge			adj	to be
90	was a vicious Julia Roberts to Kiefer Sutherland <b>kind of</b> dump. She broke his heart, slept with his best fr	Paris Season 5.txt	<i>Gilmore Girls</i>	2004	Paris	type	a + adj		n	
91	king after you. It's necessary you seem like the <b>kind of</b> lunkhead that would get up too soon and inadverte	Paris Season 6.txt	<i>Gilmore Girls</i>	2005	Paris	type	the		n	
92	get crazy-drunk in Cancun and flash your breasts" <b>kind of</b> way. Whatever. Later. Oh, and Lexapro is fast-	Paris Season 7.txt	<i>Gilmore Girls</i>	2006	Paris	type	adj		n	kind of way
93	onauts or apes. Will there be dancing? What <b>kind of</b> dancing? So we're talking mostly hip-hop. I	Paris Season 7.txt	<i>Gilmore Girls</i>	2006	Paris	type	what		n	
	our years ago. And don't look at me like I'm some <b>kind of</b> superstitious freak. It's	Paris Season 7.txt	<i>Gilmore Girls</i>	2006	Paris	some	some		adj + n	to be

Figure 13: Coding *kind of*

The token lines (here 11 lines out of 1299 in total) were manipulated with a macro function to highlight the relevant variable (*kind of* in bold typeface) in the TEXT column. Metadata from the CORPUS column was used to add further information (displayed here: what series, what character, which year). The main functions (hedge, type) were differentiated as the main coding goal. The columns to the right of FUNCTION are used for internal variation analyses<sup>33</sup>.

<sup>33</sup> DET + 'T' describes the determiner that precedes the token (if), 'T' + DET describes determiners used following the token but preceding the head of the phrase, as in *kind of a house*. MOD describes the type of head that is modified (noun, verb, preposition, etc.) and SPECIAL was used to track specific phrasal tokens (kind of thing, kind of stuff) or particular occurrences (forms of to be + hedge). While not part of the main analysis, this coding enabled me to find patterns of use of these variables in more general terms, adding to existing discussions on grammaticalization/pragmaticalization of these pragmatic forms.

Once the variables are fully coded for their functions, applicable occurrences are counted for each character and added to a basic frequency table.

This table includes all character information needed for the analysis:

- Character name
- Meta-data: series, word count, year
- Social factors: age, gender, nationality, species (vampire or human)
- Production factor: genre, production location

The full table comprises all 37 characters and 74 columns of character information and frequencies for variables. As previously mentioned, all variables occur in observed frequencies as well as normalized frequencies depending on the word count of the respective character. The normalized frequencies represent the use of a feature per 10,000 words and is the frequency that is used for distributional comparison in the main analysis. In comparison, Mandala, in a study focusing on characterization patterns reflected in the use of *y*-suffix adjectives in *Buffy the Vampire Slayer* (2007), presents data in raw numbers. While this exemplifies uses of the feature in a straightforward manner, it does not take into account varying speech portions of the individual characters. In the present study, this is even more important to note, as characters across the Television Dialogue Corpus speak anywhere between 5027 and 317,313 words. In order to provide maximal comparability, not just within the present study but also for any future investigation, I thus decided to present findings in normalized frequencies.

An excerpt from the frequency table for the numbers included for hedges can be seen in Figure 14 below. The full table is provided in the appendix at the end of the thesis (Appendix B).

	A	B	C	D	E	F	G	H	I	AG	AH	AI	AJ	AK	AL	AM	AN	AO
1	character	series	production	genre	gender	age	nationality	species	words	H sort of R	H kind of R	H sort of E	H kind of E	all sort of	all kind of	Dvalue sort of	Dvalue kind of	HSK
5	Xander	Buffy the Vampire Slayer	US	science fantasy	male	2	American	human	58509	17	72	2.9055359	12.3057991	19	102	89.47368421	70.58823529	9.400265
6	Dawn	Buffy the Vampire Slayer	US	science fantasy	female	1	American	human	16215	4	19	2.46685168	11.7175455	5	21	80	90.47619048	9.250695
7	Cordelia	Buffy the Vampire Slayer	US	science fantasy	female	2	American	human	13335	8	52	1.54270397	10.0275758	12	69	66.66666667	75.36231884	8.484871
8	Willow	Buffy the Vampire Slayer	US	science fantasy	female	2	American	human	60492	34	83	5.62057793	13.7208226	39	106	87.17948718	78.30188679	8.100244
9	Angel	Buffy the Vampire Slayer	US	science fantasy	male	3	American	vampire	11022	14	80	1.55719927	8.89828152	15	109	93.33333333	73.39449541	7.341082
10	Buffy	Buffy the Vampire Slayer	US	science fantasy	female	2	American	human	126682	45	134	3.55220158	10.5776669	49	175	91.83673469	76.57142857	7.025465
11	Anya	Buffy the Vampire Slayer	US	science fantasy	female	2	American	demon	22716	6	12	2.64131009	5.28262018	7	20	85.71428571	60	2.641310
12	Spike	Buffy the Vampire Slayer	US	science fantasy	male	3	British	vampire	32228	3	7	0.93086757	2.17202433	7	19	42.85714286	36.84210526	1.241156
13	Giles	Buffy the Vampire Slayer	US	science fantasy	male	3	British	human	48505	15	5	3.09246469	1.03082157	32	12	46.875	41.66666667	-2.06164
14	Rory	Gilmore Girls	US	dramedy	female	2	American	human	183463	18	139	0.98112426	7.57645956	22	184	81.81818182	75.54347826	6.595335
15	Sookie	Gilmore Girls	US	dramedy	female	3	American	human	41820	9	32	2.15208034	7.65184122	9	47	100	68.08510638	5.499760
16	Luke	Gilmore Girls	US	dramedy	male	3	American	human	84922	15	47	1.76632675	5.53449047	16	59	93.75	79.66101695	3.768163
17	Lorelai	Gilmore Girls	US	dramedy	female	3	American	human	317313	46	133	1.44967272	4.19144504	57	236	80.70175439	56.3559322	2.741772
18	Paris	Gilmore Girls	US	dramedy	female	2	American	human	40036	7	10	1.74842642	2.49775202	9	18	77.77777778	55.55555556	0.749325
	Emily	Gilmore Girls	US	dramedy	female	4	American	human	75047	12	5	1.59899796	0.66624915	20	26	60	19.23076923	-0.932741

Figure 14: Frequency table for hedges

The frequency table was then exported to a comma-delimited file that was readable in RStudio (.csv) where most of the distributional analysis, multivariate analysis, as well as plotting was performed.

#### 4.4 R & Rstudio

The software used for the analysis is R and, embedded in that, open source software RStudio, both used in various packages<sup>34</sup>. R is a programming language enabling researchers to perform mathematical and statistical tests, data manipulation, and graphical representations. R is a command line interface which means that it runs through the input of commands, such as

```
> sum(4, 4) yielding [1] 8 as a result.
```

<sup>34</sup> R: v3.0.0 (2013), v3.1.1 (2014) & v3.3.2 (2016), RStudio: v0.98.493 (2014) up until v1.0.136 (2016)

RStudio is an embedded environment that offers user-friendly handling of the R interface, providing a console for command input and package libraries that offer various tools for creative plotting of data.

The data table (in csv format) is imported into RStudio and manipulated so that all categories are represented in accurate data types (categorical or numerical)<sup>35</sup>.

#### 4.4.1 *Distributional analyses*

The main part of analysis describes the distribution of frequencies of the features. These analyses are done directly in RStudio and are represented in table and plot form<sup>36</sup>. Here, individual variables are compared in use across speakers and different speaker factors, as well as across variables as a whole where appropriate. Cross-tabulations here are useful tools to see if factors interact (age and gender for instance) or if frequencies are possibly skewed by uneven character distributions for some of the features. This part will form the main body of the analysis. Here, observable differences in language use are mapped onto characterization patterns. Differences in language use (higher or lower frequencies of certain variables) might indicate that the features are actively used to distinguish characters from one another, or at least, that these distinctions are subconsciously used for characterization purposes.

---

<sup>35</sup> For category age group for instance RStudio automatically suggests integer types (because of the numerical values in the cells). Because these are not countable frequencies though but rather groupings similar to the binary gender group, I changed the category to character (for categorical types). Newer versions of RStudio offer this data frame manipulation with the import of tables.

<sup>36</sup> Plots are predominantly created and modified with packages ggplot2 (v2.1.0 2016), gridExtra (v2.2.1 2016), gtable (v0.2.0 2016), and extrafont (v0.17 2014)

#### **4.4.2 *Statistical testing***

In order to see if distributions across different speakers and speaker groups are meaningful or by chance (indicating whether television language for instance exhibits language differences by gender), the data is tested for significance. The test that is used throughout the thesis is the Mann-Whitney U nonparametric test.<sup>37</sup> I include references to the tests within the respective analysis sections, as well as in summary-form in Appendix C.

In addition to statistical tests as described above, the intensifier section also includes a multiple regression analysis using Rbrul (Johnson, 2009). This type of analysis is used in variationist sociolinguistic studies where variants are part of a closed and clearly defined set (for in-depth discussions on how variables can be defined in different contexts see for instance Buchstaller, 2009; Labov, 1972b; 1978; Pichler, 2010). The aim of this type of analysis is to test variant choice against a set of constraints (social such as gender and age, or linguistic, such as syntactic context) to determine which factors might be most relevant for the direction of variation. Further information on interpretation of statistical values are given in section 5.5.

#### **4.5 Limitations**

In contrast to other studies (e.g. Mandala, 2007; Tagliamonte and Roberts, 2005), I am including a relatively large variety of linguistic features in my analysis. The main reason for this is the aim to trace possible interaction of features that may correlate with the creation of characters and character types. A draw-back in widening the linguistic scope is that the individual features cannot be as thoroughly analysed as in other, comparable

---

<sup>37</sup> Reasons for the choice of test lie in the non-normal distribution of the data as well as its robust performance with outliers.

studies. This is particularly noteworthy with regards to pragmatic functions of multifunctional items, such as *like* or *you know* (further described in chapter 5). However, as described above as well as in the corresponding sections for the individual features, all tokens from the corpus are manually coded for basic pragmatic functions. This, in all cases, is based on previous sociolinguistic studies on possible quantitative approaches to pragmatic forms and thus complies with common methods of investigating pragmatic forms. This, compared to Quaglio's approach (2009) for instance, allows for replicability and possible expansion. In contrast, Quaglio includes a set of features with limited manual coding and little acknowledgement of possible discrepancies in functions. My study offers a broad-stroke analysis of features within a specified set of pragmatic functions. While a detailed functional analysis that takes concrete discourse contexts into consideration would be undoubtedly insightful, the scale of included characters and scope of the thesis itself negates such an approach.

## 5 Variation and Characterization analysis

This chapter is separated into five subsections, each detailing one group of linguistic markers and how they might be used in fictional television series. My aim is to answer the following overarching questions for each of the marker categories:

- Does the distribution of markers reflect characterization patterns within and across series?
- Does the distribution indicate indexical and/or stereotypical characterization and if so, what characters or character groups are represented like this?
- Are distributional outliers and unexpected patterns indicative of specific characterization choices?

A summary at the end of each section will attempt to present a comprehensive overlook of characterization patterns across features. The goal here is to tease out in what ways character types are supported through a combinational effort of linguistic markers – introducing a first method for finding stylistic language patterns across characters and series.

The focus in this part of the analysis is on variation and how different linguistic choices made by the characters (or, outside the fictolinguistic framework, the script-writers) reflect differences in their personas. Analysis B (chapter 6) will focus on the notion of language change and character development.

## 5.1 Pragmatic Markers

### 5.1.1 *Feature definition*

For the first feature group I have chosen three pragmatic markers (*I mean, you know, and like*), each with slightly differing functions and varying indexical meanings. This is important to note, as the group of pragmatic markers cannot be defined as a sociolinguistic variable, i.e. different ways of saying the same thing (see above). Variation that is of interest here is not first and foremost the preference of one pragmatic marker over the other, but rather the relative frequencies of each pragmatic marker by the individual character and how this frequency compares across the Television Dialogue Corpus.

Pragmatic markers (elsewhere also called discourse markers (Schiffrin 1987, Schourop 1999), discourse particles (Schourop 1985), or pragmatic expressions (Ermann 1987)) are part of a group of discourse features that is still relatively ill-defined in (socio-)pragmatic research. Aijmer (2015:195) alludes to this, saying that “they [pragmatic markers] have a large number of functions, there is no consensus about the linguistic model needed to describe them, and the relationship between form and function is complex”. Schiffrin (1987:31), in her seminal work on discourse markers, defines the feature group as “sequentially dependent elements which bracket units of talk”. That means that they “can occur quite freely within a sentence at locations difficult to define syntactically” (Schiffrin, 1987:32), which contributes to their functional flexibility. Indeed, previous studies have assigned a plethora of functions to the individual features and the contexts in which they appear: markers can act as turn-taking orientation, politeness and face-saving acts, vagueness markers, hesitation devices, and/or repair signs. The multifunctional nature of pragmatic markers can cause difficulties in determining general meanings and, as Beeching summarizes, “precise categorisation and percentages are thus

difficult to ascertain, and a certain fuzziness is inevitable” (2016:34). A solution to this is the notion of core meanings, a concept, defined below, that I will apply to the present study.

Generally speaking, pragmatic markers are defined through their lack of propositional function to an utterance: if removed from the sentence, the truth value still stands. An example (26) from the Television Dialogue Corpus is given below:

- (26) Adam: And I hope you don't take it the wrong way but, *you know*, just cause it's the music business, *you know*, it doesn't mean that, *you know*, we expect you -- or that you have to dress a-a certain way. (PH)

In this utterance, Adam tries to talk to his assistant about the way she dresses, which makes him clearly uncomfortable. He is trying to mitigate (“don't take it the wrong way”), pre-emptively giving her an excuse (“just cause it's the music business”), hesitating (“a-a”), and being vague (“a certain way”) in order to save her face. In using pragmatic marker *you know* three times here, Adam acknowledges the fact that he might be overstepping their professional relationship and so hedges his approach. All tokens of the pragmatic marker could be omitted from this utterance without the propositional message being changed, although it would arguably make his statement too forward and possibly face-threatening.

Pragmatic markers, as seen in the example above, “do not convey social and/or expressive meanings” (1987:318), rather they function on a discourse managing level with what Schiffrin referred to as ‘core meanings’ and Fox Tree and Schrock (2002) proposed as ‘basic meanings’. Holmes (1995) claimed that a distinction can be made between self-oriented and other-oriented markers, depending on the personal pronoun that is part of

the marker. With regards to the markers chosen for the present study, *I mean* is a self-oriented marker, whereas *you know* can be described as other-oriented. This is comparable to the basic meanings suggested by Fox Tree and Schrock (2002), who said that *I mean* serves the speaker in forewarning adjustments to the utterance (2002:744) and *you know* is used to invite “addressee inferences” (2002:744).

Sociolinguistic research into pragmatic markers has yielded inconsistent results, with some studies claiming that pragmatic markers are used most frequently by female speakers (e.g. Coates, 2013 on hedges; Macaulay, 2002 on *you know*), while others have found male speakers to lead in the use (e.g. Holmes, 1990 on *you know*). Age is a much more straightforward predictor in most studies, in that younger speakers are the predominant users of new, incoming pragmatic markers (e.g. Tagliamonte, 2005), while well-established pragmatic markers (such as *you know*) are used increasingly with age (e.g. Beeching 2016). Additionally, pragmatic features were described to be more common in American English (e.g. Biber et al., 1999 on discourse markers in general; Erman, 2001 on *you know*) than British English.

While they do not carry propositional meaning, the features can be used as indexical markers. Beeching (2016:13) claims that they “may (...) serve functional and identity-constructing purposes”. This is particularly important for the current study, as identity-construction and stylization for characterization happens on a much more conscious level (by the writers and actors) and pragmatic markers might be used to support the identities that are being portrayed. While broad social categories of gender, class or age might not be clear predictors, they do inform indexical meanings and serve as stylization markers. An example of that can be seen in the use of pragmatic marker *like*. Outside of linguistic research, this marker attracted attention through its strong association to a particular speaker group: young and female, typically from California (*Valley Girls*). This

association turned the marker into a stereotype (Labov, 1972a), or according to Silverstein (2003) a third-order indexicality marker. Despite stereotypes not always reflecting actual language use (the use of pragmatic marker *like* for instance is far from being used exclusively by young and female speakers), they contribute to a very specific impression of speaker style. In terms of stylisation for television characters it can be expected that stereotypes such as the high use of pragmatic marker *like* by young female speakers are used to portray characters that fall into the respective scheme. This creates character types that are easy to recognize for the audience. Not all pragmatic markers are stereotypes, although research has shown that they can all carry indexical meaning (Johnstone, 2010) which might be used for characterization purposes.

### **5.1.2 *Token inclusion & coding***

All tokens of *I mean, you know* (as well as *y'know*) and *like* were extracted from the main corpus and coded manually for their function within the context of the interaction. Where concrete functions were unclear from the token context (50 characters before and after the token), the transcript was used to determine the function. This was the case in some instances of *you know*, where it was used as a minimal response and only the full transcript was able to show whether it was used as a pragmatic marker or a propositional utterance.

#### *You know*

I found four main functions of *you know* in the Television Dialogue Corpus: the propositional use (27) where the knowledge of the interactant is mentioned and the phrase oftentimes co-occurs with an object (here 'that'), the emphatic use (28) where the propositional meaning is highlighted, the pragmatic marker use (29) where the phrase does not add to the propositional meaning of the utterance and the utterance is made non-committal, and finally the pragmatic marker use (30) where the marker implies certainty.

Here, the interactional function is not hedging or non-committal but rather condescending, what Beeching called impositional (2016:106).

- (27) Lorelai: He's gonna come and go as he pleases, babe, *you know* that.  
(GG)
- (28) John: You see? *You know* things. (SH)
- (29) Angel: It doesn't keep me alive, but, *you know*, sometimes I get a hankering. (A)
- (30) John: I'm not stupid, *you know*. (SH)

The distinction between the last two functions of pragmatic marker *you know* is discernible mostly through the context of the conversation and intonation (with intonation rising for the hedging pragmatic marker and falling for the impositional use).

Pragmatic marker *you know* can be found clause-initially, mid-clause and clause-final.

For the present analysis only markers that increased non-committal of the speaker (29) were included to allow for comparison across speakers and series.

### *I mean*

Similarly to *you know*, *I mean* can have propositional meaning in conversation. While propositional *you know* can occur without an object (as a minimal response for instance), propositional *I mean* almost always co-occurs with an object, either preceding the phrase (31), or following it (32). Pragmatic marker *I mean* (33) is not syntactically restricted, although it seldom appears in clause clause-final position in the Television Dialogue Corpus.

- (31) Rory: It's British. You know what *I mean*. (GG)
- (32) Angel: Be careful. *I mean* it. (A)
- (33) Giles: Yes, yes, so it seems. I-*I mean*, um, uh, you did indeed. (BVS)

The analysis includes all pragmatic marker functions of *I mean*, discarding propositional *I mean*.

### *Like*

*Like* has a number of propositional and pragmatic functions. In the Television Dialogue Corpus, *like* occurred with propositional meaning as conjunction (34), preposition (35), nouns (36), verbs (37), as well as in pragmatic marker contexts as approximators (38), quotatives (39) and finally pragmatic markers (40).

- (34) Angel: Looks *like* you're hurt. (BVS)
- (35) Lorelai: I feel *like* one of those cats that's bred to have no legs. (GG)
- (36) Giles: Y'know, uh, feeding patterns a-and the *like*. (BVS)
- (37) Sherlock: Maybe you liked his wife; maybe you don't *like* his drinking. (SH)
- (38) Crosby: You haven't played in *like*, three months. (PH)
- (39) Lorelai: And you're *like* 'Hey, his heart should be in his chest'. (GG)
- (40) Gwen: It's stupid, but I always sort of think, *like*, you know, white light and all that. (TW)

According to Beeching the last function can be further categorized into three subcategories, that are, alongside quotative and approximator, all linked to the primary function of a hedging device (2016:128):

Exemplifying function, where the pragmatic marker is closest resembling the propositional meaning of prepositional *like* (41). The focuser function, where *like* is used to highlight a following point, oftentimes found at the end of a clause (42). Hedge *like*, where the pragmatic marker is used to reorganize the utterance to avoid offending the listener and saying unwanted things (43).

- (41) Anya: You could uh, could have *like* a world without shrimp. (BVS)
- (42) Lorelai: I'm just curious, was she pretty? *Like*, what kind of pretty?  
(GG)
- (43) Amber: It's, I mean, *like*, I deserve it. I get it. (PH)

Functions of exemplification, focuser and hedge are combined for the purpose of this study. As both quotative and approximators arguably carry some propositional content, the comparison to other pragmatic markers *I mean* and *you know* will be carried out with pragmatic marker *like* only. In order to see if there is a correlation with the other functions however, I will include quotative and approximator *like* in a separate analysis.

### 5.1.3 *General distributions*

The initial extraction of *I mean*, *you know* and *like* amounted to 17,251 tokens, including both grammatical or propositional and pragmatic functions. Through manual coding for pragmatic functions I excluded 10,387 tokens (approximately 60% of the initial output) that do not have the core functions of pragmatic markers as outlined above. The Television Dialogue Corpus thus totals 3,317 tokens of *you know*, 2,591 tokens of *I mean*, and 956 tokens of *like*.

For the sake of comparability, as mentioned previously, I will present main findings in normalized data sets as per 10,000 words spoken. This will allow for direct comparisons across all speakers and speaker groups. The following table presents all occurrences of tokens that were, following the above-mentioned coding frames, included in the analysis. A table including relative frequencies alongside raw counts can be found in Appendix B.

Table 4: Pragmatic markers distribution, per 10,000 words

Character	Series	<i>like</i>	<i>I mean</i>	<i>you know</i>
Angel		1.00	12.68	19.6
Cordelia	<i>Angel</i>	2.89	12.53	10.4
Spike		1.46	5.12	2.9
Xander		3.76	18.97	14.4
Dawn		8.02	20.35	18.5
Cordelia		8.25	17.25	12.7
Willow		4.79	30.42	18.7
Angel	<i>Buffy the Vampire Slayer</i>	0.00	11.79	17.2
Buffy		2.68	15.71	23.2
Anya		1.76	36.10	24.2
Spike		0.00	6.21	16.8
Giles		0.00	8.45	6.8
Rory		1.31	16.62	9.9
Sookie		0.72	19.37	14.8
Luke		0.47	18.25	21.4
Lorelai	<i>Gilmore Girls</i>	1.20	12.29	15.3
Paris		0.50	12.99	8.0
Emily		0.00	5.06	5.9
Richard		0.00	1.12	5.6
Julia		4.51	14.67	48.0
Crosby		11.49	20.98	51.7
Amber		47.88	27.50	55.0
Zeek		5.10	58.66	57.1
Haddie	<i>Parenthood</i>	77.71	29.64	20.0
Sarah		14.36	29.61	74.0
Camille		2.24	14.55	88.4
Max		4.46	4.46	14.3
Adam		2.58	10.31	35.6
Kristina		14.20	23.92	31.0
Sherlock	<i>Sherlock</i>	0.00	0.59	4.4
John		0.00	5.51	8.6
Jack		0.00	2.53	3.4
Owen		1.31	4.58	17.0
Toshiko	<i>Torchwood</i>	1.06	3.19	1.1
Rhys		3.98	9.95	21.9
Ianto		0.00	2.86	4.3
Gwen		0.97	7.77	3.9
mean overall		6.23	14.93	21.78
	<i>Angel</i>	1.78	10.11	10.9
	<i>Buffy the Vampire Slayer</i>	3.25	18.36	16.9
mean by series	<i>Gilmore Girls</i>	0.60	12.24	11.56
	<i>Parenthood</i>	18.45	23.43	47.51
	<i>Sherlock</i>	0.00	3.05	6.51
	<i>Torchwood</i>	1.22	5.15	8.58
mean by gender	Female	10.27	18.40	25.42
	Male	1.98	11.28	17.94
mean by age	Young	13.67	20.21	19.11
	Middle	2.82	10.98	19.99
	Older	1.84	19.85	39.26
mean by nationality	American	8.22	18.46	26.61
	British	0.88	5.42	0.00

#### 5.1.4 *Distributions across categories*

The following section will outline distributions of pragmatic markers per 10,000 words across the individual speaker categories. Age and gender are combined to test for possible interaction of both factor groups. This structure is repeated for all following linguistic feature sections in this chapter.

##### *Age and gender*

Looking at gender distributions (Figure 15), all three pragmatic markers show that female characters are leading in use, if only marginally for *you know*. This confirms expectations gained from previous studies as well as points towards pragmatic markers indexing gender. However, if we investigate who the outliers are, then it becomes clear that there is not one speaker group that is being indexed through all pragmatic markers. For *like*, the outliers are Haddie and Amber from *Parenthood*, the two young women who might represent the stereotype of the *Valley Girl* persona, strengthening the indexicality of *like* for this particular speaker group. Surprisingly, they do not lead the use of the other two pragmatic markers.

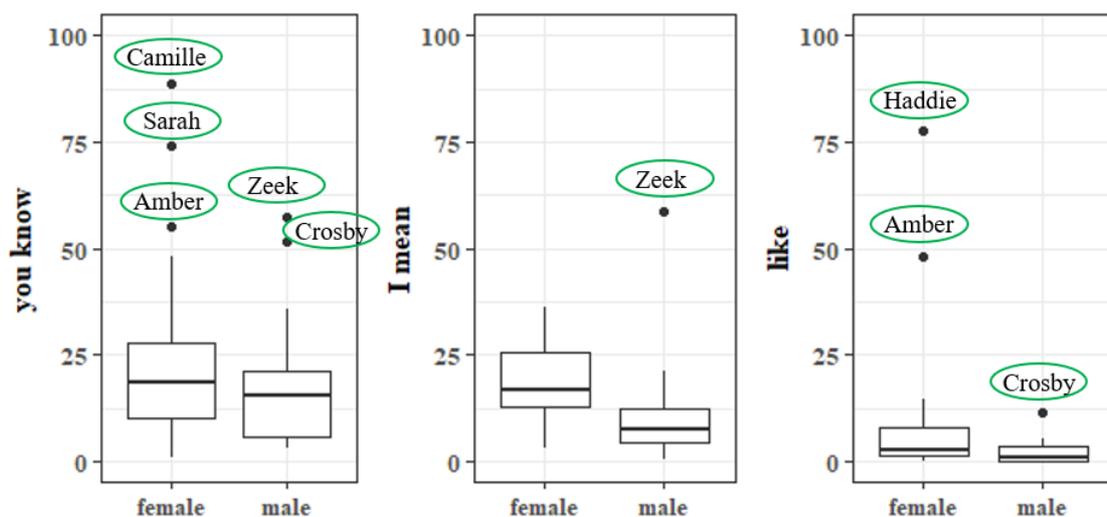


Figure 15: Pragmatic markers across gender, per 10,000 words<sup>38</sup>

*You know*, the inference-seeking marker of shared knowledge, has its outliers exclusively in the series *Parenthood*, albeit not limited to one gender or age group. This highlights the distinctive production procedure and accounts for the argument that television series as a group cannot be linguistically classed together but must always be disentangled into series, genres, and production types. *You know* does not appear to be strongly favoured by one gender, a possible indication that as a pragmatic marker, *you know* is not stereotyped as particular for one gender over the other and possibly further indicates a broad acceptance into language use (as opposed to, for instance, *like*, discussed further below).

*I mean* shows the biggest difference between male and female characters, which reflects previous studies' results and might be due to the stereotyped characterization of female

---

<sup>38</sup> The outliers are colour-coded by series:

*Parenthood*: green, *Gilmore Girls*: light green, *Buffy the Vampire Slayer*: blue, *Angel*: light blue, *Torchwood*: red, *Sherlock*: pink

characters as more tentative and repairing in their communication. This in particular is important to consider when investigating individual characters and series to see if those with strong female leads show a different distribution. In fact, as table 6 shows, direct comparison of female lead characters with their male co-cast show that the women indeed have lower rates of *I mean*, as also exemplified in table below. In comparison, *Buffy's* Willow, often portrayed as timid and shy, is among the leaders in the series' use of *I mean*. Likewise, *Torchwood's* Gwen uses *I mean* more than main character Jack.

Table 5: Use of *I mean* (per 10,000 words) main characters in *Buffy* and *Gilmore Girls*

character	Series	<i>I mean</i>
Buffy	<i>Buffy the Vampire Slayer</i>	15.70
Willow		30.41
Xander		18.97
Luke	<i>Gilmore Girls</i>	18.25
Lorelai		12.29
Jack	<i>Torchwood</i>	2.53
Gwen		7.77

Age, as previously mentioned, is a speaker category that shows varying distributions for pragmatic markers, in particular in terms of new features coming in to use, with most frequent use in younger generations. We can see that trend repeated in the Television Dialogue Corpus, with pragmatic marker *like* showing high rates in the youngest age group (which goes up to the age of roughly 25) and decreasing steadily with increasing age. A cross-tabulation between age and gender shows that age is the main determining factor here (at  $p < 0.01$ ). It seems then that *like* on the Television Dialogue Corpus predominantly indexes young age. The *Valley Girls* schema is used as a stereotype for some of the characters though it seems that the feature is broadening into other speaker groups as well.

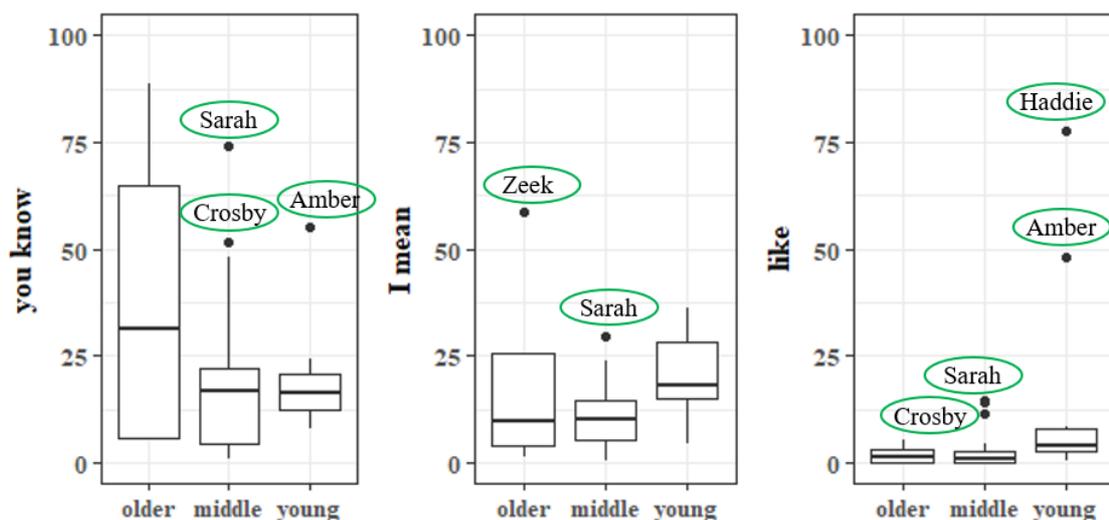


Figure 16: Pragmatic markers across age, per 10,000 words

For *I mean* we find again the youngest age group leading in use. This is surprising, as research has shown that established pragmatic markers will occur more with older speaker groups (c.f. Beeching 2016:208). It might be that the use of the feature is determined by the series, possibly influenced by the high rates in *Parenthood*. *You know* shows a relative level distribution across the age groups, with three outliers in the youngest age group (Amber) and the middle age group (Sarah and Crosby respectively).

#### *Nationality*

Finally, in terms of regional differences, previous research claimed that pragmatic markers in general are used more frequently in American English and that incoming features into British English are caused by American English, and, incidentally, television (Macaulay 2001:17).

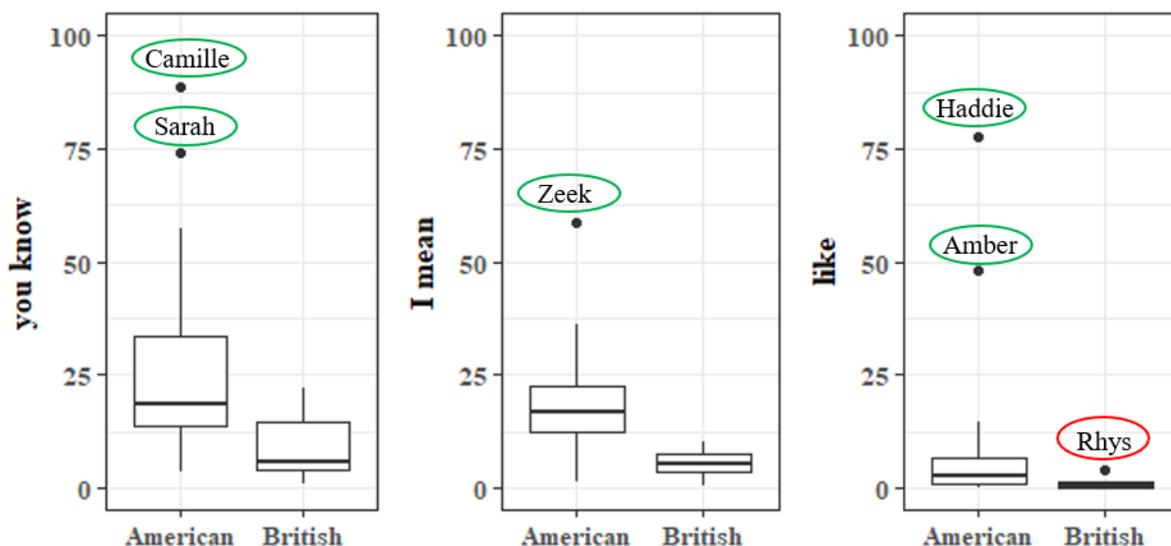


Figure 17: Pragmatic markers across nationality, per 10,000 words

We can see similar patterns in the Television Dialogue Corpus in Figure 17 above, where all three pragmatic markers are most frequently used by American characters. Patterns of British characters are not clearly marked however. Apart from the fact that British characters appear to have fewer occurrences of pragmatic markers overall, there does not seem to be an indexical use of one marker over the other implying Britishness. Nevertheless, we can see the pattern of upper social class in American characters is being indexed through British patterns. *Gilmore Girls*' Richard and Emily (and to some degree, Paris) are closest to other British characters of the Television Dialogue Corpus with low rates of pragmatic markers overall. However, with low rates of pragmatic markers also indicating diminished pragmatic competence (as we have seen with Max in *Parenthood* and Sherlock in *Sherlock*), this interpretation needs further evidence.

### 5.1.5 Pragmatic markers as characterization devices

Pragmatic markers *you know*, *I mean* and *like* follow the expected distributional patterns. *Like*, as a perceived incoming marker, is used predominantly to index young, female speakers but can be found used by male characters as well, albeit with lower frequencies. *You know* and *I mean*, as pragmatic markers that are more established, are used much more evenly distributed across age groups, although the stereotypical higher use by women can be observed for all three markers. It could be argued that markers are used for support of specific character traits, such as a low use of marker *I mean* for characters that are particularly self-assured, such as Sherlock from *Sherlock* or Richard from *Gilmore Girls*. *Like* in particular, is used to evoke the stereotype of a *Valley Girl*, with most frequent rates by the same speaker group of young, female, American characters: Haddie and Amber from *Parenthood* as well as Cordelia and Dawn from *Buffy*.

The main result from pragmatic marker use from the Television Dialogue Corpus is the series-specific distribution. In all runs of analyses, the series category turned out to be most significant with *Parenthood* showing distinctive patterns due to their production process.

The series-specific results support the argument that television series as such cannot be classed as an all-inclusive genre. While previous research found that overall patterns on television series are equal, the distinctive use of pragmatic markers that were found in this chapter argue for a more refined categorization of television genres in linguistic analyses.

In terms of the main questions for the analysis of characterization patterns in fictional television, I will summarize findings in the following before turning to the next feature: hedges.

- Does the distribution of markers reflect characterization patterns within and across series?
  - The effect of age, gender, and to some degree, social class, was evident across series. Nationality seemed meaningful in *Buffy*, but to a lesser degree in the British series *Torchwood*. Individual characterization patterns were also reflected in pragmatic marker use, such as self-assurance in characters (e.g. Richard, Amber), pragmatic competence (e.g. Max), or leading/power qualities (e.g. Buffy, Lorelai).
- Does the distribution indicate indexical and/or stereotypical characterization and if, what characters or character groups are represented like this?
  - The most striking stereotypical patterning was that of the *Valley Girl*. This is unsurprising with the character scheme as common and recognizable as it is. A stereotype that needs further investigation is the relation of British patterns with upper class American characters.
  - Character groups, with reference to social networks, show that Amber and Haddie (*Parenthood*) are exhibiting very similar patterns. With both characters being of same age and sex, as well as close to each other socially, it is likely that they are not only indicating the aforementioned *Valley Girl* type, but also a close-knit friendship. Further features will determine whether this is a consistent pattern. A second pair that stands out here is comprised of Sarah and Crosby (*Parenthood*). This is interesting as they are, generation-wise, grouped with Adam and Julia. By linguistically splitting these four characters up, attention is drawn to their distinct personalities. Where Adam and Julia are in control of life, Crosby and Sarah are shown struggling and generally more casual in

their approach to life. This last point in particular matches stereotypical increased uses of pragmatic markers by speakers who are more colloquial.

- Are distributional outliers and unexpected patterns indicative of specific characterization choices?
  - Max, as characterized specifically through his ASD, shows a slight distinction of pragmatic marker use when compared to other *Parenthood* characters. Xander might be an example of a character being used to counter stereotypes, in particular with reference to Buffy, who is in many ways characterized opposing expectations of the typical blonde helpless girl (i.e. *Valley Girl*). Further feature investigations are needed to see if the pattern of these two characters using ‘vice versa’ expected linguistic features holds true.

The next section focuses on the use of hedges *sort of* and *kind of* across the Television Dialogue Corpus.

## 5.2 Hedges

### 5.2.1 Feature definition

George Lakoff first established linguistic features such as *sort of* and *kind of* as part of a linguistic category he called ‘hedges’: “words whose meaning implicitly involves fuzziness<sup>39</sup> – words whose job it is to make things fuzzier or less fuzzy” (1973:471). These hedges can be classed as adverbs that pre- or postmodify a word (or the whole phrase) as ‘fuzzier’, vague, and imprecise.

Example (44) below shows two instances of one of the two hedges that are of interest in this section, *kind of*, from the *Gilmore Girls* sub-corpus. In both cases the hedge modifies the following word and reduces the force of the utterance.

- (44) Sookie: You know, I have a little confession to make. I *kind of* asked you and Rory to both be godmothers because I thought it might *kind of* bring you back together, patch things up. (GG)

Sookie is careful in making a statement and vague language features such as the repeated use of *kind of*, but also of the discourse marker *you know*, of the diminisher *a little*, and modal *might* frame the utterance as hesitant and non-committal.

Both Kay (1984) and Aijmer (1984) claimed that functionally *sort of* and *kind of* are the same, although Aijmer referred to variation in use across varieties of English with *kind of* being preferred in North American English and *sort of* used more frequently in Britain (1984:118). This distinction is also noted by Biber et al. (1999:562) and Gries and David (2007). Quirk et al. (1980) defined the features as compromisers (within the group of

---

<sup>39</sup> Other lexical items he discussed here that are included in the present analysis are pragmatic markers (8.1) and modal adverbs (8.4) on the side of fuzzy, and degree adverbs (8.5) on the side of less fuzzy.

downtoners) and claimed that both *kind of* and *sort of* are more frequently used in American English than comparable compromisers *quite* or *rather*.

Syntactically, these hedges can appear in pre-modification of nouns (45), verbs (46), adjectives (47), as post-modifiers (48), as well as minimal responses (49) (cf. Aijmer, 1984:120-121).

- (45) Xander: It was *kind of* a blur. (BVS)  
(46) Luke: I *kind of* feel like a weight's been lifted, you know. (GG)  
(47) Jack: We're *sort of* busy. (TW)  
(48) Amber: I play guitar, *kind of*. (PH)  
(49) Jack: 24 years old.  
Ianto: *Sort of*.  
Gwen: Well, he's either 24 or he's not.  
Ianto: Depends on how you work it out. (TW)

Quaglio (2009:157), in his study on vague features in television series *Friends* in comparison to naturally occurring language, found that *kind of* was used 895 times per million words and *sort of* 107 times per million words. The Television Dialogue Corpus shows similar overall numbers with *kind of* at roughly 637 per million, and *sort of* at approximately 212 per million words.

His results and the overall distribution in the Television Dialogue Corpus imply that previously mentioned differences in use by American and British speakers might be relevant and observable in scripted language. *Friends*, a series with almost exclusively American characters, shows a higher relative frequency of *kind of*, the American preferred variant. Having characters and series from both varieties as part of the Television Dialogue Corpus, the analysis enables me to see whether the observed regional preference for one variant over the other might be used as a stylistic mark. Included in that is the question of how salient hedges are in terms of stylizing (i.e. whether they are used as

indexical markers). The analysis focuses on the factor groups that may play a role in the use of these hedges overall, as well as the choice between the two variants.

### 5.2.2 *Token inclusion & coding*

For the extraction of the feature from the sub-corpora I decided that it would be impractical to differentiate between the forms *kind of* and *kinda*, as well as *sort of* and *sorta*. With the transcription convention of the individual transcribers not consistent across all episodes, seasons, and series, I collated both variants of each hedge.<sup>40</sup>

In terms of coding the feature for inclusion in the analysis, I again differentiated between tokens that show purely pragmatic functions, and tokens that carried propositional meaning.

In example (50) for instance, *kind of* describes a specific type of love (“the one you deserve”), rather than a vague reference to the concept of love. Here, the feature is not used as a hedging device, but rather a device of specification or typification.

- (50) Angel: And maybe you’ll find your way to *the kind of* love you deserve. (A)

This distinction between hedge and typification is exclusive to tokens occurring in noun phrases, for both features. With regards to tokens that modify noun phrases, Fetzer summarizes the scaled fuzziness component of the feature in terms of its immediate syntactic context: “the anchoring of the object at hand to the scale is not arbitrary but

---

<sup>40</sup> Future investigations might benefit from an analysis of both forms for each variant which includes a detailed phonological transcription. An increased use of the reduced forms may well correspond with an increase of the pragmatic form and confirm processes of delexicalization (cf. other reduced forms with similar trajectories: *you know* (*y’know*))

depends on the semantics of the hedge” (2010:51). She differentiates between tokens that appear with determiners before the hedge, such as “this *sort of* X” or “the *kind of* Y”, which are identifying and describe closeness to prototypicality, whereas hedges with “higher degrees of indeterminateness”, such as “some sort of” or “Ø *kind of* Y” are closer aligned with a peripheral identification which in turn implies vagueness.

This part of the analysis focuses exclusively on the last use (hedge), with a closer look at the relative distribution between propositional and pragmatic uses analysed in chapter 6.

### **5.2.3 *General distributions***

A total of 1450 tokens were included in the analysis, with 1091 tokens for *kind of*, and 359 tokens for *sort of*. A majority of characters included in the main corpus are American, which reflects findings from previous studies about the preferred variant for American English being *kind of*.

Table 6: Overall distributions hedges, tokens per 10,000 words

<b>Character</b>	<b>series</b>	<i>sort of</i>	<i>kind of</i>
Angel		1.56	8.9
Cordelia	<i>Angel</i>	1.54	10.03
Spike		2.93	1.46
Angel		1.81	9.98
Anya		2.64	5.28
Buffy		3.55	10.58
Cordelia		3.75	5.25
Dawn	<i>Buffy the Vampire Slayer</i>	2.47	11.72
Giles		3.09	1.03
Spike		0.93	2.17
Willow		5.62	13.72
Xander		2.91	12.31
Emily		1.6	0.67
Lorelai		1.45	4.19
Luke		1.77	5.53
Paris	<i>Gilmore Girls</i>	1.75	2.5
Richard		2.52	0.56
Rory		0.98	7.58
Sookie		2.15	7.65
Adam		1.03	2.75
Amber		0.95	11.38
Camille		1.12	4.48
Crosby		0	12.49
Haddie		0	9.61
Julia	<i>Parenthood</i>	0	13.54
Kristina		6.57	6.84
Max		0	1.78
Sarah		2.24	5.83
Zeek		0	9.69
John		4.9	1.84
Sherlock	<i>Sherlock</i>	1.19	1.48
Gwen		5.83	1.46
Ianto		5.71	2.86
Jack		2.96	8.02
Owen	<i>Torchwood</i>	2.62	4.58
Rhys		1.99	1.99
Toshiko		2.13	3.19
mean overall		2.28	6.08
	<i>Angel</i>	2.01	6.79
	<i>Buffy the Vampire Slayer</i>	2.97	8.00
mean by series	<i>Gilmore Girls</i>	1.75	4.1
	<i>Parenthood</i>	1.19	7.84
	<i>Sherlock</i>	3.04	1.66
	<i>Torchwood</i>	3.54	3.68
mean by gender	female	2.44	7.13
	male	2.11	4.97
mean by age	young	2.18	8.48
	middle	2.37	8.82
	older	2.52	5.13
mean by nationality	American	1.96	7.51
	British	3.13	2.21

## 5.2.4 Distributions across categories

### Age and gender

Exploring how hedges are used across gender, *kind of* shows an overall frequency preference with female characters (per 10,000 words), albeit not to statistically significant levels.

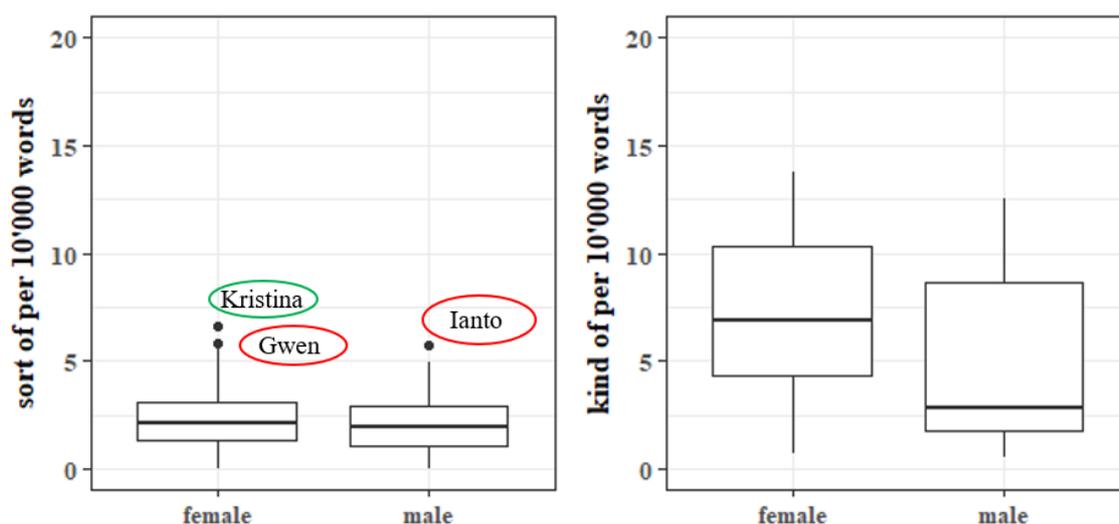


Figure 18: Hedges across gender, per 10,000 words

*Sort of*, in comparison, does not show any preference for either gender. We might find that *sort of* as a vagueness feature is, if used, not as indexically marked as *kind of* (this is further explored in chapter 6). *Sort of* also includes three outlier users (characters that use forms statistically frequent or infrequent when compared to the average use). As a form that previous studies have said to be a British-marked variant, the *sort of* outlier positions of Gwen and Ianto, both from Welsh-based series *Torchwood*, are unsurprising. Kristina, from NBC's *Parenthood*, in comparison, is unexpected. Contextualizing the statistics within the background of the series, one suggested reason might be that the numerical outlier might reflect a literal outlier status of Kristina within the rest of the ensemble cast. The series is



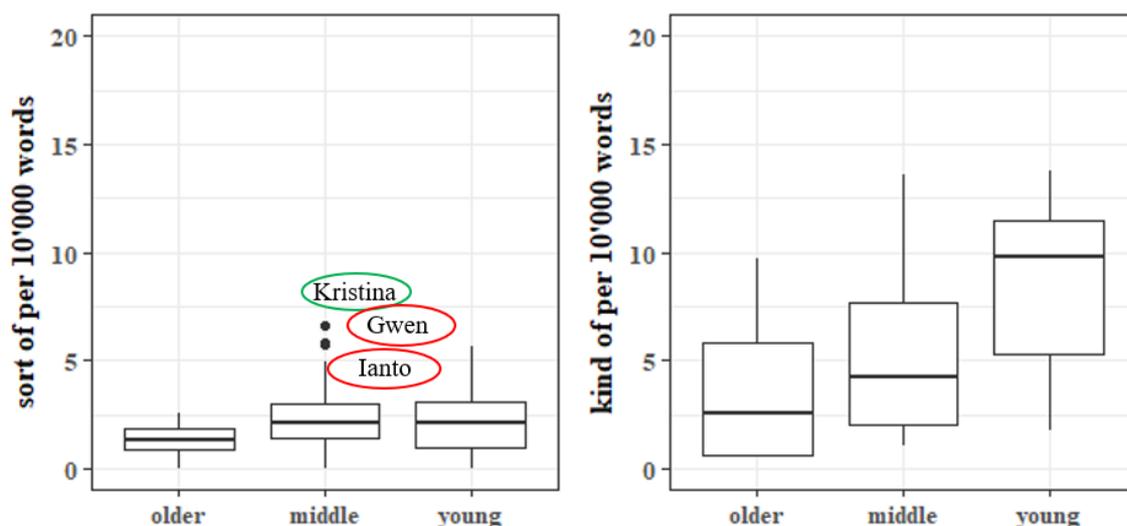


Figure 19: Hedges across age groups, per 10,000 words

With age and gender oftentimes coinciding as factor groups, I combined both categories to see whether there is interaction between the two categories. A cross-tabulation of age and gender shows that female characters are almost always leading in their use of hedges, with the exception of the older characters when using *kind of*. As seen in table 7, as well as Figure 20, *sort of*, as mentioned above, is much less stratified across either gender or age.

Table 7: Hedge use per 10,000 words across gender and age

	female		male	
	<i>sort of</i>	<i>kind of</i>	<i>sort of</i>	<i>kind of</i>
<b>young</b>	2.32	8.76	1.45	7.05
<b>middle</b>	2.91	6.10	2.32	4.65
<b>older</b>	1.36	2.57	1.26	5.13

The surprising incline in the older male character group for *kind of* is grounded in *Parenthood's* Zeek. With this character group only consisting of two characters (Richard from *Gilmore Girls* as well), it might be that rather than representing the older, male character group, the high use by Zeek is dominating this group.

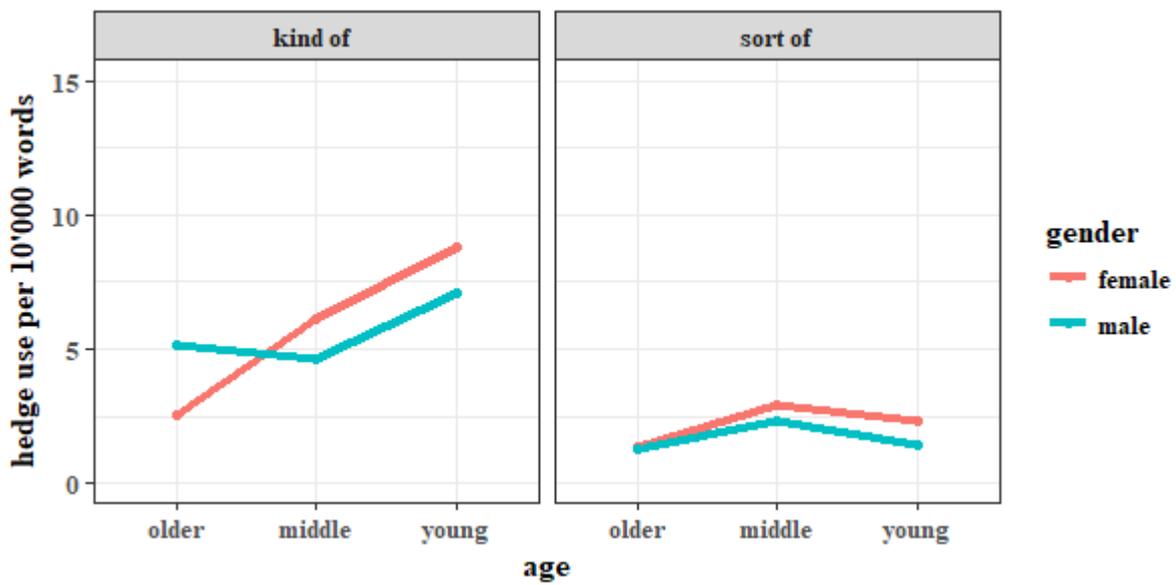


Figure 20: Stratification of hedges across gender and age, per 10,000 words

### Nationality

Regional background of the characters (see Figure 21), as previously indicated, shows *kind of* with a clear American preference, while British characters use more *sort of*. This distribution includes British characters on American series as well as American characters on British series.

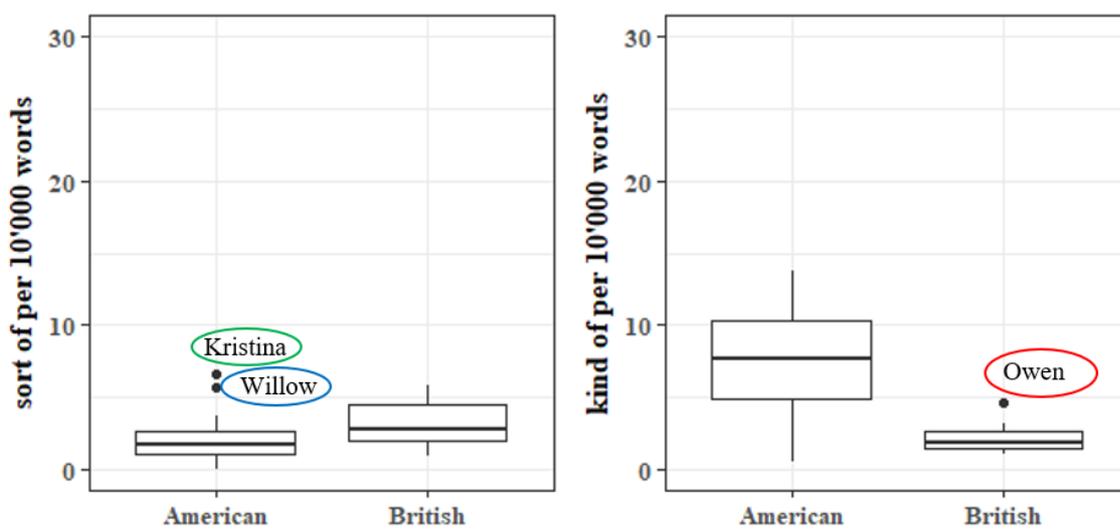


Figure 21: Hedges across nationality, per 10,000 words

For *kind of*, the characterization categorization is much stronger than that of *sort of*, with the former at statistical significance of  $<.001$ . This supports the claim that the two variants *sort of* and *kind of* differ in their saliency and thus in the use as stylization devices. *Kind of*, then, might be a more salient characterization feature, as we have seen that stratification is much more apparent with this feature. *Sort of*, a much more established feature, offers little variation in character groupings, albeit clearly linked to British character backgrounds. When investigating how these two differing variants can be used in interplay with each other for characterization purposes, further character patterns become visible.

Figure (22) displays the preference (per 10,000 words) of each character in using *sort of* and *kind of*. Those on the high end will be more likely to use *kind of*, while those in the negative have a preference for *sort of*.

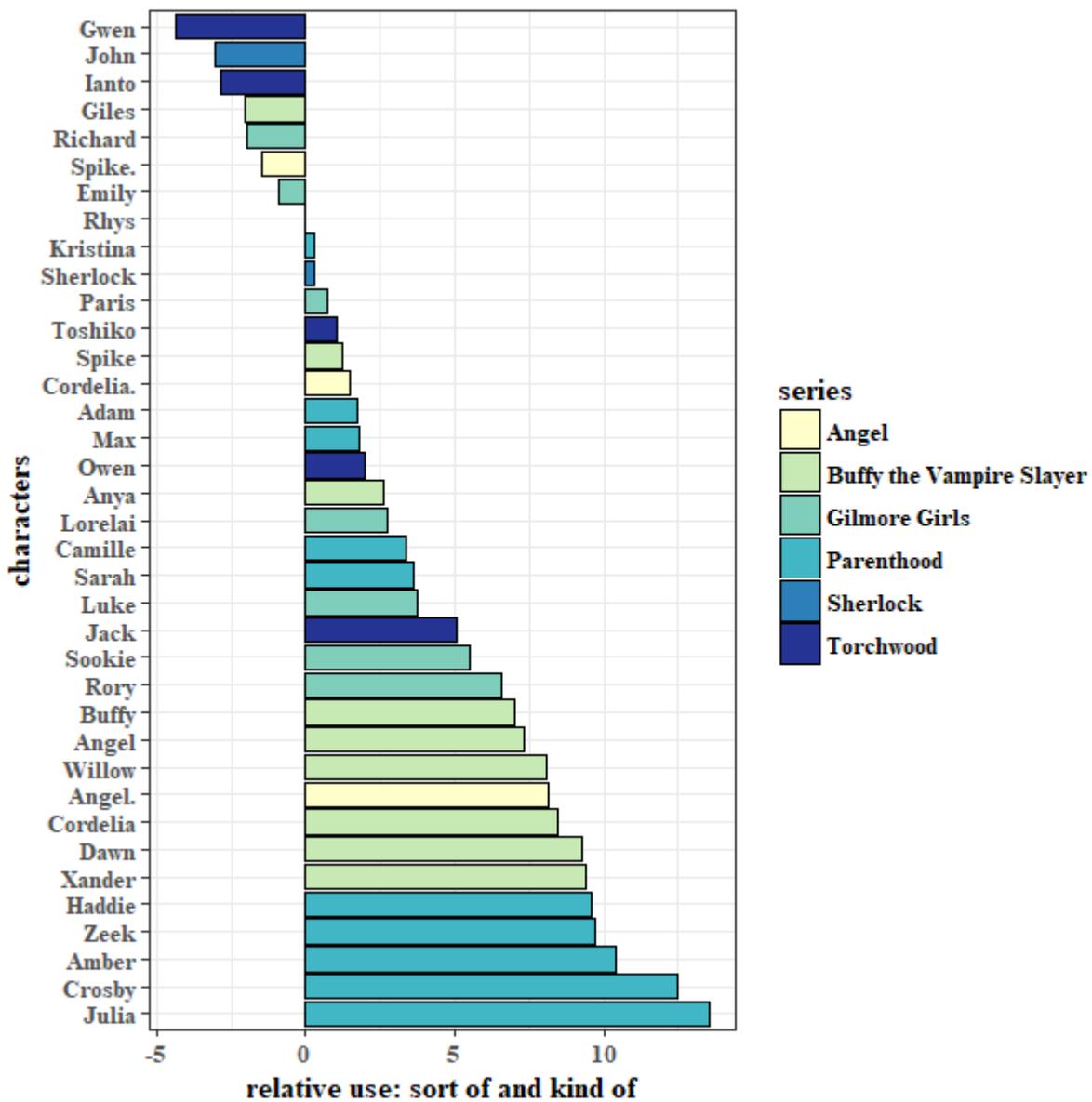


Figure 22: Hedges compared: *sort of* relative to *kind of*

The negative preference (i.e. a more likely use of *sort of* over *kind of*) in the plot above includes characters from five of the six series, an indication that patterns are not series-dependent but rather dependent on similar characterization patterns across series. In connection to previous studies it is revealing to see that most of the characters on this end of the spectrum are English (John, Giles, Spike), or Welsh (Gwen, Ianto, Rhys), supporting again the idea that *sort of* is preferred by British English speakers. This also

further underlines the claim that hedges are salient features that are used knowingly as indicators for Britishness: two of the series (*Buffy* and *Angel*) which show this patterning are American series. While their British characters are presented at the top of the graph (preferring *sort of* over *kind of*), the American characters are found on the other end of the scale – reflecting a clear distinction between nationalities despite being written by the same writing team. In addition to this, we also find Richard and Emily from *Gilmore Girls* favouring *sort of*. This supports the notion of British patterns in American characters indexing socio-economic status as both characters are distinctively upper class and more formal than other characters on the series. This interpretation is partially supported by upper-class Paris. She is much younger than Richard and Emily and would thus be more likely to use the variant that is predominantly used by the youngest generation: *kind of*. Her actual hedge use, however, is much more levelled than that of other characters her age: Rory, Buffy, or Willow. I suggest that her social class status (and the associated British-influenced forms) have an effect on her linguistic patterns here.

On the other end of the scale most characters are American (including the American Jack from the UK-based *Torchwood*), mirroring expectations of a preference for *kind of* over *sort of*. It is interesting to see that for hedges we find a distinctive pattern separating *Torchwood*'s Jack from the British ensemble while this distinction was not visible for the previously analysed pragmatic markers. It might be that hedges are a more salient feature for the writers in distinguishing characters' nationalities. The scale also shows that the characters who use *kind of* the most are from *Parenthood*. The series' production process is much less script-true than other series and actors are found to improvise lines frequently. With regards to research saying that scripted language is less vague than naturally occurring language (cf. Bednarek; 2012:48-49 and Quaglio; 2009:68), this might be a case of production-hybridization, where the dialogue is scripted but actors are

adding spontaneous features to the performance (as we have also seen with pragmatic markers). In fact, *Gilmore Girls*, which has been discussed as being very script-true, shows a much lower overall rate for hedging than the comparable family-centred series *Parenthood*, with 2.92 to 4.51 per 10,000 words respectively.

A final interpretation that can be drawn from this visualization is the development of character. We see Cordelia shift from being in the lower third of the plot (using more *kind of* with respect to *sort of*) when in *Buffy* to the upper part (using less *kind of* with respect to *sort of*) when in *Angel*. A possible reason for this shift might be character development over time. With a shift that was observed for pragmatic markers as well, we can assume that linguistic changes are purposefully used to imply character developments. Further analyses of other features are needed to fully confirm this.

#### **5.2.5 Hedges as characterization devices**

In summary, the distribution of hedges *kind of* and *sort of* follows expected patterns: American speakers favour *kind of*, whereas British speakers favour *sort of*. Further, this distinction can be traced across series, indicating that this is a general variable that is used to index national backgrounds. British characters on American series, and an American character on a British series also adhere to this patterning, so it could be argued that the distinctive use is a feature script-writers employ to distinguish characters' backgrounds. Additionally, we find a shift of overall hedge use across age groups, with younger speakers more likely to use hedges overall. This likely correlates with stereotypes of perceived youth language that are also visible in distributions of other discourse features (pragmatic markers in the previous chapter, as well as intensifiers in chapter 5.5). The preference for the British variant by some American speakers might point to an additional characterization tool; by using the seemingly more formal variant that is not associated

with American teenage talk, i.e. *sort of* over *kind of*, characters are portrayed as more formal and upper class in comparison to their fellow cast. A closer analysis of further features will highlight if this is a conscious effort put into linguistic patterning or if it is limited to hedge use only. With reference to the main questions for the analysis of characterization patterns in fictional television, I will summarize findings for hedges *sort of* and *kind of* in the following before turning to the next feature: general extenders.

- Does the distribution of markers reflect characterization patterns within and across series?

Main characterization patterns for hedges concerned national background. British characters showed a preference for *sort of* over *kind of*, vice versa to American speakers' patterns. Patterns of formal, upper class American patterns correlating with British feature uses, as seen with pragmatic markers before, were followed here as well.

- Does the distribution indicate indexical and/or stereotypical characterization and if, what characters or character groups are represented like this?

As suggested in previous research, certain character groups were more likely to use higher frequencies of hedges (female speakers overall, younger speakers overall, American characters specifically *kind of*).

With regard to social networks and character groupings that refer back to the previously introduced sociograms, the analysis of hedges exemplify predominantly out-groups (e.g. Kristina from *Parenthood*, further described below). A remarkable social group that distinguished itself from other characters from the series are *Gilmore Girls'* Richard and Emily. Their

unique use of hedges not only established them as very similar to one another, but equally as distinct from the other generations present in the series.

- Are distributional outliers and unexpected patterns indicative of specific characterization choices?

Previously investigated outgroups were again characterized through hedge use differing from other character. There are two types of out-groups that can be observed here. With Kristina from *Parenthood*, I found her use of hedges against expectation. While in line with her overall representation as outside the Braverman clan, sociolinguistically her use should have been close to that of Sarah or Julia. Here, the linguistic choice seems largely based on her social styling. Opposed to that, Giles from *Buffy* uses hedges as distinct from other characters but in line with sociolinguistic expectations (being English in an otherwise overall American ensemble). Here, styling appears to be based not only on his outgroup status, but also on known dialect differences. Outliers are indicative of differing characterizations, albeit the kind of difference can alter from series to series.

The next section analyses the use of general extenders *or something, or anything, and stuff*, and *and everything* across the Television Dialogue Corpus.

### 5.3 General extenders

#### 5.3.1 Feature definition

General extenders (GE), also called vague category identifiers (Channel 1994), extension particles (Dubois 1992), or approximation markers (Ermann 1995), are clause-final markers that “generalize from a preceding referent to the larger group of items to which that referent belongs” (Tagliamonte & Denis 2010: 335).

GEs combine a conjunction (usually *and* or *or*) with a generic noun. The inclusion of the conjunction, while most frequent, is not prerequisite for the GE function, as examples (52)-(56) show. Additionally, some GEs will also include qualifiers preceding the noun and/or comparatives following the construct.

Examples from the Television Dialogue Corpus:

- (52) Lorelai: Stella, do something. Show yourself. Molt or chirp *or* something. (*GG*)
- (53) Dawn: Just, you know, with, with the lips and, and the pressing together *and stuff?* Big expert here. (*BVS*)
- (54) Zeek: Yeah, I mean, you know, I, it was a good job. I really liked it, but heck, you know, it's not like I want to do it the rest of my life *or anything*. (*PH*)
- (55) Angel: Everybody always trying to expand their horizons, actuate their potential, and all that other touchy-feely crap. (*A*)
- (56) Sherlock: Some secret societies used to send dried melon seeds, orange pips, things like that. (*SH*)

In example 55, we find Lorelai looking for an escaped baby chick (Stella), imploring it to take an action a chick would be associated with (*molting, chirping*) but not providing a full list of chick-related activities (for instance *pecking*). Previous research has claimed

that GEs are undergoing grammaticalization/pragmaticalization processes in which their original meaning is lost (bleached) and their functions become increasingly pragmatic (cf. Cheshire (2007), Tagliamonte & Denis (2010), Pichler & Levey (2011)). In brief, a GE construction is not merely used to imply an incomplete category set, but to multi-function within the communicative event as, for instance, a signal for turn taking or topic shifts. Further, they can be used to imply shared knowledge between the speakers, as well as affirm group memberships. GEs are oftentimes included in vagueness marker classifications, as they do not provide a specified continuation of a list but rather a general pointer to the incompleteness (in terms of counts as well as in terms of specification) of what the speaker is providing. In example (56), the GE is used to highlight Dawn's awkwardness when it comes to talking about kissing. She is vaguely describing what constitutes a kiss and her statement is intended to humorously juxtapose her following statement of being the 'big expert'.

While the contexts in which GEs can appear are quite varied, the overall structure is similar and the core meaning (cf. core meanings in pragmatic markers, as discussed above) can be summarized as approximation of categorization that implies increased vagueness or non-committal to preciseness (cf. Dines 1980:19).

Sociolinguistically, GEs are claimed to be "conditioned by social factors such as age, sex, education, and socioeconomic class (Dubois 1992; Stubbe & Holmes 1995)." (Tagliamonte & Denis 2010: 336). With the exception of education, the Television Dialogue Corpus will be able to shed light on all these variables and in what way they might be used strategically for characterization purposes.

### 5.3.2 *Token inclusion & coding*

Similarly to other pragmatic features included in the present study, I will provide only a sample of features representing the non-exhaustive list of possible general extender constructions. Lists of common GEs from Cheshire (2007) and Tagliamonte and Denis (2010) were taken as initial selections for feature extraction. In order to investigate reoccurring patterns in characters' speech, I then limited the present analysis to the four most frequently occurring features:

- *or something*
- *or anything*
- *and stuff*
- *and everything*

*And everything* and *and stuff* both describe adjunctive variants (meaning that the GE is extending the general list), whereas *or anything* and *or something* are disjunctive (meaning that the GE is giving an alternative to the general list).

### 5.3.3 *General distributions*

The overall distribution as shown in table 9 below already indicates that GEs are not as frequently used as some of the other features discussed in the present research. It is for that reason that while initial analysis will investigate each of the GE individually, it might be fruitful to also have a general GE use to compare across characters ('GE general').

As indicated in Table 8 below, Amber, Haddie (*Parenthood*), as well as Cordelia and Dawn (*Buffy*) are leading in the use of GEs overall. This possibly indicates yet another young, female, and American indexical use. Contrary to previously analysed features, however, the variant preference is different for the respective characters, i.e. there is not one GE that is used to index the above characteristics. How the use of GEs is distributed across different social character groups is further discussed below.

Table 8: General extenders distribution, per 10,000 words

character	series	<i>or something</i>	<i>or anything</i>	<i>and stuff</i>	<i>and everything</i>	GE total
Angel		1.11	0.33	0	0	1.4
Cordelia	<i>Angel</i>	1.54	1.35	0.4	0.4	3.7
Spike		1.46	0	0	0	1.5
Angel		0.91	0	0	0	0.9
Anya		3.52	0.44	0	0.4	4.4
Buffy		1.89	0.39	0.6	0.5	3.4
Cordelia	<i>Buffy the</i>	2.25	1.5	2.2	3	9
Dawn	<i>Vampire</i>	3.08	1.23	3.7	0.6	8.6
Giles	<i>Slayer</i>	0.62	0	0	0.2	0.8
Spike		0	0	0.3	0	0.3
Willow		3.31	0.5	2.5	1.5	7.8
Xander		1.37	0.34	0.3	0.5	2.6
Emily		0.53	0	0	0.1	0.7
Lorelai		2.24	0.63	0.3	0.4	3.6
Luke		2.94	0.47	0.1	0.5	4
Paris	<i>Gilmore Girls</i>	1.25	0.75	0	0	2
Richard		0.56	0	0	0.3	0.8
Rory		2.02	0.55	0.2	0.8	3.5
Sookie		1.43	0.24	0.5	0	2.2
Adam		1.55	0	0	0	1.5
Amber		6.64	0	3.3	0	10
Camille		0	2.24	0	0	2.2
Crosby		5	0.25	1.2	0	6.5
Haddie	<i>Parenthood</i>	6.41	3.2	0.8	0	10.4
Julia		2.26	0.56	0	0.6	3.4
Kristina		1.05	0.26	0	0	1.3
Max		1.78	0	0	0	1.8
Sarah		2.47	0.67	1.3	1.1	5.6
Zeek		2.04	0.51	0	1.5	4.1
John	<i>Sherlock</i>	1.22	0	0	0	1.2
Sherlock		0.3	0	0	0	0.3
Gwen		0	0.49	1	0.5	1.9
Ianto		2.86	0	0	0	2.9
Jack	<i>Torchwood</i>	0	0	0	0.4	0.4
Owen		0	0.65	0	0	0.7
Rhys		1.99	0	2	0	4
Toshiko		0	0	0	0	0
mean overall		1.83	0.47	0.56	0.36	3.23
	<i>Angel</i>	1.37	0.56	0.13	0.13	2.19
	<i>Buffy</i>	1.88	0.49	1.08	0.75	4.2
mean by series	<i>Gilmore Girls</i>	1.57	0.38	0.15	0.3	2.39
	<i>Parenthood</i>	2.92	0.77	0.67	0.32	4.68
	<i>Sherlock</i>	0.76	0	0	0	0.76
	<i>Torchwood</i>	0.81	0.19	0.49	0.15	1.64
mean by gender	female	2.2	0.79	0.89	0.52	4.4
	male	1.43	0.14	0.22	0.19	1.98
mean by age	young	2.92	0.85	1.18	0.64	5.59
	middle	1.4	0.22	0.32	0.18	2.11
	older	0.78	0.69	0	0.49	1.96
mean by nationality	American	2.19	0.61	0.65	0.47	3.92
	British	0.85	0.11	0.33	0.07	1.36

The most frequently used GE of the four variants is, throughout all series, *or something*; in *Sherlock* it is the only GE used. In comparison, Quaglio (2009), who included three of the variants in his research on language in television series *Friends*, found similar distributions (see comparison table below).

Table 9: Comparison GE use: Television Dialogue Corpus and Quaglio (2009)

	<i>or something</i>	<i>or anything</i>	<i>and stuff</i>
<i>Angel</i>	1.37	0.56	0.13
<i>Buffy the Vampire Slayer</i>	1.88	0.49	1.08
<i>Gilmore Girls</i>	1.57	0.38	0.15
<i>Parenthood</i>	2.92	0.77	0.67
<i>Sherlock</i>	0.76	0.00	0.00
<i>Torchwood</i>	0.81	0.19	0.49
<i>Friends</i> (Quaglio 2009:157)	1.30	0.46	0.68

He states that GEs belong to the “most obvious markers of vagueness in conversation” (2009:74) and, while not as frequent in scripted contexts as in naturally occurring language, they are used in, I suggest, characterization building capacities. Similarly to the discussion of pragmatic marker *you know* and mentioned in the feature definition above, GE are used to evoke shared knowledge between the speaker and listener. In television the listener role is two-fold: Firstly, the characters on screen that are communicating with each other; secondly, the characters on screen that are communicating for an audience. The latter communication channel is not reciprocal, which makes any construction of shared knowledge difficult. However, implied shared knowledge (and thus a level of familiarity and intimacy) is necessary for the television series to create in order to keep audiences engaged. The lower rate of vague language we find in scripted fictional television is, I claim, not only due to it ‘slowing down’ the dialogue and adding seemingly unnecessary words to a timed conversation, but also because the writers (and actors) need

to be consciously aware of how much shared knowledge with the audience can be assumed (cf. Quaglio 2009:78).

#### **5.3.4 Distributions across categories**

##### *Age and gender*

The below Figure (23) displays the distribution for each of the variants across gender. The outliers are again highlighted with the name of the character as well as colour-coded for the series.

All four variants show a higher frequency use of GEs by female characters. It is the outliers here that are of particular interest. As with features in previous sections of the analysis, characters from *Parenthood* are dominating the outliers. As with hedges and pragmatic markers, the use of GEs by characters further supports a claim that *Parenthood* is a hybrid form between naturally occurring language and scripted dialogue. Notably, for the adjunctive variants (*and stuff*, *and everything*), we also find characters from *Buffy* and *Torchwood* in the outliers.

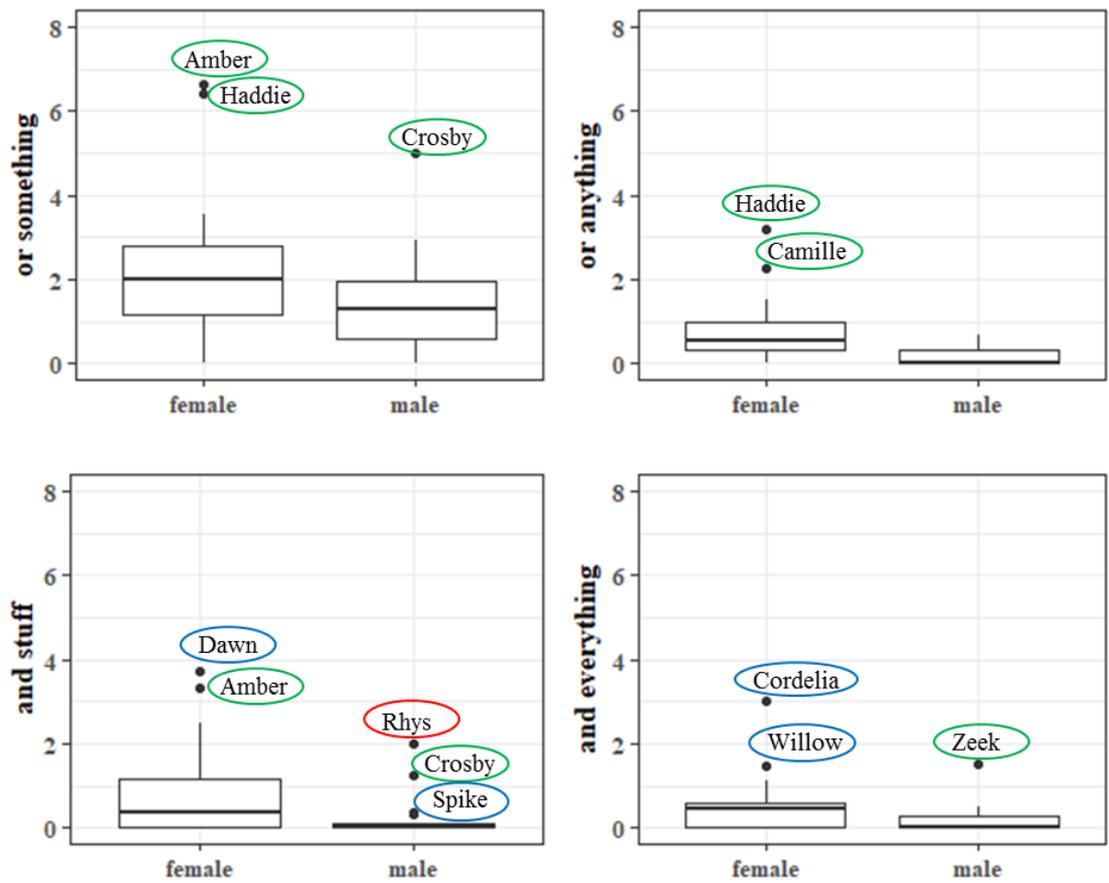


Figure 23: General extender use across gender, per 10,000 words

Considering these findings in combination with possible patterns across the three age groups gives some support to the claim that GEs are features increasingly used by younger characters. Particularly with *or something*, we can see a steady decline with age.

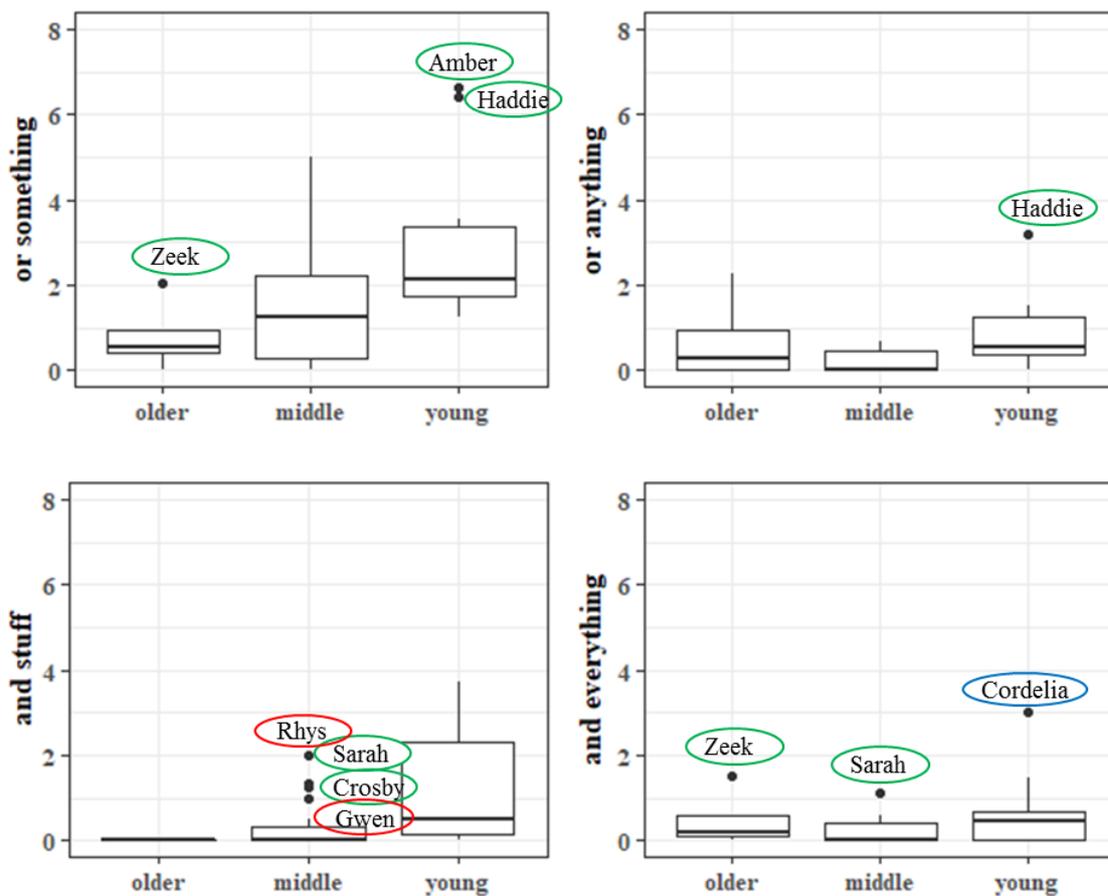


Figure 24: General extenders across age, per 10,000 words

The outliers for *and stuff* for the middle age group might be an indicator that age is not a characterization category for this variant in general; similarly unclear are distribution patterns for *and everything*.

### Nationality

Nationality as a characterization category has been found, for previous features, to be a rather salient factor group. The general pattern for GEs is that American characters use GEs more frequently, though no variant shows alternative preference patterns as we have seen with hedges for instance.

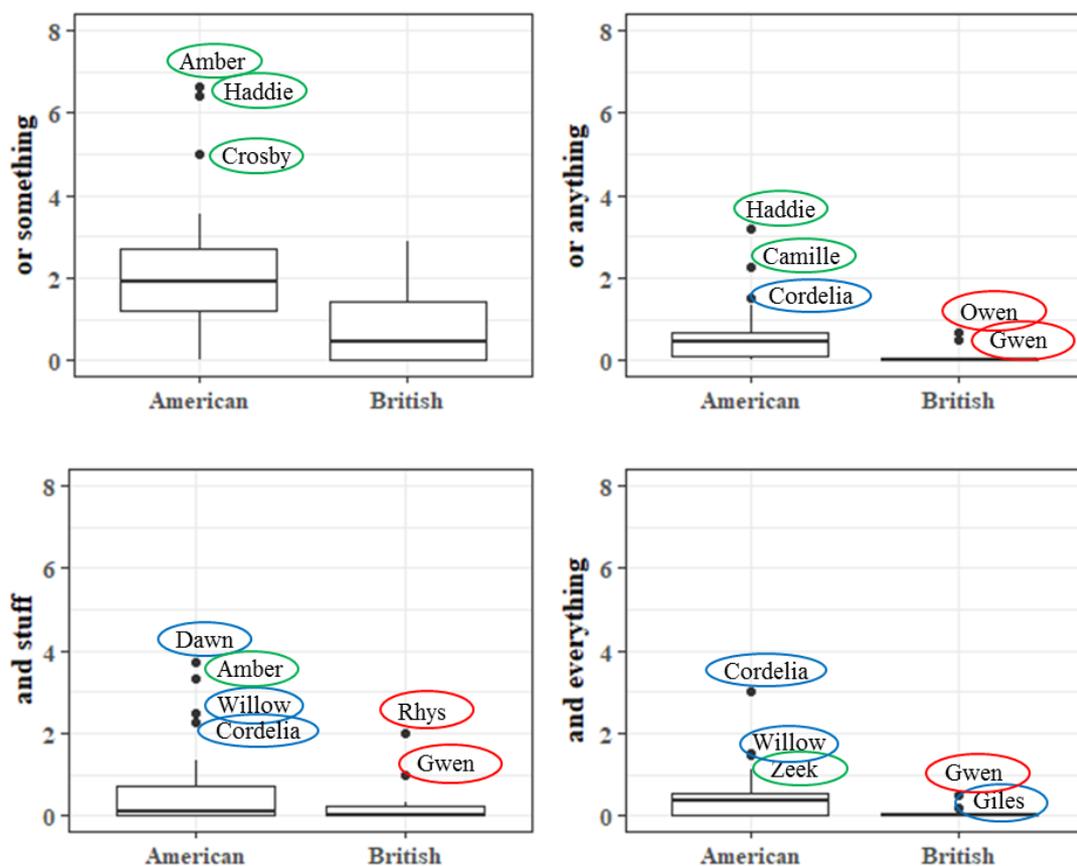


Figure 25: General extenders across nationality, per 10,000 words

A characterization pattern that is worth investigating here is the continuous distinction between American and British characters. The distributions we have seen so far, including the present GE distribution, shows that British characters on British series are using pragmatic features differently (in terms of which variants, as well as at what frequencies). This might point towards a possible distinction in how British series are scripted, or how British writers are creating their dialogue: mostly in accordance to what British English would appear like in naturally occurring language. However, this interpretation does not account for the pattern we find for American characters on British series and vice versa. Britishness is clearly indexed through distinctive patterns on American series and these patterns map onto what we find in British series by British characters. Thus, what

television series present in terms of national background is not only part of the local dialect (for British series) but also what is perceived to be British (for American series with British characters).

Patterns of national background, clearly adding to characterization in the fictional television contexts, are thus observable across television series and not, such as other factor groups, depending on the kind of series and how it is produced.

A further support for this claim can be seen in the below plot (26) that gives the character-based use of GE overall for all series. Similarly to the hedge distribution in section 5.2 (plot (22)), we can see that the upper quarter of the plot is dominated with British characters from three series (*Torchwood*, *Sherlock*, *Buffy*) as well as upper class American characters (*Gilmore Girls*), a characterization pattern congruent with previously investigated features.

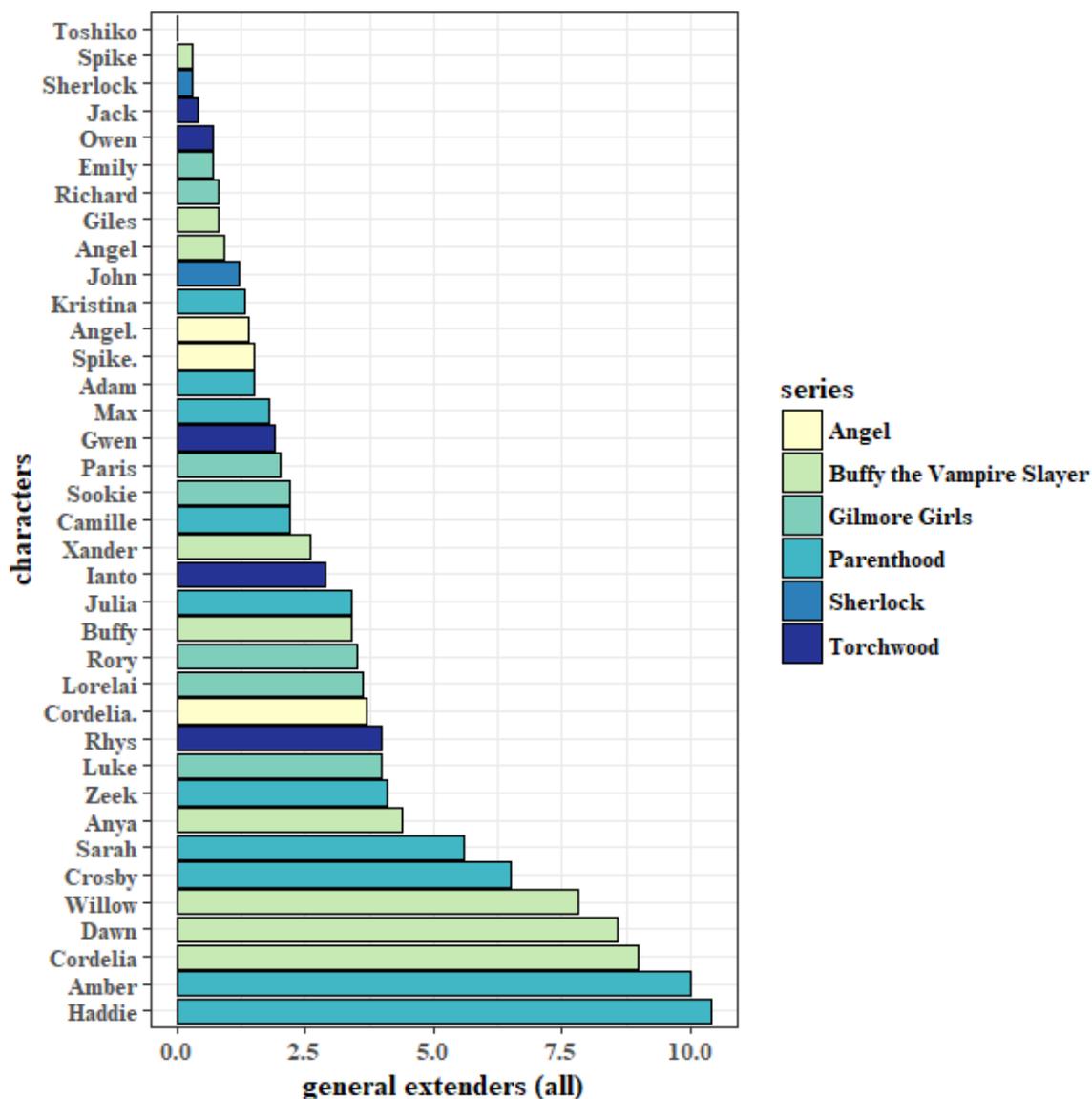


Figure 26: General extenders across characters, per 10,000 words

The characters producing the most GEs are *Parenthood's* Amber and Haddie, as well as *Buffy's* Cordelia and Dawn, as previously noted, indexing the particular character group of young, female, and American *Valley Girl*. Interestingly, and in line with what was found with hedge *kind of*, Cordelia shows character development in her use of GEs, while in *Buffy*, and still stereotyped as the typical popular Californian cheerleader, she uses significantly more GEs than when in *Angel*. A similar development can be seen with

Spike, who, in *Buffy*, is very much characterized as the rebellious (at times evil) outsider of the group. His vagueness features are rather low, partly due to his British background, but arguably also due to his characterization of being straight-forward (to the point of rudeness) and diminished mitigation/politeness attempts. In *Angel*, as can be seen in the above plot, he shifts his character to being part of the group and less of a rebel outsider, showing similar GE use to Angel.

Interesting here is that Max, who was characterized throughout with differing frequencies from the rest of the *Parenthood* ensemble, shows very similar rates to his immediate family Kristina and Adam. Possibly, GEs do not carry as much indexical possibility as hedge *kind of*, or pragmatic markers, and are thus not used here to characterize Max's diminished pragmatic competence.

### 5.3.5 *General extenders as characterization devices*

General extenders have further supported some of the characterization patterns we have found with the previously analysed features. The main factor group that is consistently reflected in distinctive language use is a character's national background. In fact, as I have shown, national background is shown not through individual series' production location, but is apparent for nationalities independent of where the series is created. Further, patterns are observable across series and genres, indicating that there appears to be a general perception of what British characters and what American characters sound like. General patterns of age and gender are the increased use of vague features by younger, as well as by female speakers. Character-specific uses of GE additionally showed the possibility to indicate character development over time (as with Cordelia and Spike in *Buffy* and *Angel*).

- Does the distribution of markers reflect characterization patterns within and across series?
  - Characterization patterns of nationality are observable across series. Gender and age, to a certain degree, follow previously seen patterns for *Valley Girl* stereotypes. Varying production means in terms of scriptedness are seen here as well, with many of *Parenthood*'s characters leading in GE use regardless of other character categories (such as age and gender).
- Does the distribution indicate indexical and/or stereotypical characterization and if so, what characters or character groups are represented like this?
  - The stereotype of *Valley Girl*, as mentioned above, bears out in the use of GEs. A development of character away from stereotypical characterization can be observed through a decrease/increase of GE use (*Buffy*'s Cordelia and Spike respectively).
  - With regard to social networks, I again found Amber and Haddie (*Parenthood*) to be quite similar in use. In addition to that, Cordelia and Dawn from *Buffy* have comparable frequencies which leads me to believe that the use of GEs by these four characters (all young and female) is indexing a character type (see above). Other social relatedness is observed with Sarah and Crosby (cf. pragmatic marker use) and Emily and Richard (cf. hedge use). Repeating patterns of social networks across various feature groups indicate that the characterizing closeness among characters is a salient marker that is linguistically expressed.

- Are distributional outliers and unexpected patterns indicative of specific characterization choices?
  - Outliers here are congruent with characterization choices of stereotypical language uses (*Valley Girls*: Amber, Haddie, Cordelia (*BVS*), Dawn), as well as production means (*Parenthood*).

## 5.4 Modal adverbs

### 5.4.1 Feature definition

The final category of features indicating vagueness or a lack of commitment is the category of modal adverbs (e.g. *maybe, perhaps, probably*). Terminology, again, is varied for these pragmatic devices and includes, among others, epistemic modal markers (Kranich 2011), epistemic stance adverbials (Biber et al., 1999), plausibility shields (Prince et al., 1982), or modal adverbs (Coates, 2003; Fraser, 2010), as they will be called here.

Biber et al. (1999: 854-856) categorize modal adverbs as part of a group of stance adverbials indicating doubt and certainty. In comparison, hedges discussed in previous sections (*sort of, kind of, like*) are categorized as imprecision markers. The main difference here is the source of fuzziness. Consider examples (57) and (58) below:

(57) Angel: Our friend, she's under some *sort of* spell. (A)

(58) Haddie: Like, *maybe* her friend liked you. (PH)

In (57), Angel hedges the likeliness of the spell, acknowledging that there is some semantic membership of 'spell' that is fulfilled, yet not precisely attained. In (58), conversely, Haddie hedges the likelihood, or probability, of her utterance. She does not call into question the proposition value, but rather adjusts her own positioning or belief toward it. Both hedges attenuate (part of) the utterance, although in slightly different ways. Fraser (2010:22) distinguishes between these hedges as propositional and speech act hedges respectively, where the latter are used to "signal a lack of [...] full commitment to the force of the speech act being conveyed". This distinction can be traced to Prince et al. (1982), who categorize these two types of hedges as approximators (propositional

hedges) and shields (speech act hedges). Their notion of approximators is further separated into adaptors (such as *sort of*, see section 5.2) and rounders (such as approximator *like*, see section 5.1). Shields<sup>41</sup> (speech act hedges) have the effect “that the speaker has implicated that s/he is not fully and personally committed, i.e. committed in the usual or ‘unmarked’ way, to the belief that the relevant state of affairs actually obtains” (1982:11). Brown and Levinson (1978), who developed the notion of speech act hedging, refer to the use of these hedges as “a primary and fundamental method of disarming routine interactional threats” (1978:146, in Fraser 2010:19). More detailed pragmatic functions and how they relate to characterization processes are explored further down.

While there does not exist an exhaustive list of shield markers, modal adverbs are a relatively contained group of features that can be functionally defined as equivalent. The features under investigation in this section are *probably*, *maybe*, and *perhaps*<sup>42</sup> (other, less frequent items are, for instance, *possibly*, *presumably*, *conceivably*).

Previous work on modal adverbs has mainly focused on semantic analyses (cf. Suzuki, 2015) and how they can be defined along the scales of epistemic modality and evidentiality (Cornillie, 2009). Across linguistic studies, assignments of concrete functions and possible syntactic contexts are called into question and re-categorised at various points (see for instance Suzuki’s summary of distinctions between subjective and

---

<sup>41</sup> Prince et al. (1982) further distinguish between plausibility shields and attribution shields. The latter refer to hedges that transfer the source of epistemic stance onto a third entity (e.g. *according to*, *consistent with*, etc.). These markers will not be relevant in this study.

<sup>42</sup> The initial analysis also included *possibly*. Due to very low frequencies however, I decided to omit it from the final version of this section. While adding to the functional class of vague modal adverbs and thus providing additional insights into the use of these features, found frequencies could not be interpreted as speaker patterns as the numbers were too low to disregard the notion of chance.

objective stance (2015: 1368-1369)). Despite varying terminology<sup>43</sup> and disagreements on detailed function, modal adverbs are generally investigated with reference to their epistemic stance and resulting attenuation of speech. While the hedges *sort of* and *kind of* modify the head of the clause in terms of prototypicality, the modal adverbs here are indicating the speakers' position (or stance) toward their utterances as a whole.

In the examples (59) to (61) below, varying effects of attenuation are apparent.

- (59) Lorelai: *Maybe* Milan Kundera is the Robin Williams of the Czech Republic. (GG)
- (60) John: I think he was *probably* joking. (SH)
- (61) Giles: If there were deaths, then, uh, *perhaps* we're dealing with a fairly ... standard haunting. (BVS)

In using *maybe*, *probably*, and *perhaps*, the characters are not only acknowledging the likelihood that what they are saying might be untrue, but they are also using mitigation techniques to help the communicative flow. Using *maybe* in example (59), Lorelai is introducing an absurd statement for comedic effect. While she probably would not have caused offense without the modal adverb in place, the inclusion indicates that she is rather obviously speculating and does not expect her utterance to be taken at face value. For comparison, example (60) shows how a modal adverb can help release possible tensions between speakers. John could have instead stated: "He was joking", fulfilling the same truth value as in the example above. However, in using hedging devices *I think* and *probably*, he offers the listener the chance to catch up with a previously misunderstood joke without them losing face. What is apparent in this example is that the multifunctional

---

<sup>43</sup> Fraser (2010: 21) provides an extensive list of terms that hedges have been previously referred to, including adaptors, agent avoiders, bushes, diffusers, politeness markers, etc.

aspect of modal markers is important to consider. The speaker is not always providing a semantic reference marker to their uncertainty, but a discourse managing device that allows all participants to evaluate the utterances under non-threatening circumstances. Even with absolute certainty that what John is referring to is indeed a joke, the modal adverb provides vagueness for the sake of the listener (both within the fictional and the actual world). In (61) then, Giles is indicating a certain probability of his utterance being true, showing that he cannot fully commit to his suggestion. Here, the function of the modal adverb is to provide a clear marker of uncertainty from the speaker's perspective. The distinction between speaker- and listener-oriented hedging (as in (60) and (61) respectively) is not always clear and oftentimes functions of hedging are overlapping. In (62) below, Julia uses the modal adverb *maybe* in a clause ('it's corny') that justifies her emotional stance ('it meant a lot to me').

(62) Julia: And *maybe* it sounds corny, but it meant a lot to me. (PH)

The marker thus enables her to distance herself from a possibly face threatening statement (in that it is personal and intimate) by acknowledging its emotional weight. At the same time she is offering the listener the possibility to regard her admission as 'corny' or not serious. *Maybe* functions as a strategic device to lessen the commitment that Julia makes by adding a sense of vagueness.

In terms of characterization, modal adverbs can be used to suggest that some speakers are more or less willing to commit to their utterances with full force, whether this is because of uncertainty about the propositional content, or as an interaction managing device. Characters that have heightened levels of insecurity (or are more aware of possible face threatening situations) are more likely to make use of these modal adverbs while characters that are very self-assured and convinced of their thoughts may use modal

adverbs to a lesser degree. Landert (2017:498), who looks at stance in character dialogue, notes that markers of uncertainty make the character “appear less confident” and that, overall, epistemic stance markers’ “strategic use has quite a distinct effect” (2017: 499) in terms of characterization of insecure or hesitant personas. She however does not specify the epistemic stance markers and while modal adverbs are included in the general category of characterizing devices that are under investigation there, their concrete role is yet to be analysed.

With reference to pragmatic competence, a characterization aspect found important for other markers, Fraser (2010: 30) writes that “hedging, in American culture (and I assume most others) is necessary in many circumstances, lest the speaker be perceived as impolite, offensive, or arrogant”. Extrapolating this observation, a low frequency use of modal adverbs might also point to the characters that were previously found to struggle with interpersonal communication and/or have been established as diagnosed with ASD. This adds to previous sections where social and relational characteristics were explored and found to be meaningful for characterization purposes.

While there are numerous studies that include these devices in their investigations, surprisingly little is known of the sociolinguistic patterning of *probably*, *maybe*, and *perhaps*.

Some studies on epistemic stance markers, including modal adverbs, claim that they are more frequently found in women’s speech (cf. Coates, 2003), but concrete quantitative evidence is most of the time given through other features (Holmes, 1984 on *you know* and *sort of*, Coates 1987 on a range of forms including *I mean*, *well*, *just*, *I think*). Reasoning for this gender distinction is claimed to be due to the conversational structure and different nature of topics discussed rather than what Lakoff (1975:54) claimed to be a sign of

women's unassertiveness and socialization through some form of assumed femininity status. However, Lakoff's notions might point toward stereotypes of gender in society that are replicated in the media.

Biber et al. (1999: 569) find that of the three markers, *probably* is used most frequently in conversational English, followed by *maybe* and *perhaps*. In terms of regional variation, previous studies have found modal adverbs (particularly *maybe*) are more frequently used in American English. *Perhaps*, in contrast, appears to be used more in British English, albeit still with low overall frequencies in both varieties.

The following analysis thus not only focuses on how modal adverbs are used by each character and within each series, but also whether there are distinct patterns of speaker preference for each individual marker or if an overall modal adverb use (including all markers) is indicative of characterization.

#### **5.4.2 *Token inclusion & coding***

Modal adverbs can occur in clause initial, medial, and final positions, as well as in minimal responses. Because these pragmatic devices can be delimited to clearly defined lexical items that are unambiguously modal adverbs (unlike other features discussed previously), there were no particular exclusion criteria for any token types.

Albeit, as mentioned above, the individual features can be multifunctional, I refrained from a detailed functional categorization. The main reason for that lies in the interpretation of functionality and the fact that the devices are seldom used with one concretely definable communicative purpose.

### 5.4.3 General distributions

Overall, I included 3339 tokens in the analysis with 2350 tokens for *maybe*, 838 for *probably*, and 151 for *perhaps*.

A shift in preference is visible when comparing the raw frequencies found in Biber et al.'s (1999) analysis of modal adverbs in the Longman Grammar Corpus. In the conversational data that is under investigation in Biber et al. (1999:869), preference for *probably* is above *maybe* (for both American and British English; both varieties combined in the below table) whereas the Television Dialogue Corpus shows *maybe* over *probably*. For both data sets *perhaps* is least frequently used.

Table 10: Each □ represents 100 tokens per million words, table adapted from Biber et al. (1999: 869)

	CONV (AmE & BrE)	Television Dialogue Corpus (AmE & BrE)
<i>probably</i>	□ □ □ □ □ □ □	□ □ □ □ □
<i>maybe</i>	□ □ □ □ □	□ □ □ □ □ □ □ □ □ □ □ □ □
<i>perhaps</i>	□ □	□

A point to note here is that this shift is not necessarily context- or genre-bound, in that naturally occurring conversational speech shows different patterns from scripted television language, but that the composition of the data set might be a significant factor here. Quaglio (2009: 36) created a sub-corpus of the Longman Grammar Corpus to compare television series *Friends* with naturally occurring language. This subcorpus represents roughly 14% of the complete data set, but is more appropriately aligned with the reference corpus (*Friends* data) that he was using. In his analysis, he found that conversational American English prefers *maybe* (1044 per million) over *probably* (880 per million), contrary to the data in Table 10 above. Taking Quaglio's subcorpus of American conversational English for reference, the Television Dialogue Corpus of the

present study could be said to reflect naturally occurring language patterns accurately more so than when compared to the complete Longman Grammar Corpus (see above). This is a striking example of how sub-setting larger sized corpora can at times lead to flawed comparisons. Incidentally, this also underlines the argued heightened value of using transcripts of complete television series for comparative studies. Studies focusing on incomplete representations of speech ultimately carry with them the possibility that they are not ever entirely representative of a speaker's use of language.<sup>44</sup> Television language transcripts offer the unique opportunity to investigate the complete speech (or in any case, a considerable portion) that a speaker produces over the course of their fictional existence. Thus, while the above comparison to naturally occurring language may provide some insight into how fictional language relates to other registers, any results have to be interpreted with caution. Conversely, comparing the Television Dialogue Corpus with Quaglio's findings of *Friends* (2009: 73) offers the opportunity to compare two sets of corpora that are readily comparable (as they are both more or less complete sets of the language they claim to represent).

The below table presents Quaglio's findings (including comparison frequencies of the subset of the Longman Grammar Corpus of conversational American English) for *probably*, *perhaps* and *maybe* as well as frequencies found in the Television Dialogue Corpus' six series. Frequencies are provided as per 10,000 words.

---

<sup>44</sup> This is only ever an issue if the claim is made that findings are 100% representative. Most sociolinguistic studies rightfully delimit the scope of their findings, making clear that any result should only ever be seen within the context of the data obtained and broad generalizations need further investigation.

Table 11: Quaglio's *Friends* data and Television Dialogue Corpus, modal adverbs per 10,000 words

		<i>probably</i>	<i>perhaps</i>	<i>maybe</i>
<b>Quaglio (2009:157)</b>	<i>Friends</i>	3.86	0.35	13.61
	Conversation	8.8	0.26	10.44
<b>Television Dialogue Corpus</b>	<i>Angel</i>	5.6	0.11	17.01
	<i>Buffy the Vampire Slayer</i>	4.81	1.85	16.9
	<i>Gilmore Girls</i>	3.91	1.16	9.63
	<i>Parenthood</i>	5.19	0.2	11.55
	<i>Sherlock</i>	6.62	1.19	8.78
	<i>Torchwood</i>	1.72	0.62	10.73

The comparison shows that the overall pattern for scripted language of *maybe* > *probably* > *perhaps*, is consistent throughout all seven series. In what ways conversational American English (unscripted) can be compared is, as discussed above, debatable.

The following table displays the use of modal adverbs by speaker across the six series included in the Television Dialogue Corpus, as well as by speaker factor group (age, gender, nationality). Frequencies are, similar to previous sections, per 10,000 words spoken.

The right-most column summarizes all three modal adverbs used together. Since it is not clear whether individual stance markers carry any social meaning, an overall account of modal adverb use might add insight into how these features are used for characterization purposes.

Table 12: Modal adverb use across speaker groups, per 10,000 words

<b>character</b>	<b>series</b>	<i>probably</i>	<i>perhaps</i>	<i>maybe</i>	<b>modal adverbs total</b>
Angel		7.01	0.33	20.4	27.7
Cordelia	<i>Angel</i>	5.4	0	20.4	25.8
Spike		4.39	0	10.2	14.6
Angel		6.35	0	20	26.3
Anya		5.28	0.44	19.8	25.5
Buffy		6.95	0.16	21.6	28.7
Cordelia	<i>Buffy the Vampire Slayer</i>	4.5	0	7.5	12
Dawn		1.23	0	23.4	24.7
Giles		2.47	14.64	3.5	20.6
Spike		3.41	0.93	10.9	15.2
Willow		6.45	0.17	27.9	34.6
Xander		6.67	0.34	17.4	24.4
Emily		2.53	2	6.1	10.7
Lorelai		4.16	0.19	11.6	15.9
Luke		6.71	0	9.8	16.5
Paris	<i>Gilmore Girls</i>	2.5	0	8.2	10.7
Richard		2.8	5.32	4.8	12.9
Rory		5.78	0.38	15	21.1
Sookie		2.87	0.24	12	15.1
Adam		4.3	0	9.3	13.6
Amber		6.64	0.95	12.8	20.4
Camille		2.24	0	10.1	12.3
Crosby		5.99	0	20.2	26.2
Haddie	<i>Parenthood</i>	5.61	0	9.6	15.2
Julia		5.08	0.56	17.5	23.1
Kristina		5.78	0	11.6	17.4
Max		11.6	0	3.6	15.2
Sarah		3.14	0.45	11.7	15.3
Zeek		1.53	0	9.2	10.7
John	<i>Sherlock</i>	6.12	0	11.6	17.8
Sherlock		7.11	2.37	5.9	15.4
Gwen		0.49	2.43	10.2	13.1
Ianto		1.43	0	10	11.4
Jack	<i>Torchwood</i>	1.27	0	12.7	13.9
Owen		1.96	1.31	15	18.3
Rhys		1.99	0	8	9.9
Toshiko		3.19	0	8.5	11.7
mean overall		4.4	0.9	12.6	17.9
	<i>Angel</i>	5.6	0.11	17.01	22.73
	<i>Buffy</i>	4.81	1.85	16.9	23.56
mean by series	<i>Gilmore Girls</i>	3.91	1.16	9.63	14.7
	<i>Parenthood</i>	5.19	0.2	11.55	16.93
	<i>Sherlock</i>	6.62	1.19	8.78	16.59
	<i>Torchwood</i>	1.72	0.62	10.73	13.07
mean by gender	female	4.2	0.42	13.98	18.6
	male	4.62	1.4	11.24	17.26
mean by age	young	5.72	0.2	15.62	21.54
	middle	4.06	1.12	11.92	17.1
	older	2.28	1.83	7.54	11.64
mean by nationality	American	4.83	0.43	13.85	19.11
	British	3.26	2.17	9.39	14.81

With frequencies, in particular for *perhaps*, quite low across all characters, with some characters not having any tokens, any interpretation in the following analysis is intentionally (and topically) kept vague. It is however worth pointing out here, and in more detail further down, that generally low frequencies of *perhaps* are disrupted by Giles (*Buffy*) and Richard (*Gilmore Girls*), who both show exceptionally high frequencies for this particular variant.

Before analysing whether distributions of modal adverbs correspond to characterization patterns, the following analysis focuses on how these markers are strategically used across character factor groups. As previously mentioned, there is no concrete sociolinguistic account of feature preference across gender or age, and only Biber et al.'s claim that American English uses modal adverbs in higher relative frequencies. Studies on epistemic stance markers more generally highlight a preference in use by female speakers, although these assumptions have not been applied to modal adverbs as such. The analysis thus aims at finding out whether any of these patterns are indeed observable in the Television Dialogue Corpus and whether there may be patterns that have not been discussed before.

#### **5.4.4 *Distributions across categories***

##### *Age and gender*

Distributions across gender show no significant ( $p > 0.5$ ) preference for the use of any of the three modal adverbs. This refutes possible claims that modal adverbs are used to cater toward stereotypes of women as more uncertain and unassertive, as laid out by Lakoff (1975). It is possible however that the constellation of the television series included in this corpus affects these results. With strong female characters and feminist storylines found throughout the plots, it might be expected that characters are not falling into

categorical gender-based patterns of certain/uncertain or strong/weak schemata. Three possible hypotheses can be taken from here:

1. Modal adverbs are characterization devices that indicate a lessened degree of speaker commitment, but as a characteristic trait this is not gender-based.
2. Modal adverbs are characterization devices that indicate a lessened degree of speaker commitment and follow stereotypes that reflect a gender binary. However, the present corpus does not include storylines that permit any categorical gender binary.
3. Modal adverbs are not systematically used as characterization devices, thus show no significant variation across character factor groups.

Hypotheses 1 and 2 are not necessarily conflicting interpretations and a following character-specific analysis will shed light on patterns that (may) exist in the present corpus. In any case, for the series included in this study, the use of modal adverbs is not based on categorical gender stereotyping and is thus different to other features that have been previously investigated (pragmatic markers, section 5.1 in particular).

The below plot shows the mean distributions, adjusted by mean frequency (per 10,000 words). The three variants are used at different frequencies, but consistently by both genders, meaning women and men in general follow the pattern of *maybe* > *probably* > *perhaps*. The frequencies, in particular for *perhaps*, are rather low and thus any interpretation needs to be made with caution. However, while most characters follow a similar pattern, *perhaps* shows a number of outliers that use the feature against what everyone else is doing. I will return to this in the section on character-specific use of modal adverbs to see whether this corresponds to a particular character trait.

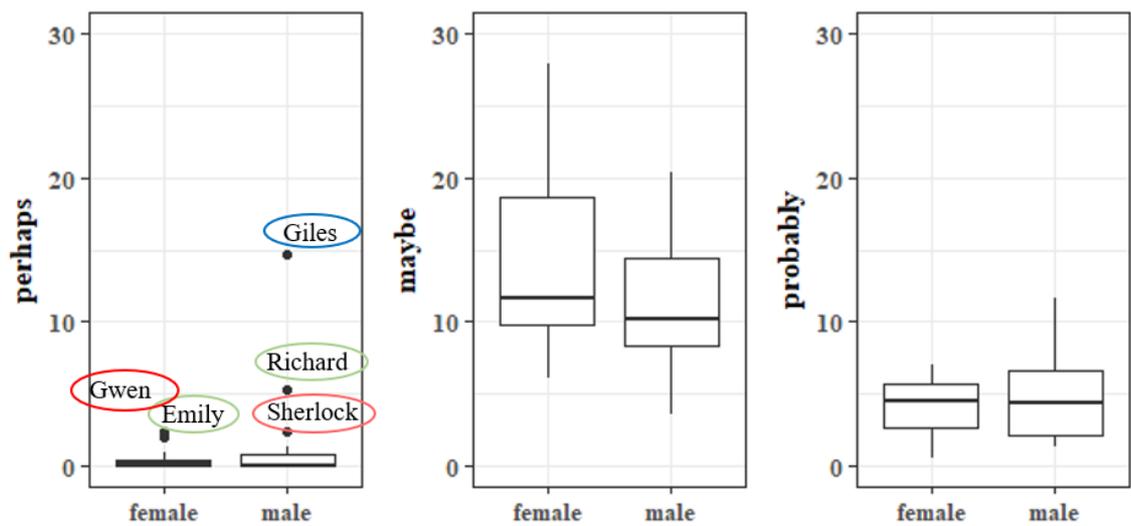


Figure 27: Modal adverb use across gender, per 10,000 words

In terms of age, stratification from younger to older characters can be seen as decreasing for *maybe* and *probably*, and slightly increasing for *perhaps*. Only the change of *maybe* is statistically significant ( $p < 0.05$ ), between the middle and older character groups.

While gender distinctions did not point to categorical character factors that might indicate modal adverbs as characterization devices, age distribution seems to be marked. Of note here is particularly the finding that patterns are not congruent for all three features and that instead the individual adverbs seem to be indicative of a certain age group.

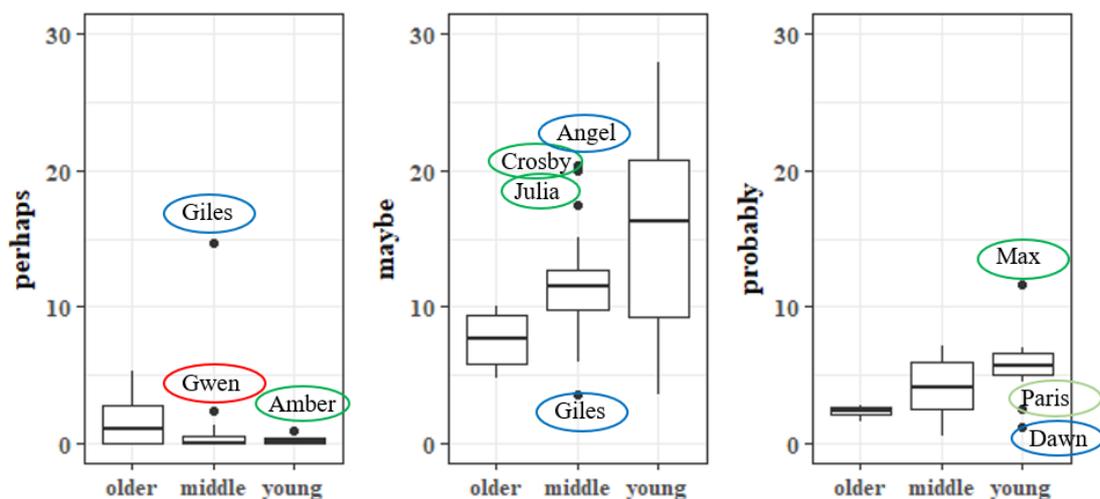


Figure 28: Modal adverb use across age, per 10,000 words

Modal adverbs are relatively stable features (rather than incoming features) so it can be assumed that the apparent time display (see plots above) describes a situation related to age grading. This is comparable to findings from pragmatic marker *I mean* (section 5.1) or hedge *kind of* (section 5.2), although, arguably, both of those features also exhibit signs of pragmaticalization which is an indicator for language change. Here we find much more stable variation and can thus infer generational preferences with more confidence. Younger characters are using non-committal markers at higher frequencies, possibly indicating personalities that are not yet self-assured and confident.

The below table summarizes the findings from this section and cross-tabulates the means for the character factor groups age and gender. As discussed above, overall patterns of preference are consistent for male and female characters. Further, the cross-tabulation highlights that for younger characters gender distinctions are more pronounced than in other age groups. Young female characters thus show a marked preference for *maybe*,

while young male characters use *maybe* less and, importantly, at almost similar frequencies to *probably*.

Table 13: Cross tabulation of modal adverbs: Age & Gender, means per 10,000 words

	young	middle	older	
<b><i>Probably</i></b>				
Female	5.03	3.53	2.38	
Male	9.13	4.32	2.17	
<b><i>Maybe</i></b>				
Female	16.64	11.85	8.10	
Male	10.50	11.96	6.97	
<b><i>Perhaps</i></b>				
Female	0.21	0.55	1.00	
Male	0.17	1.40	2.66	

A character-based analysis of marker preference will investigate if these findings are due to overall patterns of language use or whether these distributions are determined by outliers. If the latter is the case, further indexical meanings of modal adverbs might affect the characterization process. The main finding from this section is a clear age-based distribution for *maybe* and *probably*, indicating that with increasing age the use of modal adverbs decreases. For *perhaps* the opposite pattern can be discerned, albeit low overall frequencies make this finding not as clearly marked.

### *Nationality*

Biber et al. (1999) claim that American English uses modal adverbs more frequently than British English and, looking at the means for the Television Dialogue Corpus, this is seemingly the case here as well:

Table 14: Modal adverb by nationality, means per 10,000 words

		<b>probably</b>	<b>perhaps</b>	<b>maybe</b>	<b>Total</b>
<b>mean by nationality</b>	American	4.83	0.43	13.85	19.11
	British	3.26	2.17	9.39	14.81

No statistical significance (at .05) can be found, although for *maybe* p is at .056. Despite no statistically observed trend, overall patterns that correspond to previous research can be assumed.

Both varieties show a clear preference for *maybe*, followed by *probably* and *perhaps*, making *maybe* the prototypical modal adverb across all character factor groups. Interestingly, this is distinct from Biber et al.'s findings (1999: 869) that claim that *probably* is more frequently used.

The boxplots below summarize the normalized findings (per 10,000 words). Again, a number of outliers appear for *perhaps*, as well as for *maybe*. The character-based analysis following this factor group section will investigate how far this indicates systematic characterization or whether the low overall frequencies for *perhaps* in particular might skew the picture.

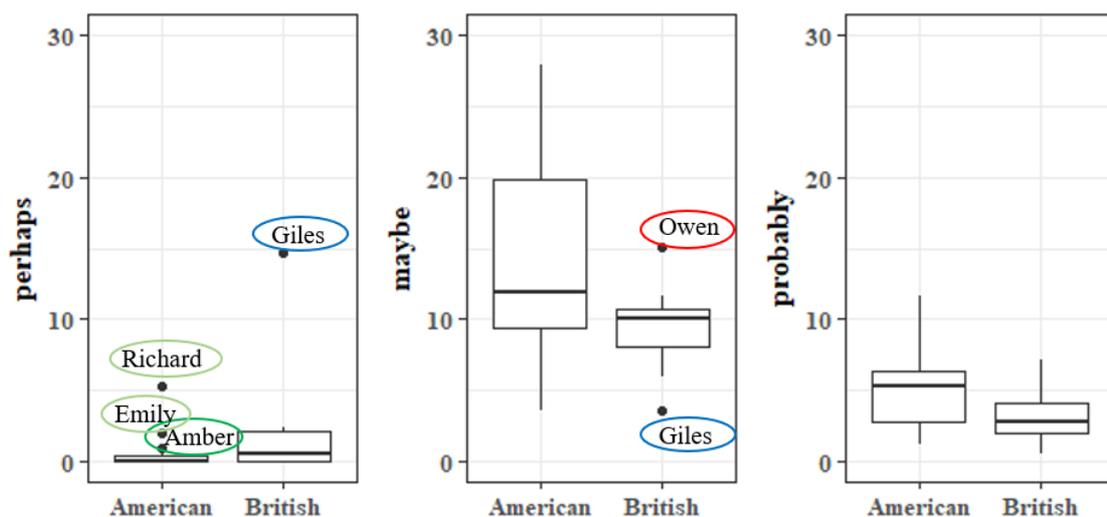


Figure 29: Modal adverb use across nationality, per 10,000 words

Previous sections have shown that British English patterns oftentimes correspond with patterns of older American characters (particularly characters of upper social class).

The corpus does not consist of British characters covering the three age groups set up through the American characters and instead represent middle aged characters exclusively. If patterns of ‘older American characters = Britishness’ hold true for modal adverbs as well, however, frequency correspondence should still be observable in cross tabulation of nationality and age.

The plot below shows mean frequencies for the three age groups for American characters for each of the modal adverbs and how British frequency means relate. For all three variants the British mean frequency (per 10,000 words) is closest to the older American speakers. This can be seen for the variants where older speakers are exhibiting a decreased use of the feature (*maybe* and *probably*), as well as for *perhaps*, where the use increases with age.

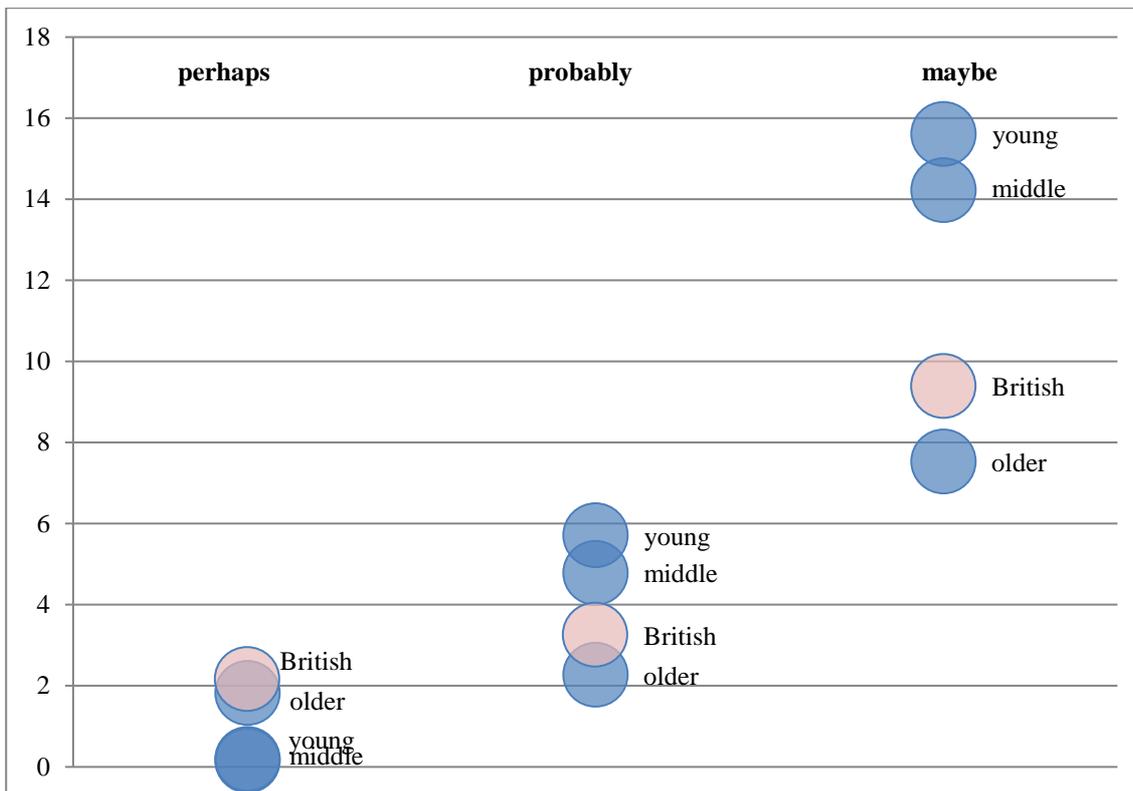


Figure 30: Modal adverb use cross-tabulation: nationality & age (per 10,000 words)

The following character-based analysis will investigate this pattern further, including a more detailed look at how the outliers within the nationality distinction relate to overall patterns.

So far, modal adverbs seem to predominantly reflect character developments in terms of age with possible indexical meanings of Britishness patterns that have been found in previous features as well.

#### 5.4.5 *Character-based distributions*

This section of analysis will further investigate some of the patterns found in the overall character factor groups and attempt to uncover character-specific uses of modal adverbs that are not visible through broad categories of gender, age, or nationality.

The plots used in this section, again, give frequencies as per 10,000 words spoken by each character. No means will be used to illustrate points. Important to note here is that for some features the frequencies are relatively low (in particular for *perhaps*), and equally, some characters use modal adverbs at low frequencies overall. While patterns can still be found, any interpretation based on low frequencies is essentially to be taken with caution.

The first part of this section takes each variant individually and examines usage patterns by character for each series. This makes it easier to compare the use within the same context (i.e. series). Characters for each panel are ordered in terms of frequency of use from top to bottom, from least frequent to most frequent respectively.

The first plot shows the use of *maybe* across the series and characters. What can be seen here is that there is not one characterization pattern (or index) observable, but several.

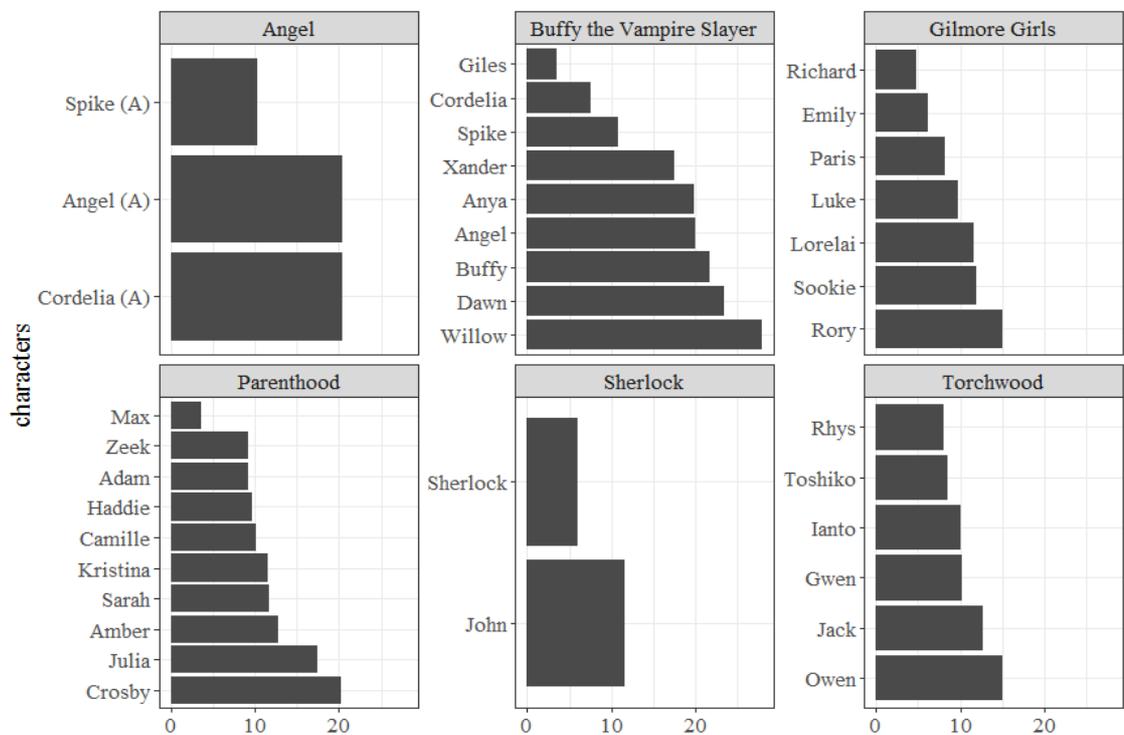


Figure 31: Modal adverb *maybe* across series and characters, per 10,000 words

## 1. Commitment

As previously mentioned, modal adverbs are possible linguistic devices that indicate insecurity or non-committal of a character. Thus, a low use of modal adverbs (here, *maybe*) might indicate that the character is speaking with higher levels of assurance, confidence and/or security. Characters that have previously found to use a lesser degree of vagueness indicators are, for instance, Spike (both in *Angel* and *Buffy the Vampire Slayer*), Cordelia (particularly in *Buffy the Vampire Slayer*) or Paris (*Gilmore Girls*). The plot below confirms these tendencies as these characters are found on the low-use side of the respective scales.

Comparably, the other end of the scale is not as clearly indexed. While it can be argued that the high use of modal adverbs by Willow (*Buffy the Vampire Slayer*) or Rory (*Gilmore Girls*) exhibits higher levels of insecurity, other characters do not fit that characterization pattern. Crosby (*Parenthood*) or Owen (*Torchwood*), for instance, are confident characters within their respective series. This suggests that *maybe* is not reflecting clear indexical meanings.

## 2. Pragmatic competence

As highlighted by Fraser (2010), using modal adverbs also indicates a level of pragmatic competence. Using hedging devices helps manage communication through saving face and providing levels of politeness. A lower use in these features might point towards characters that are less equipped in handling social interactions. Previous sections have found Max (*Parenthood*) in particular to be a character who is oftentimes portrayed as lacking social skills and pragmatic competence. As a stereotyped presentation of ASD, his dialogue exhibits a number of distinct language patterns that are different from the other characters on the series. His use of modal adverb *maybe* here confirms that.

Markedly fewer instances of the variant point towards a conscious scripting of hedging devices for characterization purposes. A similar pattern can be seen for Sherlock (*Sherlock*), who not only shows a lack of social skills but also portrays an incredibly self-assured persona.

### 3. Britishness reflects upper class American English

As indicated in the section on nationality variation in general, patterns of correspondence between Britishness and upper class American English that have been previously found are again reflected here. Richard and Emily (*Gilmore Girls*, also, to a degree, Paris), as well as Giles (*Buffy the Vampire Slayer*) show similar uses of modal adverbs. This is equally the case for *perhaps*, see plot below.

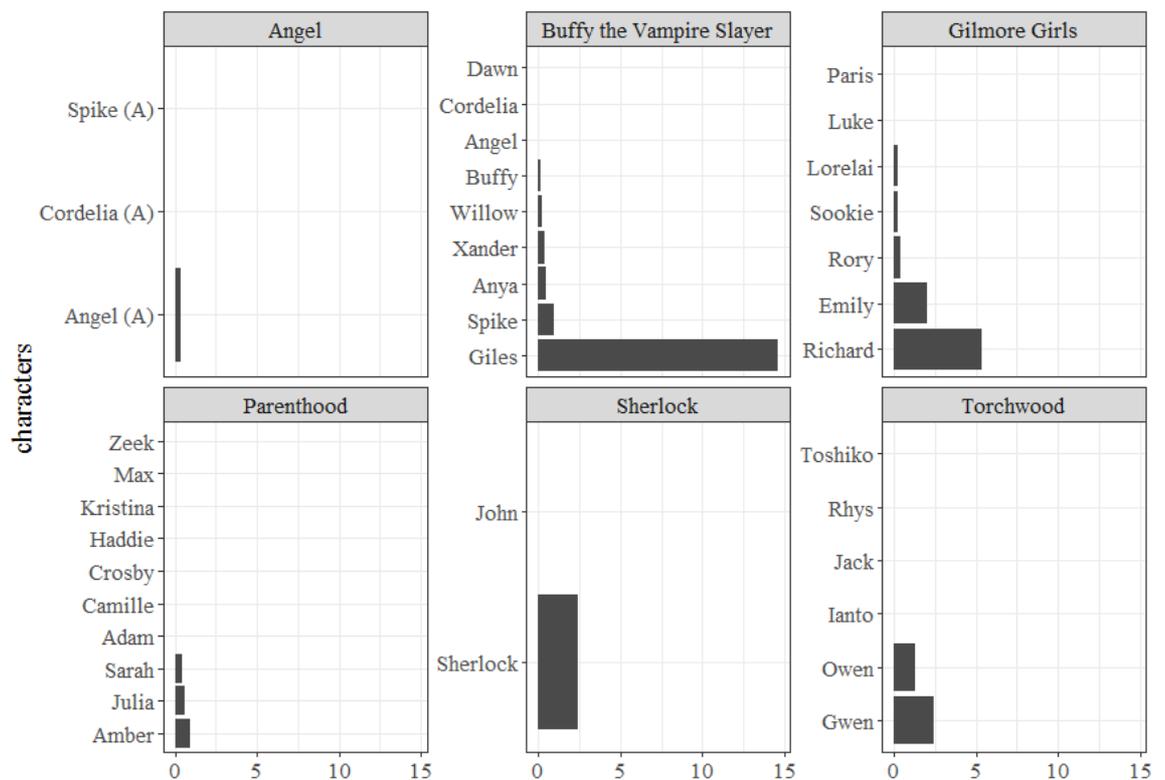


Figure 32: Modal adverb *perhaps* across series and characters, per 10,000 words

While this plot shows the strikingly low use of the variant overall, it also draws attention to the pattern outlined above. Giles, previously low in his use of modal adverbs, shows a high use of *perhaps*. His use, while not to the same degree, is mirrored by Richard and Emily, Sherlock, as well as Owen and Gwen (*Torchwood*). The low use overall and incongruent use by British characters does not fully establish *perhaps* as a marked feature of Britishness. However, seeing that its use is predominantly by British characters as well as upper class American characters, previous characterization patterns seem reflected here.

Lastly, as the distribution of *probably* (plot below) further shows that the variants are not used in similar ways and might carry differing indexical meanings.

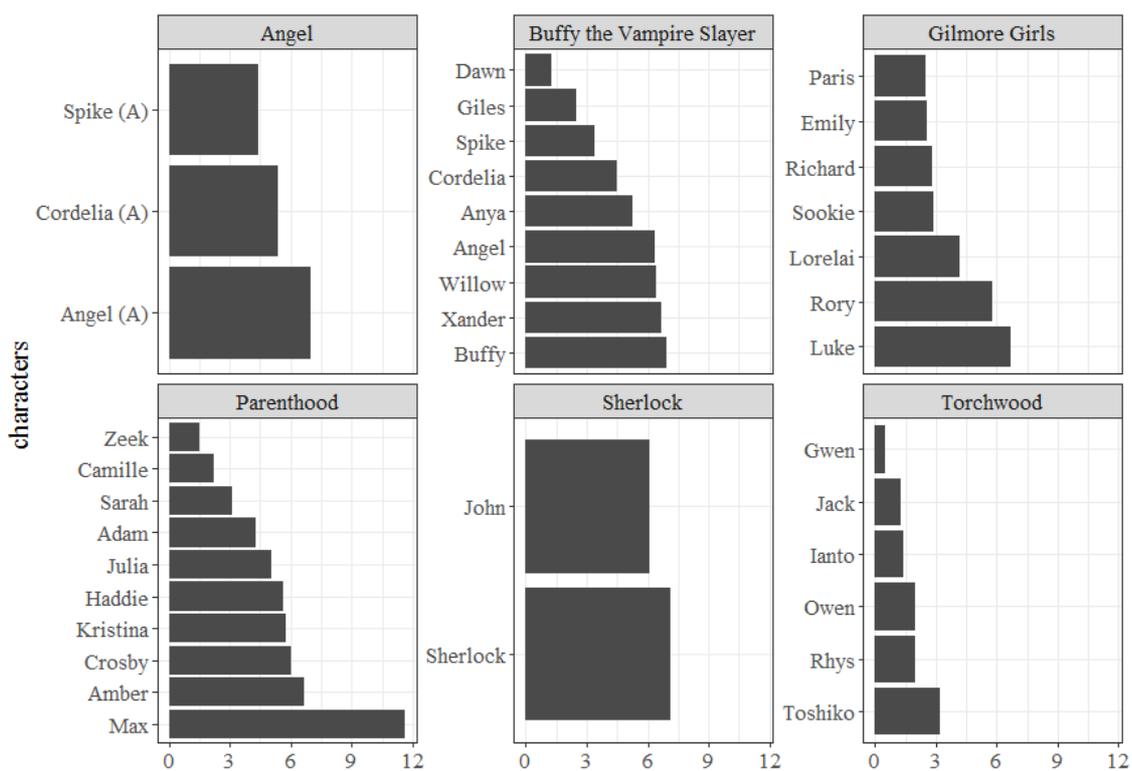


Figure 33: Modal adverb *probably* across series & characters, per 10,000 words

While a low use of *maybe* seemed to serve as an indicator for low pragmatic competence, the picture seems reversed for *probably*. Indeed, Max, as well as Sherlock, use *probably* as their predominant modal adverb and ahead of other characters in their respective series. It is possible that different semantic associations for modal adverbs might have affected the use here. This suggests that modal adverbs cannot be easily grouped together as equivalent markers of any one function. As the above analysis has shown, each variant can serve multifunctional purposes within scripted dialogue. While the basic function of non-committal marker and/or hedge can be applied throughout, it is important to investigate each variant on its own.

This is exemplified within the plot below (34). Here, all three modal adverbs are put together for each of the characters from the Television Dialogue Corpus. Referring back to some of the main functions that were discussed, a possible interpretation might be that Rhys (*Torchwood*) or Zeek (*Parenthood*) are rather self-assured and committed characters due to their low use of modal adverbs throughout. Additionally, Willow and Buffy (*Buffy the Vampire Slayer*) are the most non-committed (or insecure) characters. While the argument could certainly be made for Willow, the other characters do not fit that characterization. Characters that were found as outliers in previous sections (Sherlock, Giles, Max) are actually showing average accounts of modal adverb use which ignores some of the most indicative findings.

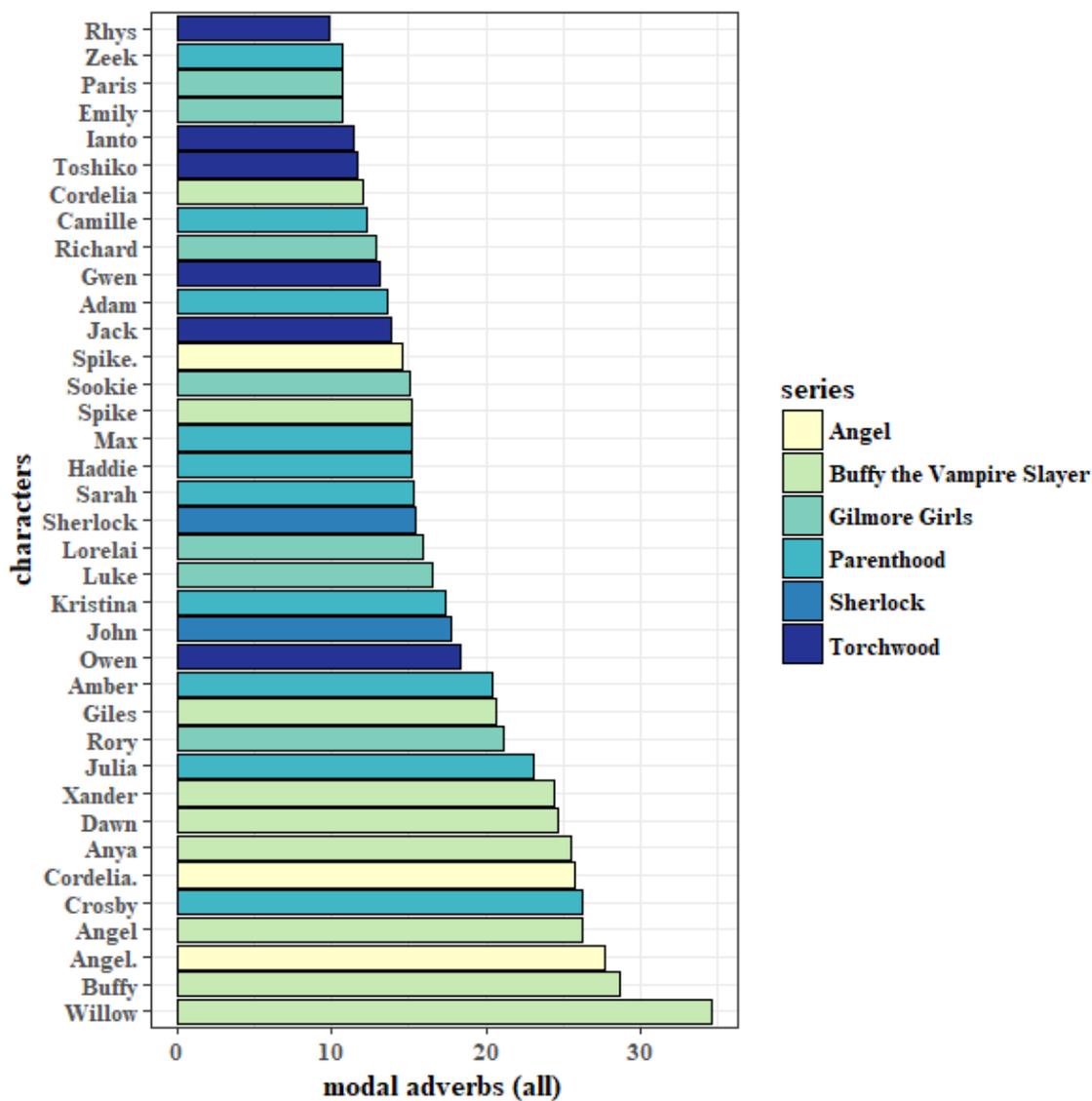


Figure 34: Modal adverbs across characters, per 10,000 words

In comparison to other features included in the study, modal adverbs are not clearly marked as indexical devices for characterization. However, certain patterns are discernible and characters are distinguishable through their use of modal adverbs.

#### 5.4.6 *General extenders as characterization devices*

- Does the distribution of markers reflect characterization patterns within and across series?
  - Partly. Because modal adverbs are multifunctional (where functions are also oftentimes overlapping), clear characterization structures are difficult to pinpoint. Certain character differences (such as age, nationality, pragmatic competence) are more likely to be indicated through modal adverb use. It is important to note that the individual variants are indexed in different ways. Even though the basic function for all variants is supposedly similar, the data shows that preferences are indeed varied.
- Does the distribution indicate indexical and/or stereotypical characterization and if, what characters or character groups are represented like this?
  - Overall distinctions for character groups (for all three modal adverbs that are included) can be found for age, where younger characters use modal adverbs *probably* and *maybe* more frequently than older characters. The distribution for *perhaps* is reversed. Another distinctive use can be detected with regards to nationality, whereby previous claims of higher frequencies of overall modal adverb use in American English (above British English) is reflected in the Television Dialogue Corpus as well. In terms of indexical meanings of uncertainty and a lack of commitment to what is being stated, findings have been mixed. For some series, a low use of one or more modal adverbs indeed reflected heightened

commitment and vice versa. Findings however are not congruent across the corpus or across features.

- Social networks are implied to a lesser degree here than was the case with previously analysed features. Notable exceptions here are, again, Richard and Emily, who use these features in a similar fashion. Further, for *maybe*, it is striking to see a gender-divide in *Parenthood* with only Crosby falling out of pattern and using the form most of all. As will be pointed out elsewhere further on, Crosby aligns closer with other female characters than with the male characters as he defies typical masculine traits. In fact, for any feature that is stereotypically associated with female speakers, he is a top-user.
- Are distributional outliers and unexpected patterns indicative of specific characterization choices?
    - Outliers found in the character factor groups (in particular nationality and age) are characters that have been previously found to be characterized in specific manners. An unusually high use of *probably* by Max for instance is unexpected as it is not only distinct within the context of the other characters on the series, but also because a high use of this variant goes against indexical meanings of social aptitude. Other outliers reflect previous findings of older American characters aligning with British patterns. Again, this is particularly marked for upper social class (where this is important for the overall plot of the series).

The final subsection of individual feature analyses investigates the use of intensifiers in the Television Dialogue Corpus. This feature is distinct from previous sections in that within the notion of scales we now move from features lessening commitment towards a proposition to features that heighten the effect.

## 5.5 Intensifiers

### 5.5.1 Feature definition

Intensifiers (also called degree words (Bolinger, 1972), adverbs of degree (Biber et al., 1999), or degree modifiers (Paradis, 1997; 2000)) are adverbs that modify the quality of the following word or phrase. Extract (63) below gives an example of how intensifier *very* modifies adjective *romantic*:

(63) Luke: It was very romantic. (GG)

Modification can be understood as a scale where the semantic meaning of the modified word (or head) can intensify upwards (increase) or downwards (decrease). General terms for these distinctions as found in previous studies are ‘amplifiers’ and ‘downtoners’ after Biber et al. (1990: 554-555) or ‘reinforcers’ and ‘attenuators’ after Paradis (1997:17-18). The notion of the semantic scale goes from positive totality to mere approximation towards the head’s quality. Generally, intensifiers can be broadly categorized into four groups:

#### i. Maximizers

Examples: *absolutely, totally, completely*

Scale the following word up to the extreme end of the scale and will usually occur with words that are bound (i.e., are part of what Paradis calls an “either-or conception” (2000:148)): *right –wrong, full – empty, etc.*

#### ii. Boosters

Examples: *so, very, really*

Scale the following word up to a degree and “reinforce the gradeable property of the adjective [or other item] they apply to” (Paradis, 2000: 148).

iii. Moderators<sup>45</sup>

Examples: *fairly, pretty, quite*

Moderators qualify the following word and diminish the propositional meaning, similar to previously discussed hedges.

iv. Approximators<sup>46</sup>

Examples: *almost, hardly*

Here, the intensifier indicates that the propositional meaning of the modified word was not reached, or “falls short of the expected limit” (Paradis, 2000: 148).

The below figure (Figure 35) illustrates the distinctions with an example utterance from *Gilmore Girls*’ Emily: “Well that sounds *very* interesting”, where *very* is the intensifier that is boosting the meaning of the adjective *interesting*. The sentence is adapted according to other variants and their functions based on their meaning along the semantic scale.

---

<sup>45</sup> Many moderators can also act as maximizers, although the unambiguous transcription of that function is somewhat difficult. Some studies therefore exclude these items altogether from an analysis that focuses on any upscaling function. However, there is no consensus over which variants should be excluded and which are part of the variable set (Tagliamonte and contributors (2003, 2005, 2008) for instance include *pretty*, while Barnfield and Buchstaller (2009) exclude *pretty*). Paradis (2000) claims that disambiguation can be achieved by the type of modified word (whether it ranges on totality or scaling), but through processes of grammaticalization and overall expansion of context, this distinction does not reflect actual language use throughout.

<sup>46</sup> There are further distinctions for downtoners. Quirk et al. (1980) for instance differentiate between approximators, compromisers, diminishers, and minimizers whereas Paradis subcategorises between approximators, moderators, and diminishers. Following the notion of scalarity, functional distinctions are not clear cut and more or less dependent on the researchers’ respective framework.

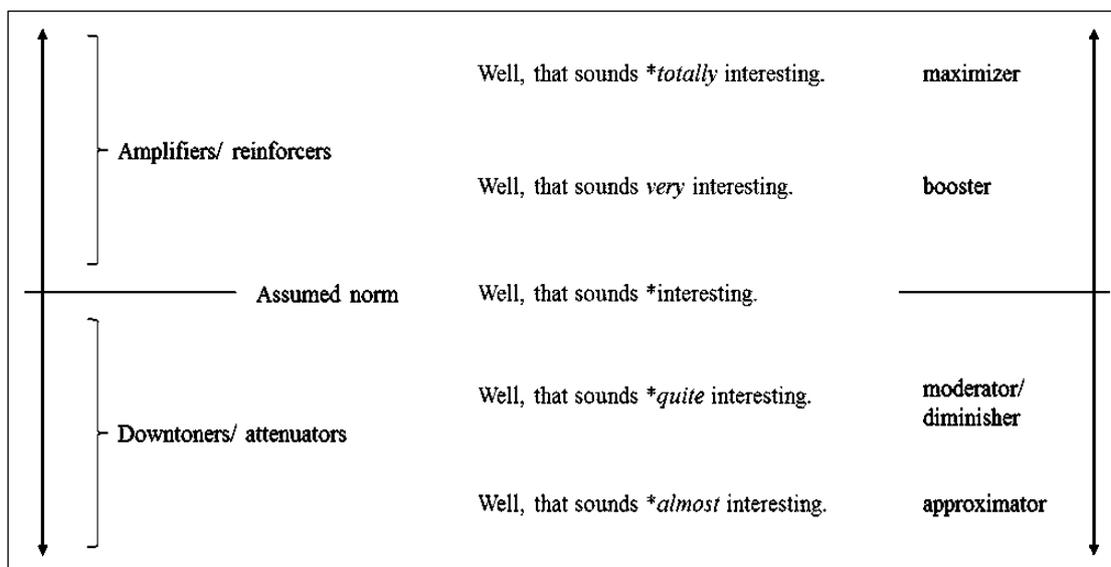


Figure 35: Intensification terminology

The variable set included in this section focuses on the upward scaling intensifiers (boosters and maximizers) exclusively. Moderators and approximators share their pragmatic functions with some of the items I investigated in previous sections (in particular hedges *sort of* and *kind of*, approximator *like*, and general extenders) and the analysis will now turn its attention towards the other end of the scale.

In line with features previously investigated in this study, intensifiers denote commitment to a speaker’s proposition (stance) and might provide further insights into the characters’ identifying qualities in terms of emotionality. In fact, Labov (1984:43) claims that while emotional states are “often expressed through peripheral, gradient systems: by prosody, vocal qualifier, and gesture”, intensifiers are the most common linguistic features that are “specifically devoted to emotional expression”.

There is a vast body of sociolinguistic research on adjective intensification that finds the use of intensifiers varied across gender, age, nationalities, and communities of practice.

In the following, I will summarize previous studies that have investigated adjective intensification, focusing on boosters and maximizers.

Tagliamonte and her contributors examined adjective intensification most thoroughly in terms of comparability and range by conducting research in three different settings from 1998 to the early 2000s: British English (York, England, 2003), North American English (Toronto, Canada, 2008), and scripted English (television series *Friends*, produced in the US, 2005). The studies give a detailed account on the historical trajectory of intensifier development in English from as far back as Old English (Ito and Tagliamonte, 2003:259-262; Tagliamonte, 2008:362-365), highlighting the fast-paced change of preferences in use. Table 15 below shows the most frequent intensifiers that were found in the three studies.

Table 15: Most frequent intensifiers in Tagliamonte and collaborators' studies

<b>Ito and Tagliamonte (2003) York, UK</b>					
	<i>very</i>	<i>really</i>	<i>so</i>	<i>absolutely</i>	<i>pretty</i>
<b>N</b>	364	287	96	30	30
<b>%</b>	38.3	30.2	10.1	3.2	3.2
<b>Tagliamonte and Roberts (2005) tv series <i>Friends</i>, US</b>					
	<i>so</i>	<i>really</i>	<i>very</i>	<i>pretty</i>	<i>totally</i>
<b>N</b>	832	464	269	115	53
<b>%</b>	44.1	24.6	14.2	6.1	2.8
<b>Tagliamonte (2008) Toronto, Canada</b>					
	<i>really</i>	<i>very</i>	<i>so</i>		
<b>N</b>	1282	651	599		
<b>%</b>	35.9	18.2	16.7		

Overall, the most frequently used intensifiers in North American and UK unscripted contexts are *very* and *really*, with *really* still being preferred in the UK, while the North American corpus shows that *really* has overtaken that position. In addition to that, the data also indicates the increased use of *so* in these contexts. The data from *Friends* actually shows *so* as the most frequently used intensifier, which leads Tagliamonte and

Roberts (2005: 288) to speculate whether this particular scripted series “may reflect the next phase in the changing intensifiers of English”.

All three studies corroborate a shift, albeit at different stages, from *very* to *really* to *so*. With regards to the *Television Dialogue Corpus* and the present study, this shift might be visible in terms of variation across age groups (with older speakers more likely to use outgoing variants). Furthermore, with the York study (Ito & Tagliamonte, 2003) the authors find *very* to be the most frequent intensifier still, nationality of the characters from the *Television Dialogue Corpus* might also indicate structured variation. In fact, it is not just data from York that points towards a slower moving trajectory of intensifier shift in the UK. Barnfield and Buchstaller (2010) find that *really* is replacing *very* in a study covering five decades of Tyneside English. Below, Table 16 summarizes their results (five most frequent intensifiers) from the oldest to the latest corpus.

Table 16: Most frequent intensifiers in Barnfield and Buchstaller (2010)

<b>Tyneside TLS<sup>47</sup></b>					
	<i>very</i>	<i>really</i>	<i>rather</i>	<i>absolutely</i>	<i>so</i>
<b>N</b>	91	12	9	4	4
<b>%</b>	65	8.6	6.4	2.9	2.9
<b>Tyneside PVC<sup>48</sup></b>					
	<i>dead</i>	<i>really</i>	<i>very</i>	<i>so</i>	<i>absolutely</i>
<b>N</b>	106	74	53	22	12
<b>%</b>	35.9	25.1	18	7.5	4.1
<b>Tyneside NECTE<sup>49</sup></b>					
	<i>very</i>	<i>really</i>	<i>so</i>	<i>dead</i>	<i>absolutely</i>
<b>N</b>	96	79	27	23	14
<b>%</b>	32.4	26.7	9.1	7.8	4.7

<sup>47</sup> Tyneside Linguistic Survey

<sup>48</sup> Phonological Variation and Change in Contemporary Spoken English

<sup>49</sup> Newcastle Electronic Corpus of Tyneside English

In the 1960s (TLS corpus), *very* is the preferred intensifier by a vast margin. The PVC corpus (collected in the early 1990s), with most of the speakers under the age of 20<sup>50</sup>, show a rapid shift where *very* falls into third place of most used intensifier. The NECTE corpus from 2007-2009 provides a further indication of the ongoing shift. While still in first place here, *very* is leading by just over 5% ahead of *really*. *So* can be seen slowly entering the intensifier system, going from under 3% to almost 10% over the course of 50 years. In addition to the longitudinal replacement involving *so*, *really*, and *very*, the PVC corpus also provides evidence for more rapid age-dependent shifts. The use of *dead* describes a short-lived trend in intensification, underpinning one of the pragmatic contexts of the functional group of intensifiers. In order to add emphasis and highlight emotional stance, some intensifiers come into use very fast, especially in young speaker groups. A similar patterns can be seen in Macaulay (2006), who found *pure* to be a popular intensifier in Glaswegian working-class adolescent communities. He claims that *pure* is used as a “sign of group identification” (2006: 276). There is no evidence whether *dead* or *pure* are used outside of their respective communities and take part in the overall shift, although most studies suggest that the main intensifiers are *really*, *very*, and *so*, if at differing preferences.

In terms of totality intensification, previous studies show *absolutely* as the most frequently used maximizer in the UK and *totally* dominating in North America.

The discussed studies all indicate the meaningful pragmatic functions of intensifiers. Furthermore, clear social distinctions with female speakers often leading in innovative variants and age reflecting the overall replacement of *very* with *really* with *so* can be seen.

---

<sup>50</sup> 19 of 35 participants where under the age of 21

Tagliamonte and Roberts (2005) also find that scripted contexts somewhat mirror what happens in naturally occurring language. However, it is not clear whether intensifiers carry indexical meaning that may correspond with characterization processes for telecinematic contexts.

A recent study by González-Díaz (2014) investigates the use of intensifier *quite* in Jane Austen's novels. She suggests that Austen did not only write her characters as stereotypically female (with reference to the 18<sup>th</sup> century language use), but that she also created character-based differentiations in how *quite* patterned across novels. According to the study, *quite* adopted new pragmatic functions around the time that Austen was writing. The new function became "socio-stylistically marked" (2014: 321) and was, as González-Díaz claims, consciously used to portray certain characters "as 'deviant' and/or 'inferior' in some respect"<sup>51</sup>. In particular, she writes that the pragmatic functions of intensifiers have been "consistently exploited for the creation of social identity" (2014: 311). Generally, the use of *quite* as an intensifier of the time was mostly associated with women. This stereotype was used in the novels, for instance, when creating dialogue for male character Mr. Woodhouse from Austen's novel *Emma*, who was styled as 'the most unmanly of the males' (Walker 2009: 215). Other studies that investigate intensifier use in Austen's novels include, for instance, Barchas (2007) and Burrows (1987). Both comment on the fictolinguistic appropriation of existing stereotypes for characterization purposes and find intensifier use reflective of characters and character groupings. From these studies, we can see that intensifiers are used to index certain social categories and

---

<sup>51</sup> She writes, in the case of *Pride and Prejudice* for instance, Jane and Elizabeth are using the old unmarked form while Mrs Bennet and Lydia prefer the new and marked function. (2014: 321)

that intensifier variation is used for character stylisation. The following analysis will test whether this is equally the case with scripted television language.

In particular, based on what previous studies on adjective intensification have found, I will pay close attention to:

- i. Indication of language shift with a changing preference from *very* to *really* to *so*
  - a. Whether this is visible across age (with older forms corresponding to older characters)
  - b. Whether this is visible across gender (with new forms being used earlier/more frequently by female characters)
  - c. Whether the trajectory is different across nationality (with British characters using fewer new forms)
- ii. Indication of variant preference across nationality (American and British characters use different variants, e.g. *totally* vs. *absolutely*)
- iii. Indication of group membership (with close-knit character groups using similar variants or outsiders using markedly different forms)

Before presenting distributions of the intensifiers found in the Television Dialogue Corpus, I will briefly summarize the methods behind token inclusion and coding.

### **5.5.2 *Token inclusion & coding***

This analysis differs from previous analysis sections in so far as that it does not start out with an existing list of variants for intensifiers. Rather, any forms that occur in the Television Dialogue Corpus and follow the pragmatic function of heightening the meaning of a following adjective are determined possible items to extract. It thus follows traditional sociolinguistic methods in considering all variants in order to account for the complete context of variation, or, all “elements that vary along the same dimensions in response to the same state of affairs” (Labov, 1978:10). The advantage of this is the thorough insight into linguistic choices, i.e. the full picture of variation rather than a

possibly limited excerpt. By including all possible variants, an investigation can not only tell that a variant is changing, but how it changes with reference to all other variants. This is particularly helpful when analysing intensification. As previous studies have shown, the variants consistently undergo change with preferences not only for new forms, but also older “recycled” variants that have been used with this function before. An analysis that captures these movements across features as well as social categories will ultimately be able to tell more about possible characterization through linguistic preferences.<sup>52</sup>

The full variable set of intensification included in this study follows two prerequisites:

- i. Functional equivalence: Following Rickford et al. (2007:8) “all adverbial strategies speakers have at their disposition to boost or reinforce the property denoted by their heads” are included
- ii. Structural equivalence: the modified head is an adjective (within a verb or noun phrase) and is not a) negated, b) comparative, c) superlative, or d) forming a question

These points are based on other studies that analysed social preference in intensifier choice. In particular, the syntactic context under which I investigate these features is adapted from Tagliamonte and contributors (2003, 2005, 2008) who provide considerable points of comparison by including North American and British English, as well as fictional contexts. They however differed in one important aspect which reflects the definition of the variable context. By their definition, all possible variants for intensification include cases where intensification of adjectives could have occurred, but

---

<sup>52</sup> As mentioned in the individual analysis sections, for most pragmatic features this type of analysis is difficult if not impossible to do. With multifunctionality and ambiguous syntactic contextualization, the variable is difficult to circumscribe. Therefore, most socio-pragmatic studies rely on normalized frequency analyses instead.

did not (so-called zero cases). This poses the question whether these cases describe equal functional contexts. Taking the first prerequisite (see above) as the definitional context for the variable set, I argue that non-intensification does not serve the function of adding heightening quality to an adjective and thus zero cases are not included in the set of variants in the present study. This is in line with data inclusion methods found in Rickford et al. (2007) and Barnfield and Buchstaller (2010).

Cases where the variable context is abridged were also dismissed. These include interruptions, false starts, or ellipses. The fact that the complete syntactic construct is not given and can thus not be compared consistently is the reason for exclusion from the analysis.

In cases of multiple intensification, only the intensifier context of ‘token + adjective’ is counted.

For the initial extraction of possible token lines, the *Television Dialogue Corpus* was automatically tagged for parts of speech<sup>53</sup>. This process attaches tags onto each word of dialogue, enabling a search for syntactic function and position, rather than for lexical representation only. Example (64) below shows a tagged line spoken by Dawn from *Buffy*:

(64) Dawn: Big\_JJ square\_JJ building\_NN1 filled\_VVN with\_IW  
boredom\_NN1 and\_CC despair\_NN1 .\_. (BVS)

The syntactic context of intensification is relatively unambiguous and found to be accurately represented through the parts-of-speech tagging. In many cases this sort of

---

<sup>53</sup> Using <http://ucrel.lancs.ac.uk/claws/trial.html>, using tagset C7

automatic determination of feature contexts is not without problems, especially in terms of the automatic detection of pragmatic functions. A concrete variable set-up as well as clear coding guidelines however eases the use of automation.

In the concordance software used previously, rather than searching for lexical items, I queried for any word that was tagged as an adverb followed by any word that was tagged as an adjective. As most intensifiers were tagged as adverbs, this method resulted in the most relevant hits. To account for any mistaken tags, I also extracted token lines for ‘adverb + adverb’ and ‘adjective + adjective’.

After being extracted, the tokens were manually checked for their intensifying function. Some variants had multiple functions in discourse and demanded extra attention, while other tokens were unambiguously boosting the head adjective.

The following analysis, as sections before, will investigate the use by characters and character groups. Furthermore, I am including a multivariate analysis to establish which social character factor is dominant in terms of intensifier choice. This will offer some further insight into which character background quality (age, gender, and nationality) is most recognized as a stylistic index.

### **5.5.3 *General distributions***

Overall, I found 74 distinct variants adding up to a total of 5998 intensifiers in the *Television Dialogue Corpus*. 90% of intensifiers consisted of the top five variants and 20

variants occurred only once<sup>54</sup>. To put this into perspective, Rickford et al. (2007) found only distinct 31 variants in their data set combining four corpora<sup>55</sup>.

Table 17 below displays all variants that occurred at least ten times in the corpus (with intensifiers occurring fewer times collected in the ‘other’ category).

Table 17: Complete set of intensifiers in the Television Dialogue Corpus

<b>Intensifier</b>	<b>N</b>	<b>%</b>
<i>so</i>	2077	34.6
<i>very</i>	1759	29.3
<i>really</i>	1429	23.8
<i>completely</i>	129	2.1
<i>totally</i>	125	2.1
<i>extremely</i>	60	1.0
<i>incredibly</i>	56	0.9
<i>bloody</i>	41	0.7
<i>absolutely</i>	37	0.6
<i>all</i>	33	0.6
<i>super</i>	31	0.5
<i>awfully</i>	22	0.4
<i>entirely</i>	22	0.4
<i>damn</i>	20	0.3
<i>seriously</i>	16	0.3
<i>terribly</i>	15	0.3
<i>unbelievably</i>	14	0.2
<i>dead</i>	12	0.2
Other	100	1.7
Total	5998	100

For the analysis I merged variants that were semantically equal, such as *real* and *really*, *absolute* and *absolutely*, or *complete* and *completely*. The present analysis focuses on external variables and found the separation of the respective forms not vital to the results.

---

<sup>54</sup> Including astonishingly, bug-shaggin’, magnificently, jolly, mightily, tantalizingly, ultra, etc.

<sup>55</sup> Stanford tape-recorded corpus (STRC), Wimmer/ Fought tape-recorded corpus (WFTRC), multisource all corpus (MASC), google newsgroups corpus, introduced in Rickford et al. (2007: 5-6)

In the case of *real* and *really* and *absolute* and *absolutely*, Tagliamonte (2008: 369) and Barnfield and Buchstaller (2010) followed a similar route and combined both variants. A detailed analysis of internal syntactic variables and a focus on delexicalization would benefit from keeping these forms separate however.

Looking at the overall distribution (Table 17), the three most frequently used intensifiers are *so*, *very*, and *really*, as expected from previous studies. Following these, the next four variants used are all maximizers. This is striking, as it seems that the main boosters have very little competition. *Dead*, the most popular intensifier in the Tyneside PVC corpus, occurs only 0.2% of the time, whereas *pure*, the intensifier Macaulay (2006) found most popular in Glasgow, does not occur at all. A preliminary interpretation might be that the set of variants from which the scriptwriters choose is limited to either the main forms (*very*, *really*, and *so*) or marked forms that are used to make the characters stand out (tokens from the ‘other’ bin that have little reoccurrence overall but are considerably more frequent than was found in other studies, see above).

With frequencies remarkably distinct between the “popular” variant set of *very*, *really*, and *so*, the initial distributional analysis will focus on intensifiers that occurred at least 60 times, or made up at least 1% from the overall intensifier use. Lower frequency intensifiers are unlikely to contribute to characterizing patterns that index social meaning.<sup>56</sup> For the detailed multivariate analysis (section 5.5.6), all intensifiers are considered again.

---

<sup>56</sup> Although it is worth pointing out that some variants stand out for their innovative single-time use. However, considering the overall aims of the thesis, this will not be a focus here. An example for a marked use that indicates a particular aspect of characterization is the marked use of *bloody* by Buffy in a specific episode that signals that she is actually a programmed cyborg, further discussed in Reichelt and Durham (2017).

As with other sections before, I have summarized the use of the main intensifiers in Table 18 below.

Table 18: Intensifier use in the Television Dialogue Corpus (per 10,000 words)

<b>Character</b>	<b>Series</b>	<i>very</i>	<i>really</i>	<i>so</i>	<i>completely</i>	<i>extremely</i>	<i>totally</i>	<b>total</b>
Cordelia		3.86	5.98	10.99	0.2	0.19	0.19	24.10
Angel	<i>Angel</i>	3.78	2.67	5.67	0.0	0.11	0.33	13.46
Spike		2.93	4.39	2.93	0.0	0.00	0.00	19.03
Giles		20.41	0.62	2.89	1.6	2.27	0.00	34.22
Anya		8.80	6.60	11.89	0.4	0.88	0.88	33.02
Spike		7.14	2.48	7.45	0.0	0.00	0.31	24.82
Xander	<i>Buffy the Vampire Slayer</i>	6.84	4.96	8.55	0.5	0.00	0.68	24.10
Angel		5.44	6.35	5.44	0.0	0.00	0.00	20.87
Willow		4.96	10.91	13.06	0.0	0.00	0.99	31.91
Dawn		4.93	9.87	17.88	0.0	0.00	1.23	35.77
Cordelia		4.50	10.50	20.25	0.7	0.00	3.00	40.49
Buffy		4.26	6.95	7.58	0.3	0.08	0.63	22.50
Richard		32.21	1.12	5.88	0.6	1.68	0.00	44.81
Emily		21.99	0.53	15.46	1.3	0.93	0.00	42.64
Lorelai		12.92	9.39	13.27	0.9	0.38	1.17	40.28
Rory	<i>Gilmore Girls</i>	12.26	12.37	14.23	1.5	0.33	0.71	43.17
Sookie		10.52	8.85	18.65	1.4	0.72	0.48	42.56
Paris		7.24	5.25	7.99	1.2	0.75	0.75	23.98
Luke		6.48	7.77	5.18	1.1	0.12	0.24	22.37
Amber		16.12	20.38	32.24	0.9	0.00	2.37	73.48
Sarah		14.36	19.74	28.71	0.4	0.00	0.67	65.05
Max		11.60	9.81	9.81	0.0	0.89	1.78	34.80
Zeek		10.71	3.57	4.08	0.0	0.00	0.00	19.89
Kristina	<i>Parenthood</i>	10.52	20.51	36.02	1.3	0.26	1.84	72.04
Adam		9.28	9.45	6.01	0.5	0.00	1.20	28.35
Camille		8.96	14.55	24.63	0.0	0.00	0.00	51.49
Julia		8.46	16.92	31.03	1.7	0.00	2.26	64.88
Crosby		8.24	17.23	15.73	0.5	0.00	0.75	44.71
Haddie		0.80	30.44	20.03	0.8	0.00	2.40	57.68
Sherlock	<i>Sherlock</i>	11.85	2.07	6.52	0.6	0.59	0.30	26.08
John		11.02	1.84	6.74	0.0	0.00	0.61	24.50
Ianto		7.14	1.43	4.29	0.0	0.00	0.00	14.29
Owen		6.54	4.58	5.89	1.3	0.00	0.00	20.94
Toshiko	<i>Torchwood</i>	4.25	1.06	8.51	1.1	0.00	1.06	17.02
Jack		3.38	2.53	3.80	0.0	0.42	0.00	10.13
Gwen		1.94	3.40	10.68	0.5	0.00	0.00	19.43
Rhys		0.00	1.99	7.96	0.0	0.00	0.00	25.86
mean overall		8.83	8.08	12.38	0.58	0.29	0.73	33.91
	<i>Angel</i>	3.52	4.35	6.53	0.06	0.10	0.18	18.87
	<i>Buffy</i>	7.48	6.58	10.55	0.41	0.36	0.86	29.74
mean by series	<i>Gilmore Girls</i>	14.80	6.47	11.52	1.14	0.70	0.48	37.12
	<i>Parenthood</i>	9.90	16.26	20.83	0.62	0.12	1.33	51.24
	<i>Sherlock</i>	11.44	1.96	6.63	0.30	0.30	0.45	25.29
	<i>Torchwood</i>	3.88	2.50	6.85	0.48	0.07	0.18	17.94
mean by gender	female	8.51	11.27	18.06	0.78	0.24	1.09	42.18
	Male	9.17	4.72	6.38	0.37	0.34	0.34	25.18
mean by age group	Young	7.18	11.17	14.54	0.56	0.26	1.30	37.08
	middle	7.93	6.92	11.11	0.62	0.23	0.53	30.99
	Older	18.47	4.94	12.51	0.47	0.65	0.00	39.71
mean by nationality	American	9.39	10.19	14.59	0.61	0.29	0.91	38.09
	British	7.32	2.39	6.38	0.51	0.29	0.23	22.62

Some preliminary results from the table above describe the overall distribution of variants according to their function. Boosters are used much more frequently than maximizers (with an overall average of 9.7 per 10,000 words for boosters and only 0.5 per 10,000 words for maximizers). This mirrors findings by Bauer and Bauer (2002:247) and many of the aforementioned studies that found main boosters well ahead of maximizers.

In terms of overall intensifier use, *Parenthood* leads with a mean of just over 51 intensifiers per 10,000 words. In comparison to that, characters on *Torchwood* use approximately 18 of the six most frequent intensifiers per 10,000 words. Intensification, according to Quaglio (2009) and Tagliamonte and Roberts (2005), is higher on television (in line with higher degrees of emotionality). What we find in the *Television Dialogue Corpus* is the series that is, following results from previous sections, most like naturally occurring language, actually exhibiting the highest intensifier use. I claim that the higher use of intensification (or emotional language in general) is not simply because it is television, but because of the type of television (or genre). Thus, many series will have a comparably low use of intensification (such as *Torchwood*, *Angel*, or *Sherlock*), simply because the content of the series is not predominantly emotional. Series such as *Friends* (where higher intensification was found by Quaglio (2009) and Tagliamonte and Roberts (2005)), as well as *Gilmore Girls* (where higher emotionality features were found by Bednarek (2012)) exhibit a scripted environment in which characters talk a lot more about themselves and their feelings. Series such as *Torchwood* or *Sherlock* are much more story-driven and characters react to external action rather than introspection. *Parenthood*, showing the highest level of intensification in the Television Dialogue Corpus, is a series that has rather little action and is mainly concerned with character exposition and emotional content (see again series summary in section). The level of intensification thus might contribute to the differentiation of particular types of television writing. For a

thorough investigation of any further correlations however, additional data (target audiences, audience of networks, etc.) would have to be taken into account.

For the present study it suffices to acknowledge the differing genres of television that will doubtlessly influence scriptwriting to a certain degree.

Whether or not intensification is particularly high or low by genre however should not prevent any investigation of possible characterization patterns. The choice of the intensifier, rather than the frequency, is the main focus of this analysis.

The next section investigates character background groups (gender, age, nationality) across the corpus. There has been considerable research on intensifiers that provides an extensive backdrop for the present results to set against. I will highlight relevant theories where appropriate but will mainly focus on how distributions may relate to the construction of characters. It is important to note then that this part will not attempt to advance sociolinguistic knowledge of how intensifiers are used in English. Rather, I will use existing knowledge as a foundation to analyse whether scriptwriters are aware of and apply socially stratified uses of intensifiers to their fictional subjects.

#### **5.5.4 *Distributions across categories***

##### *Age and gender*

Most sociolinguistic studies that investigate intensifier use across social variables find age and gender to be productive predictors of intensifier choice. Patterns of use consistently reflect sociolinguistic theories of women as innovators in linguistic change from below (change that is unmonitored and not socially marked), cf. Labov's principle II of linguistic change (1990:215). Innovation within the variable set of boosting

intensification describes, among others, the ongoing shift from *very* to *really* to *so*.<sup>57</sup> Following that, studies would be expected to show women leading in the use of *so*, as well as of *really* (although depending on how far along the shift the respective community is, this might not be marked). Indeed, Xiao and Tao (2007:249) find women leading in the use of *really* ahead of *very* (reverse pattern for male speakers). Stenström et al. (2002:142-143), in their study of teenage language in London, find that girls are, again, ahead in the use of *really*. Ahead here does not just indicate that they exhibit a change from *very* to *really* as a preferred variant, a change that is not yet visible for boys, but also ahead in terms of intensification overall. The longitudinal Tyneside study by Barnfield and Buchstaller (2010:268) finds that women are leading in the 1990s (PVC corpus) with the use of incoming variants *really* and short-lived trend *dead*. The female preference in using incoming *really* continues into the 2000s (NECTE2 corpus) (2010:270). They further report that albeit *so* is steadily increasing in use, the increase is spread across factor groups and does not present itself as an innovative feature in the UK context. Meanwhile in Toronto, Tagliamonte (2008:388) finds *so* to be marked as a newcomer with women tending “to push it [*so*] forward”. However, she further notes that Torontonians men have their own incoming intensifier, *pretty*, that they favour. She suggests that this might be men “reacting against a distinct female trend – the use of *so*” (2008:389), which has important implications for the current study. If the use is a reaction to a pattern that is associated with female speech, we can assume a level of markedness which indexes *so* with female speakers. If *so* is indeed indexed with a particular gender, its use as a characterizing cue is probable. This is somewhat confirmed in Tagliamonte

---

<sup>57</sup> The system sees other shifts as well, as studies focusing on popular variants such as *pure* or *dead* illustrate. The shift from *very* to *really* to *so* however is visible across communities and seems to indicate a longitudinal replacement in the system rather than a relatively contained/short-lived spike of a feature that is limited to speaker groups, regions, etc.

and Robert’s study of intensification in the television series *Friends* (2005). The study is the first to find *so* ahead of any other intensifier (this was not the case in Toronto or any of the UK-based studies) and shows female characters ahead of men in the use of both *so* and *really* (*very* is equally distributed across both genders). This not only presents variation of intensifiers that is ongoing in naturally occurring language reflected in scripted contexts, but the social preference of the ongoing shift is seemingly amplified by the writers.

The plot (36) below shows the use of the most frequent boosters in the *Television Dialogue Corpus* (*very*, *really*, *so* from left to right) across gender.

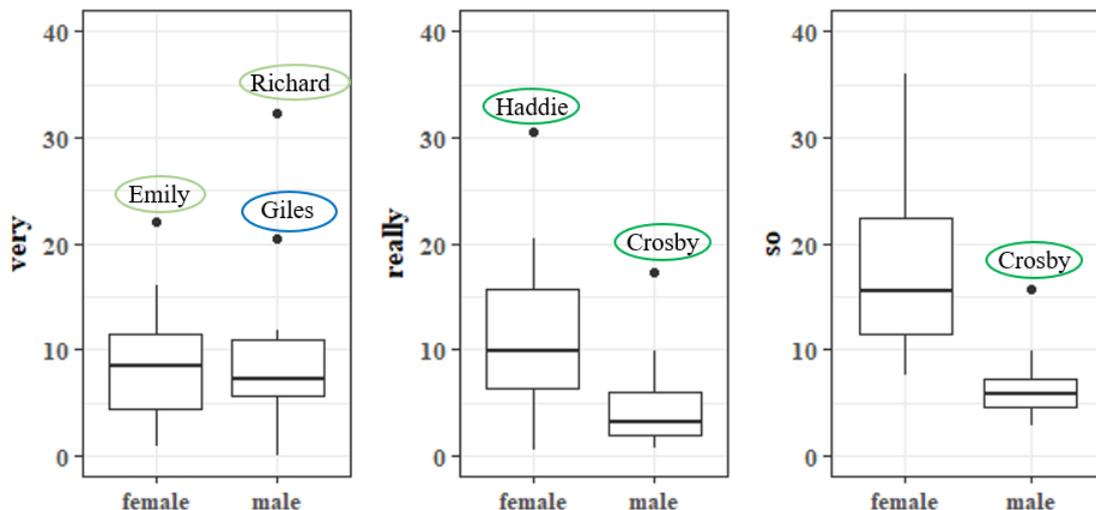


Figure 36: Main boosters across gender

The plot shows that gender stratification increases with the ‘newness’ of variant. This reflects previous findings of *very* being not particularly marked by gender. Both *really* and *so* show preference by female characters as well as overall increase of use (with *so* leading). For male characters *very* seems to be the preferred booster. Their low use of *really* is surprising. According to previous research (see above) it is the favoured

intensifier in America and rising in the UK. The low use by the male characters in the Television Dialogue Corpus thus is unexpected.

Before turning to the gender distribution of maximizers, I want to highlight some of the outliers above. *Very* is used at outstanding frequencies by three characters: Richard, Emily, and Giles respectively. This again reflects some of the findings from previous sections where these three were often aligned in their patterns of language use. Richard and Emily, the grandparents in *Gilmore Girls*, differ from the rest of the ensemble not just by their age, but their social class and formality. This formality presents itself in British patterns, particularly in speech patterns of Giles, the British librarian in *Buffy*.

British intensifier trajectories, as studies mentioned above have found, trail behind those of American English developments. Extrapolating from that, it seems sensible to see older American speakers using forms that are still very much present in British English. However, the Television Dialogue Corpus shows this not to be the case consistently for all characters, indicating that there are additional layers of characterization at play. All three characters that are outliers for *very* have in common that they are outsiders among a young and trendy ensemble. The outsider status of being ‘out of touch’ is exemplified through the heightened use of older, more formal, variants. This pattern is clearly marked with intensifiers, but can be traced back throughout the other features discussed.

For *really*, the outliers are Crosby and Haddie from *Parenthood*. Haddie, in many ways, follows stereotypical American young girl patterns (also seen with an increased use of pragmatic marker *like* for instance), which makes her high use here unsurprising. Crosby’s use of intensifiers is not just particularly high for *really*, but also for *so*. Previous features (in particular pragmatic markers) have shown him to use high frequencies of new, incoming devices. In the series he is portrayed in stark contrast to his brother Adam,

who is more serious and grown-up. Crosby uses features (by type of variant as well as frequency) that are indexed as youthful and trendy and, in the case of intensifiers, that are more associated with how female characters speak. His linguistic behaviour does not fall out of the norm of how the features are used generally, but rather out of the norm of how they are used by the other male characters.

The gender distribution for the main maximizers *completely*, *totally*, and *extremely* (Figure 37) shows female characters leading in use for *completely* and *totally*, while *extremely* is used fewer times altogether.

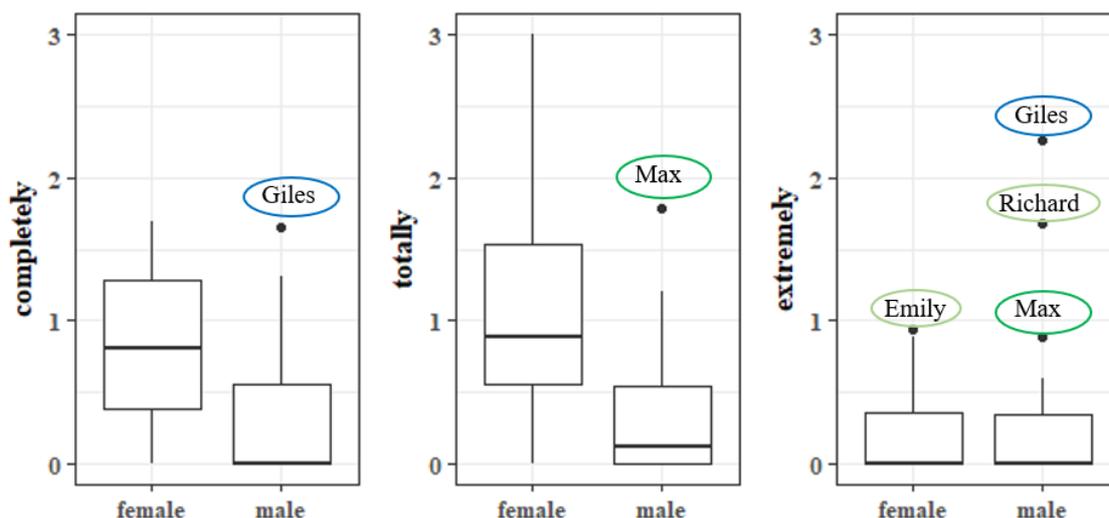


Figure 37: Main maximizers across gender

With low numbers overall, it is difficult to interpret the findings in terms of heightened intensifier use by one gender, or ongoing language change. A tentative claim would be that, similar to booster distributions, female characters make more use of maximizers overall.

Looking at the outliers, Giles (*completely* and *extremely*) and Richard (*extremely*) again stand out – further underlining previously made points on how their characterization of being outsiders within their respective ensembles shows similar linguistic realizations. Additionally, we find Max (*Parenthood*) using *totally* and *extremely* at higher rates than other male characters from the corpus. This may correspond to how his attention and level of prioritization is portrayed on the series. Much of his storyline is linked to his social struggles in being diagnosed with Asperger’s syndrome (see also discussions in previous sections). He generally demonstrated lower rates of features that are tied to pragmatic competence: hedges, mitigators, turn-taking devices. The higher uses of totality stance here might reflect the stereotypical representation of Asperger’s: being unable to deal with uncertainties and overly focusing on their requests and issues. Max certainly is portrayed this way which is reasoned with through his diagnosis. Example (65) shows Max reacting to his mother cancelling a trip because he called her ‘bitch’:

- (65) Max: This is *so* unfair! I hate you! This sucks! (...) This is *totally* unfair! You could've given me a different punishment! You are *so* mean! You are the meanest woman in the world. I cannot believe how mean you are. I wish you weren't my mother. I'm never talking to you again. (*PH*)

Much of his presence in this episode (episode 12, season 3) is defined through his anger. Linguistically, this is not just emphatically expressed through repeated use of intensifiers (*so* and *totally*), but also through superlative (*meanest*), calling upon the extreme unlikelihood of her supposedly bad behaviour (*I cannot believe how mean you are*), and giving hyperbolic consequences (*I wish you weren't my mother. I'm never talking to you again*). His use of maximizers (and intensifiers in more general terms) is aiding the

portrayal of his tantrums and a heightened use (in comparison to other characters in the corpus) shows that in these scenes his levels of emotional commitments are exceptional.

Moving on from gender distinctions to age, previous studies have found that intensifiers undergo constant change. In terms of age distribution that means that incoming features are used first by younger speakers before spreading to wider speaker groups (apparent time evidence). An important distinction that can be made here is between short-lived popular features, such as *dead* in the 1990s in Tyneside, or *pure* in Glasgow, that will be frequent in the youngest age group but will probably not expand further, and features that are entering the system and replace existing forms (such as the shift from *very* to *really* to *so*). This relates to theories of age grading and generational change, as discussed in chapter 2.

Previous studies have found the overall shift in intensifier use from *very* to *really* across North American (Tagliamonte 2008) and British (Tagliamonte 2003, Barnfield and Buchstaller 2010) contexts as reflected in different age groups. *So*, as the latest incomer, has not yet consistently made it into findings, although the trend of its rising popularity can be seen in studies such as Tagliamonte and Roberts (2005) or Tagliamonte (2008) (both discussed above). Many studies focus on young speakers as the most productive group for finding new trends (Bauer and Bauer, 2002; Macaulay, 2006; Martinez and Pertejo, 2012; Stenström et al., 2002). They include variants that are specific for young speakers, but also indicate where wide-spread change might appear to enter the system.

As indicated earlier in this section, intensifier shifts often correspond to differences in gender preference and the Television Dialogue Corpus indeed reflected this ongoing change by having female characters using newer, incoming features at higher rates than their male counterparts. The next part of the analysis will investigate whether this gender

difference indeed reflects change and what other patterns can be found, focusing on age differences.

Figure 38 shows how the top three boosters are distributed across the three age groups: young, middle, older.

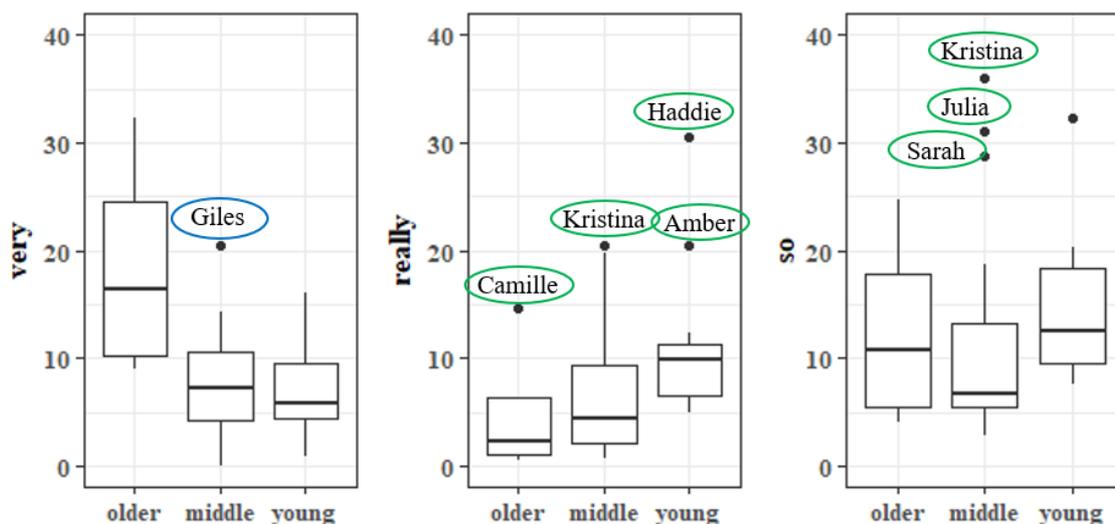


Figure 38: Main boosters across age

For *very*, the distinction between older and newer speakers is clearly visible: the older characters show a preference for *very*, not only within the variant group (across age), but within the whole variable set (across variants). The distinction between middle and younger speakers is not as visible, which may be due to interaction with other character factor groups. The outlier in the middle-aged group for *very* is, again, Giles. He uses *very* at rates that adhere much more to the older character group, which goes back to characterization patterns I have shown above.

*Really* shows a distribution closely resembling that of other studies, with a constant increase of use in time. Outliers for *really*, as well as for *so*, are exclusively female

characters from *Parenthood*. Camille, Kristina, Amber, and Haddie for *really* and Sarah, Kristina, Julia, and Amber for *so* are strikingly high in their use of intensifiers. It seems reasonable to assume that individual character patterns are a secondary reason for this distribution, and that, instead, the series itself shows a preference for female-led intensifier use. This goes back to a previously made claim on how the genre might impact how certain language features are used. *Parenthood* is one of the more emotional series within the corpus and it is thus unsurprising to find exceptional levels of intensifier use. The fact that this is mainly observable with the newer boosters *really* and *so* underlines claims about new intensifiers coming into use in order to add levels of emotionality and emphasis to utterances. All the outliers are female characters, which might correspond with indexing intensification with female language. Another possibility might be that female characters are given more contexts in which intensification is appropriate.

The overall distribution of *so* is surprisingly level across all three ages, with the middle-aged group using slightly fewer variants. Comparing age preference across all variants, it seems that the middle-aged groups is the least varied in their use of boosters. A number of reasons possibly interact here: with ongoing shift, the middle-aged group will be in-between more traditional and incoming features and will inevitably showcase use from both sides. The middle-aged group is also much less indexed than the other groups. Stereotypes on language use often singularize particular speakers and speaker communities: typical youth slang is up against traditional and archaic language. The middle-aged group is not only in between these two groupings, but it is also the group that is responsible for what is considered 'normal' and not indexed in any particular way.

We can thus assume that when there are patterns out of the norm to be found in the middle-aged character group, that these patterns correspond to other character factors (such as gender or nationality) or individual characterization.

Maximizers across age indicate three stages of use: stable, increasing, and decreasing.

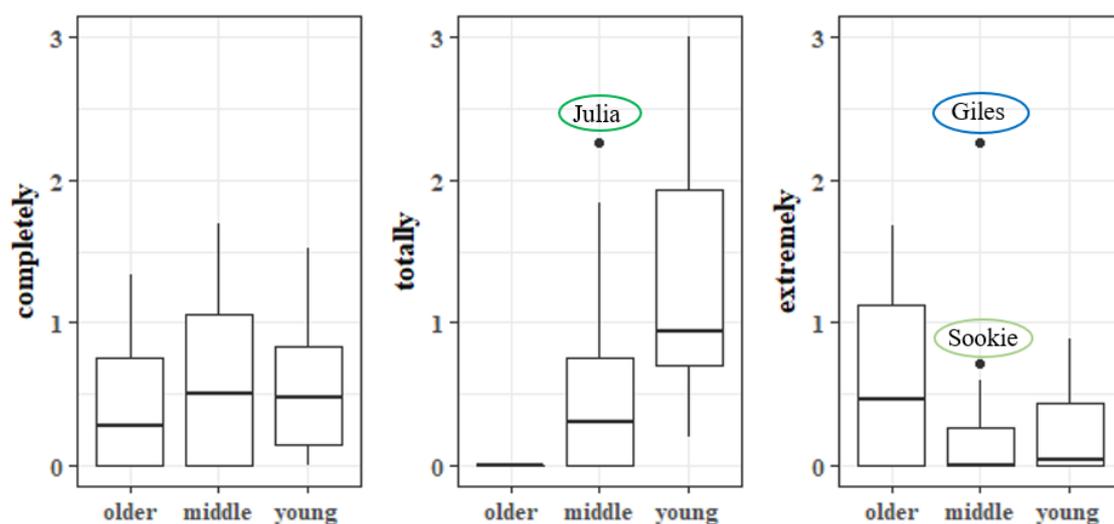


Figure 39: Main maximizers across age

*Completely* is not particularly marked across age and also shows no outliers. The previously found slight preference by female characters does not seem to indicate language change here. For *totally* we find a stark shift from older characters to youngest. The outlier in the middle age group is *Parenthood*'s Julia. It is worth pointing out here that even though not visible in the above plot, the other women in *Parenthood* are all at similar frequencies of use. This mirrors findings (and interpretations) from the boosters, discussed previously. Finally, *extremely* shows a slight decrease with time and Giles, once again, as an outlier in his character group.

Taken together, the results from this section have shown a clear picture of ongoing shifts of boosters (*very* to *really* to *so*), and to a degree, a shift of maximizers (*totally* entering the system). Further, some character-specific distributions can already be discerned from these rather broad foci. In particular, *Parenthood* stands out with a high use of intensification, possibly due to emotional storytelling. As with other features previously

discussed, there also appears to be a character-grouping of Giles (*Buffy*) with Emily and Richard (*Gilmore Girls*). The analysis of hedges (5.2) and pragmatic markers (5.1) has attributed this grouping to Britishness corresponding with American upper social class. Whether that holds up here as well will be further investigated in the following section that deals with nationality.

### *Nationality*

As mentioned above, it seems that the general intensifier shift (whereby *very* is replaced by *really* and so enters the system) happens in both North American and British English contexts, albeit at different paces. While North America, i.e. Canada (Tagliamonte 2008) and *Friends* (Tagliamonte and Roberts 2005), already shows *very* as having lost its footing ahead of other intensifiers, British English has *very* leading in all studies that include a cross section of speaker groups (not just teenagers) (Barnfield and Buchstaller 2010, Ito and Tagliamonte 2003).

Figure 40 shows the use of the three main boosters across nationality of the Television Dialogue Corpus characters.

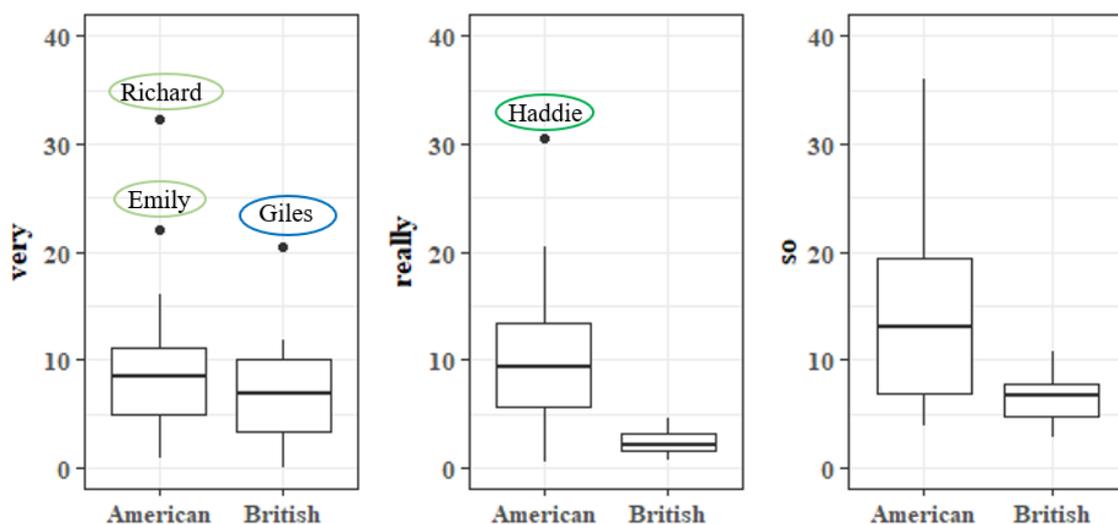


Figure 40: Main boosters across nationality

Overall, comparing the distribution across all three variants, it can be seen that American characters use more intensifiers overall. British characters prefer *very* over *so* over *really*, which is surprising. On the other hand, the lack of *really* by British characters marks them as distinct from North American contexts, where *really* is claimed to be the most used intensifier (Labov, 1984:44). The *Television Dialogue Corpus*, for North American characters, shows the expected trajectory of *very* being least frequent, followed by *really*, then *so*.

The analysis of each variant on its own repeats some of the findings already investigated in this section. The outliers for *very* group Giles with Richard and Emily, again. All three use the variant at outstanding frequencies and it can be assumed that *very* indexes particular characteristics that, even though not equivalent for all three, do match up in some common quality. The outlier for *really* is Haddie (*Parenthood*). While her high use of *really* does follow common American English patterns, expectations for a young American woman would have pointed towards *so* more so than *really*. In terms of

maximizers, it is worth pointing out that most studies have found *absolutely* leading for UK contexts. In the *Television Dialogue Corpus*, that variant trails behind other maximizer variants. Again, there is an overall higher use of intensifiers by American characters. Similarly to *so*, *totally* seems to be the most distinctly American maximizer with Cordelia (*Buffy*) using exceptionally high numbers. Toshiko (*Torchwood*) is a marked exception to the British group.

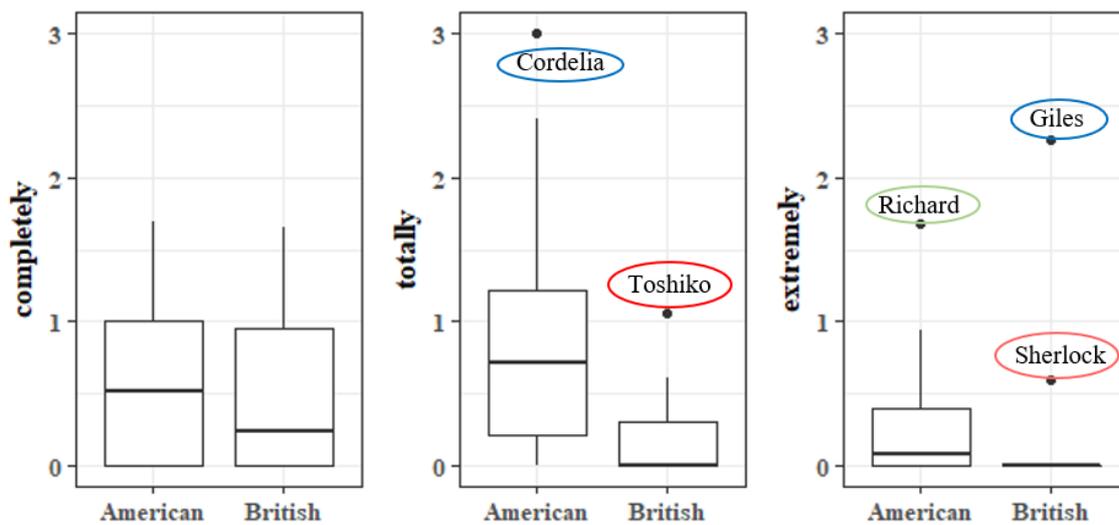


Figure 41: Main maximizers across nationality

The use of *extremely* is worth pointing out here. The previous plot of use across age has shown that *extremely* is most used by the older characters. For many features here and in previous sections, a preference by the older characters would correspond to a preference by British characters also. This however is not the case here. Merely two characters show heightened use for British characters (Giles and Sherlock) with most other characters not using the variant at all. Unfortunately, with numbers consistently low, a clear interpretation cannot be made.

The distributions across nationality have shown that patterns found in naturally occurring language are mirrored in scripted language as well. There are some patterns however, that differed (such as the low use of *really* for British characters). What these initial analyses only partly capture is how these categories (gender, age, nationality) are interacting, and how they are used as possible characterization cues. The following multivariate analysis considers all character categories in combination to find out where correlation between intensifier choice and characterization lies.

### 5.5.5 *Multivariate Analysis*

The following analysis was done in Rbrul, as detailed in the methods section (4.4). Because the numbers for maximizers were too low to allow for this sort of analysis, I will focus on the main boosters only. Each booster will be investigated separately, with the remaining set (including all other 30 variants) as the binary counterpart.

The character factors included as predictor variables (independent variables) are gender, nationality, and series. The character was set as the random intercept to account for individual speaker bias. Last but not least, the variants were separated by age group. As the analysis above has shown, the features are used differently across ages. Separating the tokens by age group thus enables me to investigate possible preferences within each age group in more detail. A similar method was used in Ito and Tagliamonte (2003) and Tagliamonte (2008).

The tables for the analysis include all “three lines of evidence” (Tagliamonte, 2006:235):

- i. statistical significance, i.e. Which factors are statistically significant at the .05 level and which are not?
- ii. relative strength, i.e. Which factor group is most significant (largest range) or least (smallest range)?
- iii. What is the order (from more to less) of factors within a linguistic feature (constraint hierarchy)?

The provided tables indicate non-significant results by square brackets around the factor weights. In line with previous comparative studies, I included the actual factor weights, as even a non-significant result can indicate a possible tendency. The range for each social variable is given underneath the factor weights. The order of the factors is predetermined, but easily distinguishable by size.

The first analysis (Table 19) shows *very* across the different character factors.

Table 19: Multivariate analyses of *very* by age group

	<b>young</b>		<b>Middle</b>		<b>older</b>	
<b>total N</b>	2085		3348		565	
<b>overall proportion</b>	0.23		0.29		0.55	
	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>
<b>SEX</b>						
<b>male</b>	.59	180	[.55]	1157	.62	199
<b>female</b>	.40	1905	[.45]	2191	.38	366
<b>range</b>	19		10		24	
<b>NATIONALITY</b>						
<b>American</b>	n/a	n/a	[.51]	2837	n/a	n/a
<b>British</b>	n/a	n/a	[.49]	511	n/a	n/a
<b>range</b>			2			
<b>SERIES</b>						
<b><i>Gilmore Girls</i></b>	[.64]	888	[.53]	1646	.64	480
<b><i>Parenthood</i></b>	[.42]	266	[.39]	1023	.56	85
<b><i>Buffy</i></b>	[.47]	806	[.61]	269	n/a	n/a
<b><i>Angel</i></b>	[.45]	125	[.39]	147	n/a	n/a
<b><i>Torchwood</i></b>	[n/a]	n/a	[.43]	135	n/a	n/a
<b><i>Sherlock</i></b>	[n/a]	n/a	[.65]	128	n/a	n/a
<b>range</b>	24		22		8	

The overall proportion shows the decrease in use of *very* from older to younger characters which was discussed above.

The factor-by-factor analysis further indicates an overall preference by male characters (significant in the older and the younger age groups). Nationality shows little preference (and no statistical significance) despite previous studies indicating a clear difference. The Television Dialogue Corpus shows an unusually high use of *very* by American characters. This was explained somewhat for the older character group where we have found that Richard and Emily use particularly high frequencies of *very*. The analysis here shows that this distribution carries through to the middle age group as well, if not at such a high level. The next analysis will add some insight to this distribution and highlight how *really* is used across character factors.

Similarly to *very*, *really*'s trajectory across time is visible in the overall proportion which is steadily decreasing with time.

Table 20: Multivariate analyses of *really* by age group

	<b>Young</b>		<b>middle</b>		<b>older</b>	
<b>total N</b>	2085		3348		565	
<b>overall proportion</b>	0.29		0.24		0.05	
	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>
<b>SEX</b>						
<b>male</b>	[.55]	180	.55	1157	[.48]	199
<b>female</b>	[.44]	1905	.45	2191	[.52]	366
<b>range</b>	<i>11</i>		<i>10</i>		<i>4</i>	
<b>NATIONALITY</b>						
<b>American</b>	n/a	n/a	.59	2837	n/a	n/a
<b>British</b>	n/a	n/a	.41	511	n/a	n/a
<b>range</b>			<i>18</i>			
<b>SERIES</b>						
<b><i>Gilmore Girls</i></b>	[.46]	888	.60	1646	.19	480
<b><i>Parenthood</i></b>	[.59]	266	.66	1023	.81	85
<b><i>Buffy</i></b>	[.49]	806	.34	269	n/a	n/a
<b><i>Angel</i></b>	[.44]	125	.48	147	n/a	n/a
<b><i>Torchwood</i></b>	n/a	n/a	.57	135	n/a	n/a
<b><i>Sherlock</i></b>	n/a	n/a	.36	128	n/a	n/a
<b>range</b>	<i>15</i>		<i>32</i>		<i>62</i>	

In contrast to *very*, it is the middle-aged group that shows a significance gender preference for *really*. Male characters are using *really* significantly more than female characters and American characters exhibit an unsurprisingly high use of *really*. British characters are using *really* at much lower frequencies than American characters, indicating that a lack of *really* in American series might be used to index an outsider status for British characters. In terms of series we can see that *Parenthood* and *Gilmore Girls* are significantly more likely to include *really* in their dialogue than *Buffy* and *Sherlock*. For the former two this preference makes sense because they are series with highly emotional family story telling whereas *Sherlock*, as a modern take on a procedural, is much more emotionally withdrawn. *Buffy*'s factor weight is likely explained through Giles who shows a very low use of *really* overall (in line with his outsider status in the ensemble). Even though not entirely clear from the table, looking at the series factor weights for the older group we can see *Gilmore Girls* at much lower (and significant) rates than *Parenthood*. While *Gilmore Girls*, for the middle-aged group, was likely to use *really*, the opposite is the case for the older characters: Richard and Emily. Again, it seems, we can detect how these two characters use features much more in line with Giles' patterns. Finally, *so* reflects an ongoing intensifier replacement that was already visible in similar ways on *Friends* (Tagliamonte and Roberts 2005).

Table 21: Multivariate analyses of *so* by age group

	<b>Young</b>		<b>Middle</b>		<b>older</b>	
<b>total N</b>	2085		3348		565	
<b>overall proportion</b>	0.36		0.34		0.30	
	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>	<b>FW</b>	<b>N</b>
<b>SEX</b>						
<b>male</b>	[.47]	180	.39	1157	.34	199
<b>female</b>	[.52]	1905	.60	2191	.66	366
<b>range</b>	5		21		32	
<b>NATIONALITY</b>						
<b>American</b>	n/a	n/a	.59	2837	n/a	n/a
<b>British</b>	n/a	n/a	.41	511	n/a	n/a
<b>range</b>			18			
<b>SERIES</b>						
<b><i>Gilmore Girls</i></b>	[.43]	888	.35	1646	[.44]	480
<b><i>Parenthood</i></b>	[.51]	266	.44	1023	[.56]	85
<b><i>Buffy</i></b>	[.49]	806	.44	269	n/a	n/a
<b><i>Angel</i></b>	[.56]	125	.56	147	n/a	n/a
<b><i>Torchwood</i></b>	n/a	n/a	.63	135	n/a	n/a
<b><i>Sherlock</i></b>	n/a	n/a	.58	128	n/a	n/a
<b>range</b>	13		28		12	

In terms of gender we can see significant preferences by female characters in the older and middle-aged groups. The range of factor weights for the oldest characters is highest and goes down the younger the characters are (with the youngest group not showing a significant difference between male and female characters anymore). This might indicate that younger characters are using *so* in broader and equivalent contexts – a sign that it might cease to index female language (but not quite the index of being part of younger language).

The variant is, as discussed previously, more typical for North American characters. However, looking at the numbers for series preference, the results are unexpected. *Torchwood*, *Sherlock*, and *Angel* are more likely to use *so* and these are the three series with arguably the least emotional storytelling. It is possible that this indicates where series

within their ensembles are varied, i.e. how distinct the individual characters are. We know that *Buffy* and *Gilmore Girls* have rather distinct character groupings (Giles vs. the younger characters, Emily and Richard vs. the more informal characters). *Torchwood*, *Sherlock*, and *Angel* are thus less varied within the respective ensemble and the variant choice seems thus more decisively in favour for one variant over others.

### 5.5.6 *Character-based analysis*

The final piece of analysis for intensifier use in the *Television Dialogue Corpus* investigates intensifier variation as a whole for each character. Here, I am interested in the overall choice each character has. In particular, I want to see which characters are limited in their variable set, as that might indicate that they are given specific variants at most times.

The plot below summarizes each character's complete use of adjective intensification with the lesser used categories on the left (other), followed by maximizers (*absolutely*, *completely*, *extremely*, *incredibly*, *totally*, and *all*), and finally boosters (*super*, *bloody*, *really*, *so*, *very*).

The rates are normalized as percentages, showing each character's complete intensification as 100% and individual variants broken down by percentage point. This enables a comparison of how patterned each character is, but also what variability can be found within and across series.

The plot orders the characters by series in the following order:

*Torchwood*, *Sherlock*, *Parenthood*, *Gilmore Girls*, *Buffy*, *Angel*

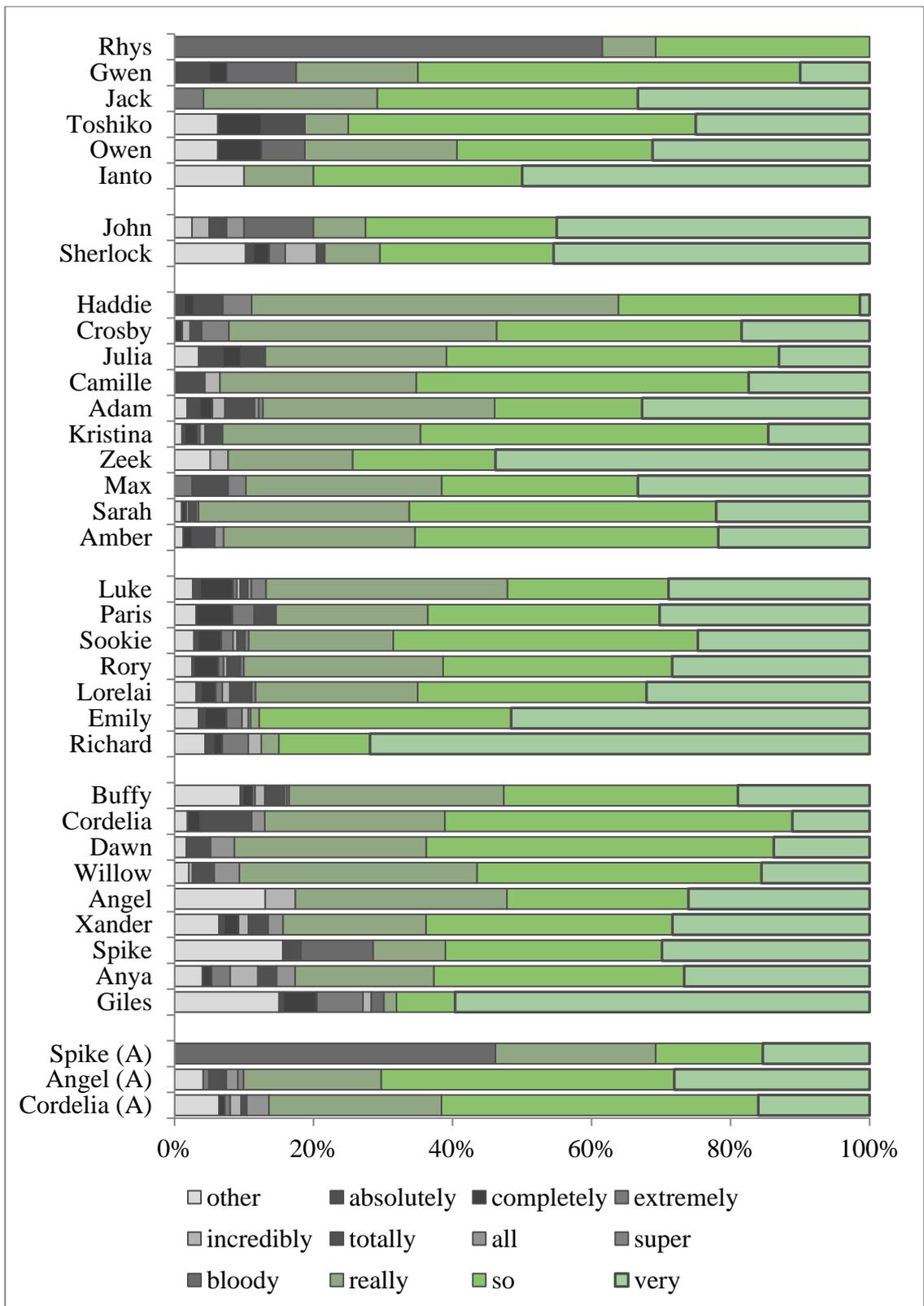


Figure 42: Intensifier variation across all characters and series

The main intensifiers are highlighted in green and it is clear that they dominate the intensifier choice throughout the Television Dialogue Corpus. Some characters however have noticeably few forms of one of the three main boosters. Giles, Spike (*Buffy*), Richard, Emily, Sherlock, John, and Toshiko have low percentages of *really*. Outside of Richard and Emily, these characters are all British, albeit other British characters using the variant prevents me from interpreting a lack of *really* as a British index. Going back to interpretations from other features, we might find formality indexed here. In *Buffy*, Britishness is portrayed with a certain level of affectedness, with Giles being stuffy and out-of-touch and Spike, especially in the beginning of his character arc, a stereotypical hyperbole of punk. While not strictly speaking formal, their characterization can be seen as English pretence. This is somewhat reflected in Sherlock as well. In contrast to highly emotional commitment by American characters, Sherlock (and, to a lesser degree, John) seems comparatively underwhelmed most of the time. This underlines perceptions of slight pre-eminence that links to formality and upper class, as we find it with the Gilmore grandparents. All of these characters are portrayed as somewhat posh and thus distinct from other characters.

Going back to the overall distribution, we find two characters that stand out for their high use of a variant that is overall not frequently used: *bloody*. Carrying some social stigma as a taboo and expletive, the variant occurs almost exclusively with British characters. Two of them, Rhys and Spike (*Angel*), use *bloody* more than any other variant and are thus not only indexing their Britishness, but also a level of non-standardness, and, certainly in Spike's case, being deviant. Throughout his appearance on *Buffy*, Spike uses fewer instances of *bloody*, although it is still his third most often used intensifier overall. Other British characters (Giles, Gwen or John for instance) use this variant to a lesser degree, albeit still noticeably so.

### 5.5.7 *Intensifiers as characterization devices*

In summary, the analysis of intensifier use in the *Television Dialogue Corpus* has shown that the variable set itself is quite extended (74 variants) but that characters do not vary consistently across all items. Only six variants are used with a proportionate frequency of at least 1% (60 occurrences overall), creating a system where choice competition is among few selected variants (i.e. *very*, *really*, *so* for boosters, *extremely*, *completely*, *totally* for maximizers).

Variant preference and change mirrors that of naturally occurring language.

Social factor group predictors are strongest for incoming features as they will be more socially marked and contribute to stereotyping. This is visible with *so*, which is significantly higher in use for female and American characters.

As with the other analysis sections, I will briefly summarize findings for intensifier use in the *Television Dialogue Corpus* with regards to the research questions.

- Does the distribution of intensifiers reflect characterization patterns within and across series?
  - As seen with other features in previous sections, certain characterization patterns are more marked with certain variants. *Very* for instance is used throughout the corpus by almost all characters, but it is the over-, or under-use that seems to contribute most to characterization (further detailed below). Other variants, such as *bloody*, only appear with specific characters and are thus marked in slightly different ways. Overall, the analysis has shown that intensifier use carries social

meaning which is exploited for the characterization in fictional television series.

- Does the distribution indicate indexical and/or stereotypical characterization and if so, what characters or character groups are represented like this?
  - Again, young female characters correlate with heightened use of specific variants (in particular ‘new’ intensifiers *so* and *totally*). We also saw a correlation of older characters using older/more established forms (which reflects ongoing language change). Indexical intensifier use is multi-layered, i.e. no one variant is indexing one characteristic. This for instance is shown with Max’s use of intensifiers which clusters in moments where he is shown to be overly emotional due to his Asperger’s diagnosis. The distribution of British intensifier use is not as easily discernible as with other features. However, as previously seen, patterns of Giles again correspond with those of older American characters, in particular with *Gilmore Girls*’ grandparents.
  - Social network representation here matches that of previous sections. Particularly consistent here are Richard and Emily (*Gilmore Girls*) and, in *Parenthood*, all female characters as well as Crosby. This is a repeated pattern that was already visible with GEs, and highlights particular character traits by Crosby (further detailed below).
- Are distributional outliers and unexpected patterns indicative of specific characterization choices?
  - As indicated above, outliers seem to be particularly telling in terms of characterization patterns. Intensifier outliers within the series usually point to character outliers within the story lines. Crosby differs from

other male characters and aligns more closely with the female characters from *Parenthood*. His character arc throughout the included four seasons revolves around him being less conservative, but also immature and free-spirited in comparison to his older brother Adam. This distinction manifests itself in Crosby's use of intensifiers (particularly *really*, *so*, and *totally*), which is closer to how his sisters Sarah and Julia, as well as his niece Amber use intensifiers. Here, we can see that intensifier use indicates group dynamics within series. Other notable outliers are Richard and Emily in *Gilmore Girls*. Their intensifier use follows predictions informed by naturally occurring language patterns (higher use of *very*, use in line with British patterns because of expected preference shift that is observed in younger generations). Outliers in this regard are cementing perceptions of characters by behaviour that is familiar to the audience.

Throughout the section above, it became clear that the ongoing shifts within the variable set are used to the advantage of characterization. Patterns that are well established (and partly stereotyped) are used to highlight fictional distinctions. Distinctions between series correlate with the emotional aspect of the genres as well as the characters part of the series (family dramas with young female characters for instance show higher overall frequencies of intensification than detective stories with middle-aged men in the centre of the plot).

## 6 Change and Characterization analysis

This chapter explores, if briefly, the notion of language change. It does so both in terms of systematic ongoing language change as well as linguistic change found within speaking patterns of the characters under investigation.

The first half of this chapter is concerned with a quantitative approach to the process of delexicalization. For this, I present a continuative frequency analysis of pragmatic markers and hedges (as introduced in 5.1 and 5.2 respectively) with references to their multifunctionality in discourse. My aim here is to further explore how linguistic features that are undergoing change in use and function are used in scripted contexts. A secondary aim is to see whether scripted language might be an appropriate source to track linguistic change more generally.

The second half of the chapter is concerned with linguistically expressed character change. With characterization patterns apparent in characters' language use, I pose a follow-up question: Is a character that is undergoing change within the plot, also showing linguistic change? The analysis in chapter 5 has shown that characters' personalities and attributes correlate with particular uses of linguistic features. This section will focus on two character types in particular to see how far linguistic patterns are adapting.

### 6.1 The discourse value as an indicator for linguistic change

The discourse value, or d-value, describes the ratio (in percentage) of pragmatic functions a linguistic form has with reference to their grammatical function (cf. Aijmer, 2002; Beeching, 2016:33; Stenström, 1990):

$$d - Value = \frac{\textit{occurrence of pragmatic functions}}{\textit{all occurrences}} \times 100$$

Markers that are used exclusively in pragmatic contexts will show a d-value of 100%, an example for that is *oh* as a pragmatic marker indicating surprise or attention or *uh/uhm* as a pragmatic marker of hesitation or turn taking management.

In that, the d-value has similar methodological scope as the sort of normalization I have employed in the previous analysis, i.e. number of occurrences per 10,000 words. In addition to this however, the d-value also provides insight into the development of multifunctional pragmatic features, as Beeching (2016:76) illustrates:

Generally speaking, we can assume that, in a representative sample, speakers will use forms for their canonical grammatical usages at a relatively stable rate. Where we see a marked rise in frequency, it is reasonable to assume that this rise is due to an increase in pragmatic marking usage, as these semantically bleached and pragmatically enriched forms are characterised by an ease of implementation and the possibility of being inserted at a number of different points in the discourse.

This means that, when comparing d-values of linguistic features across speakers or speaker groups (e.g. different age groups), varying d-values might indicate that the pragmatic context that is available to these features is undergoing some kind of change (either in terms of overall systematic language change, e.g. the expansion of possible syntactic contexts a marker can appear in, or in terms of character-based indexicality).

With regards to indexicality, Beeching (2016:77) writes that for sociolinguists, the d-value offers an opportunity to “chart the indexicality of markers as identity features and to explore how linguistic change spreads through a population, using apparent time (a comparison across different generations) as a means of assessing the extent to which a pragmatic usage is an incoming form.” In summary, using the d-value as a reference point for the analysis, I aim to see whether linguistic change (delexicalization of pragmatic devices) can be observed in the Television Dialogue Corpus.

All calculations for plots and tables in this section are based on the frequencies presented in the subsections 5.1 and 5.2. The next sections are similarly structured to previous feature-focused sections in that I will first analyse broad patterns of d-values across the features, before turning to social factors of gender, age, and nationality. Finally, the sections will end with an exploration of character-based distributions of discourse values.

### 6.1.1 D-Value of pragmatic markers

The plot below shows the d-value of the three markers *you know*, *I mean*, and *like* from the Television Dialogue Corpus. Pragmatic marker *I mean* is at almost 100%, indicating that the grammatical function of *I mean* is seldom used in scripted dialogue. *You know*, having several more propositional functions, comes to roughly 50% pragmatic uses across all series. *Like* has a relatively low d-value, which can be explained through its multifunctional uses (see 8.1 for a summary).

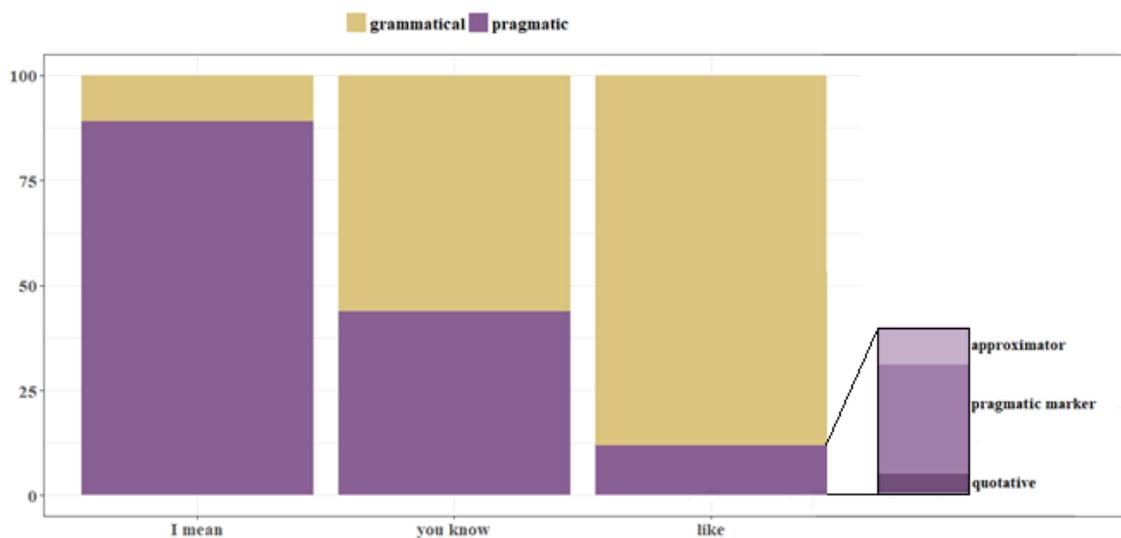


Figure 43: D-value of pragmatic markers

A detailed breakdown of *like's* pragmatic functions (purple area at the bottom right of the above figure) shows that quotative use is lowest of the functions given. This can be

explained through the medium itself. One credo for scriptwriters for television is ‘show, don’t tell’ (c.f. chapter 2) making retellings of previous happenings through reported speech a rare occurrence. Storytelling employs means such as flashbacks or the common ‘previously on’ feature to remind audiences of previous dialogue that took place.

The value does not offer opportunity to compare the pragmatic markers with each other beyond the fact that some of them have, alongside their pragmatic function, propositional meaning available to them. However, comparing the d-value of the markers to previous research can offer insights into the nature of television language and its usage in comparison to naturally occurring language, as well as possible changes in use over time. As previously mentioned, scripted language for fictional television is generally less vague. One reason for that is that discourse markers such as *you know* imply shared knowledge between speaker and listener. Consequently, they might be less used in scripted contexts where the writers not only cater for shared knowledge between the characters on screen, but also characters and audiences, a situation where shared knowledge cannot be assumed as easily.

In Table 22 the Television Dialogue Corpus is set against results which Beeching (2016) collected from subsets of the British National Corpus (*BNC*) and a corpus of role-plays from the University of West England (*UWE*)<sup>58</sup>. The nature of the Television Dialogue Corpus should relate most closely to the *BNC* subcorpus of leisure (spoken), although there will be examples of dialogue that are also part of other genres, such as institutional in *Sherlock*, business in *Parenthood*, or education in *Buffy the Vampire Slayer* and *Gilmore Girls*.

---

<sup>58</sup> Spanning a timeframe from 2011-2014, at the end point of the TV corpus’ time frame

Table 22: D-value (%) by genre, adapted from Beeching (2016:105; 191;135)

<b>d-value by genre</b>	<i>you know</i>	<i>I mean</i>	<i>like</i>
<b>Leisure</b>	86%	88%	12%
<b>UWE Role-Play</b>	88%	86%	60%
<b>Spoken demographic</b>	68%	84%	32%
<b>Education</b>	50%	88%	10%
<b>Business</b>	76%	92%	12%
<b>Institutional</b>	66%	94%	8%
<b>Television Dialogue Corpus</b>	<b>56%</b>	<b>94%</b>	<b>11%</b>

For the pragmatic markers *you know* and *like*, the Television Dialogue Corpus shows that they are much more infrequent in scripted language than in the other genres (except for education). This might relate to the point previously made; television language needs to be concrete and easy to follow for the audience and thus features that are seemingly superfluous, imprecise or vague will not be written into the dialogue. *I mean*, on the other hand, has a d-value higher than (or equal to) the other genres. As a pragmatic feature that is self-oriented (as opposed to *you know*) and used to highlight adjustments made in the utterance, television dialogue writers might use this device consciously to highlight insecure character moments. In using the marker, the character draws attention to a part of the utterance that needs to be re-phrased or clarified not only to the other character, but to the audience.

In the following, I present d-values across the social categories of age, gender, and nationality, before turning to individual characters.

Figure (44) shows the distribution of the three pragmatic markers for all speakers across the six series, divided by gender. In terms of differences in use by male and female characters, a slight difference is observable. Pragmatic marker *like* in particular shows several outliers for the series *Buffy* as well as *Parenthood* which is even overtaking *you know*'s mean d-value.

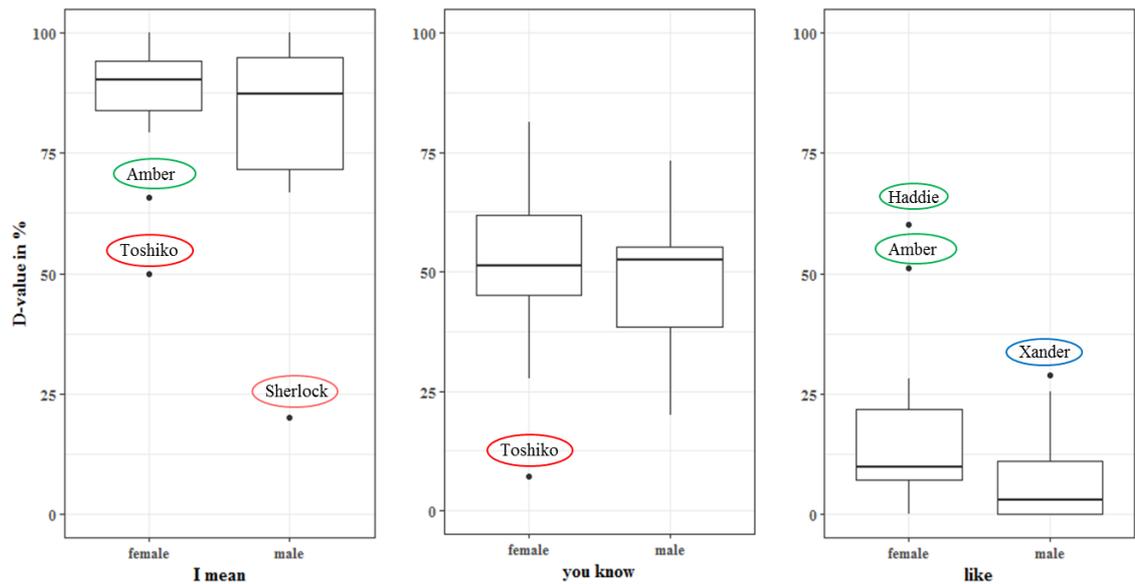


Figure 44: D-Value pragmatic markers across gender

Both *I mean*, and *you know*, have negative outliers, in that the characters here use the feature markedly less in their pragmatic functions when compared to the overall corpus. Sherlock's low use of pragmatic *I mean* might be another cue for his pragmatic competency (although, surprisingly, this is for once not paralleled with Max's use).

In terms of linguistic change or preference, *I mean* and *you know* are used at relatively stable rates (no significance in variation). In contrast, the pragmatic function of *like* is used significantly ( $p < .02$ ) more by female characters. This is also the only feature of the three pragmatic markers, that has outliers that are ahead of the use (higher than the mean), which might indicate speakers that use the feature more innovatively. To see whether this

is a pattern for linguistic change or something that implies a gender-based use, figure 45 below presents the d-value across age groups.

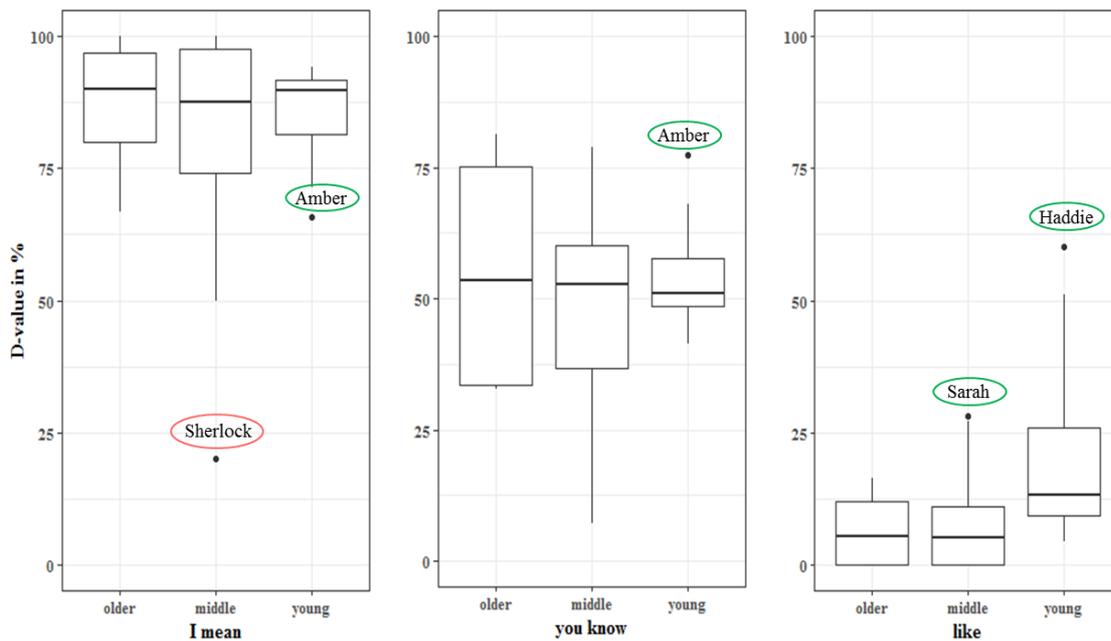


Figure 45: D-Value pragmatic markers across age

Again, pragmatic functions of *I mean* and *you know* appear to be used at stable rates (and are indeed both not statistically significant across age). *Like* shows little change between the oldest and middle age group (no significance), but increases in pragmatic use for the youngest character group (significant at  $p < .005$ ). Taken together with the previous figure, this heavily implies ongoing linguistic change, led by female characters.

Looking at the outliers specifically, Amber's use of *you know* and *I mean* is most striking. With the former, she uses the pragmatic form unusually frequently when compared to all other characters. For *I mean*, she is showing a particularly low usage for her age group. This is explored further in the character specific analysis below.

The next figure, on d-values across nationality, provides some further evidence as to how pragmatic markers are used across two main varieties of English. The previous analysis, as well as referenced previous studies (see 5.1), implied a generally more frequent use of pragmatic markers in American English.

This pattern is indeed also visible with d-value rates, albeit not at a significant level for *I mean* (significance for *you know* is at  $< .05$ , for *like* at  $< .01$ ). Interestingly here, Sherlock is, even for British rates, particularly low in pragmatic uses of *I mean*. This further cements a correlation with his pragmatic competency. Distributions for *you know* show Jack as a negative outlier in the American character group. It is unclear whether this is a conscious choice linked to social indexes, or whether it has to do with the particular setting of *Torchwood*. As the only American character in an otherwise British series (produced in the UK), it might also point to inconsistent scripting of his American identity. With his d-value much closer to the British mean, it might be that *you know* does not have American indexical meaning for the British writers and so was not employed as a characterization cue. Despite showing an overall significant distinction between American and British characters, the feature might not be realized as a salient marker for national background.

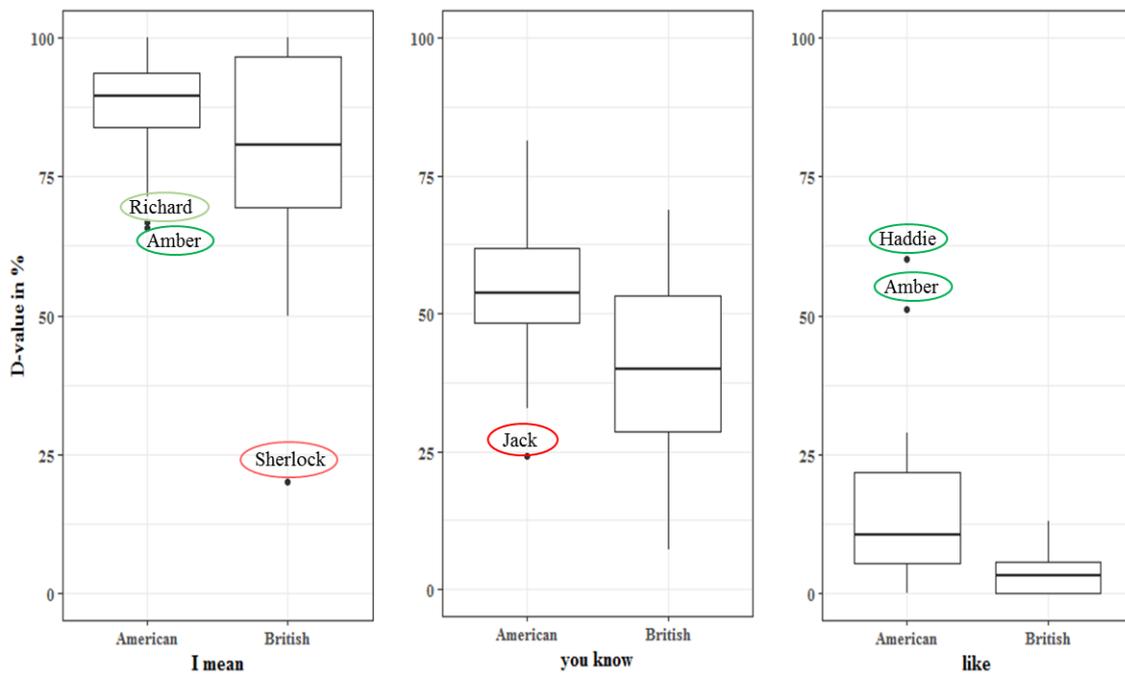


Figure 46: D-Value pragmatic markers across nationality

Finally, the low d-value for *like* and British characters might reflect the overall lesser use of pragmatic functions of these markers, or indeed reflect a lag in current change. Unfortunately the corpus does not include characters of different age groups from a British English background, so any interpretation is essentially conjecture.

The previous analysis has shown that in addition to broad social groups, the type of series might also be meaningful for variation. The following section investigates d-values across each of the characters, separated through their respective series.

The d-values in *Parenthood*, as seen in Figure 47, indicate that the overall pattern is similar in that *I mean* is used most often in pragmatic marker functions, followed by *you know*, and lastly *like*. The only exceptions to these patterns are by Haddie (who exhibits an unusually low pragmatic marker use of *you know* and an increased use of *like*), and Amber (who uses *I mean* in fewer relative pragmatic functions than *you know*). Both

characters show a heightened use of pragmatic marker *like* –an index for their group membership of stereotypical *Valley Girls* and language innovators. Further, their vice versa use of markers *you know* and *I mean* might support an individualized characterization. While both characters fall into the stereotype of *Valley Girls* with their use of *like*, their respective use of the other two features actually pinpoints where they differ. Amber, who is much more outgoing and outspoken, uses *you know*, the outwards oriented feature to a higher degree while Haddie, who is much more introverted and not as self-secure, uses the pragmatic marker of self-repair and hesitation, *I mean*, in increased patterns.

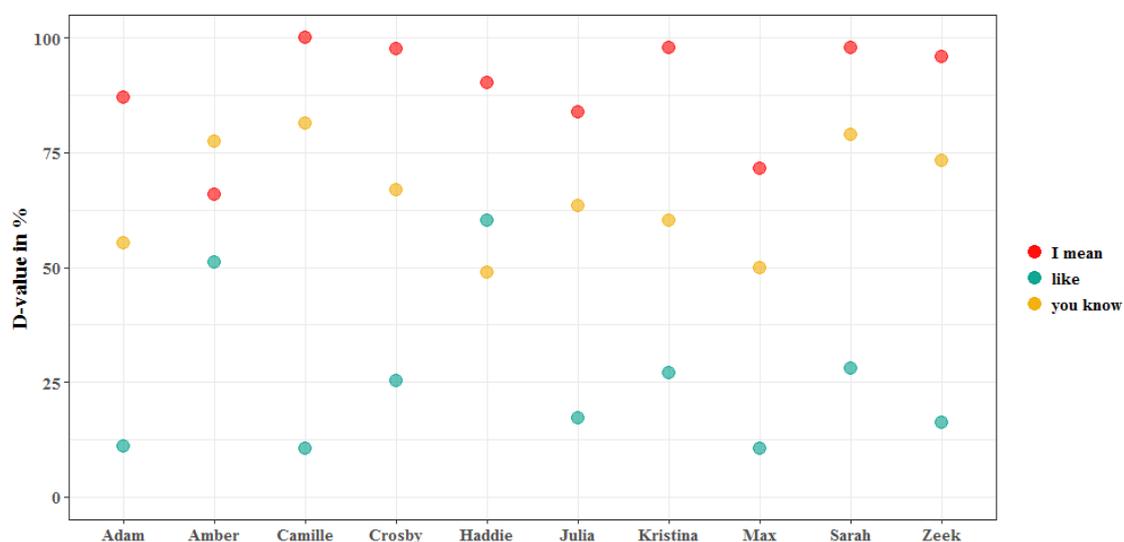


Figure 47: D-Value across characters: *Parenthood*

The lowest use of overall pragmatic marker functions is shown by Max which might be a supporting characterization cue for the ASD diagnosis that is a strong theme throughout the series itself, and his character exposition in particular. The pragmatic markers that are used to help speakers organize speech events, turn-taking, and stances such as hedging or mitigation will indicate a lessened pragmatic competence if limited or missing altogether.

The relatively low d-value is a further indication that Max, albeit using the features, is not as able to effectively communicate as the other characters within the series.

An important point to note here is that taking the whole corpus into consideration, his use of *you know*, *I mean*, and *like*, is not conspicuous at all. Indeed, the overall high use of pragmatic markers in *Parenthood* (as discussed above), is what highlights his relative low use. This shows, again, that it is important to take production means and genres into consideration and avoid conflating television series dialogue into one speech genre.

In comparison to the d-value distributions in *Parenthood*, *Gilmore Girls* seems more consistently patterned across characters. This might be because dialogue in *Parenthood* is mostly multi-authored with actors bringing in their own lines while the dialogue in *Gilmore Girls* is almost exclusively penned by the head-writer. Despite this, there are still differences depending on the characters that can be investigated.

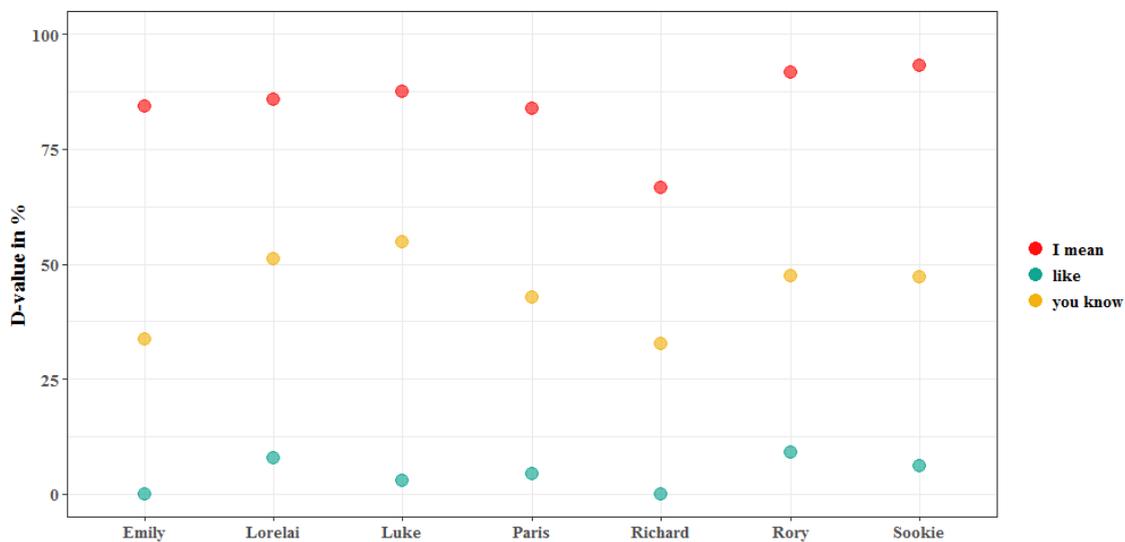


Figure 48: D-Value across characters: *Gilmore Girls*

The lowest general uses of pragmatic functions of the markers is exhibited by the grandparents, Emily and Richard Gilmore. This is unsurprising, as not only are they most different from other characters in the series (Lorelai and Rory, the main characters, in particular), but a main theme of the series which might be supported here is their inability to have ‘normal’ communication with anyone other than with each other. As explored in the previous analysis, their status in the ensemble is much more formal and upper class than that of the other characters. It is unsurprising then, to see pragmatic marker patterns that are more similar to formal speech genres than informal speech. Supporting this claim is the fact that the character closest to the grandparents’ patterns is Paris. She shares the grandparents’ socio-economic status, but is also characterized as particularly blunt.

Outside of the patterns related to class, d-value distributions in *Gilmore Girls* do not seem to indicate character categories or individualization.

Figures 49 and 50 show the d-values for the British series *Torchwood* and *Sherlock* in direct comparison.

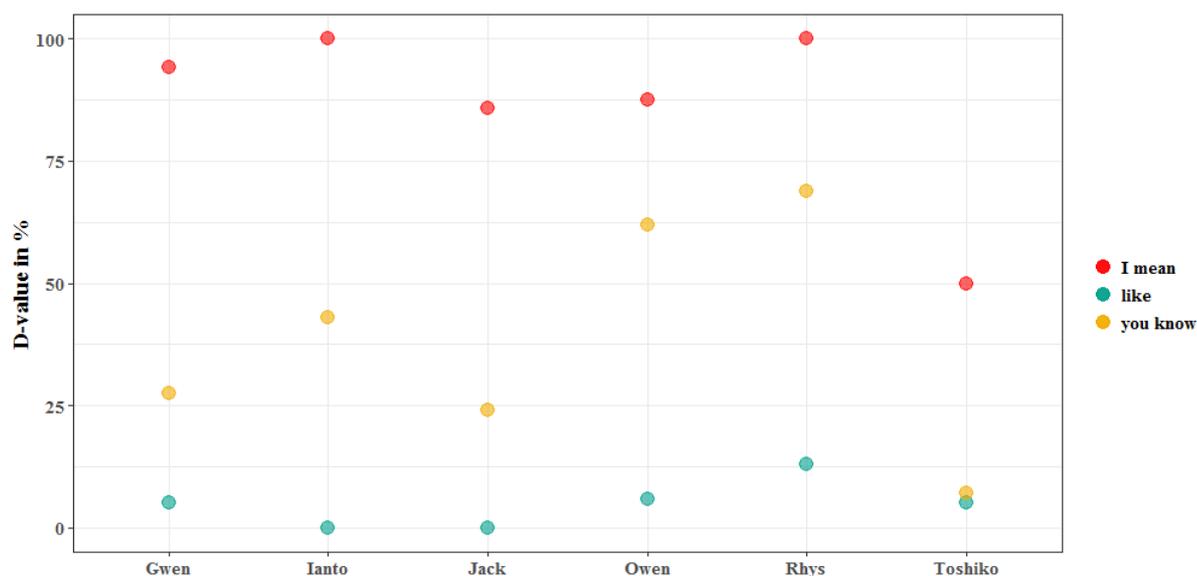


Figure 49: D-Value across characters: *Torchwood*

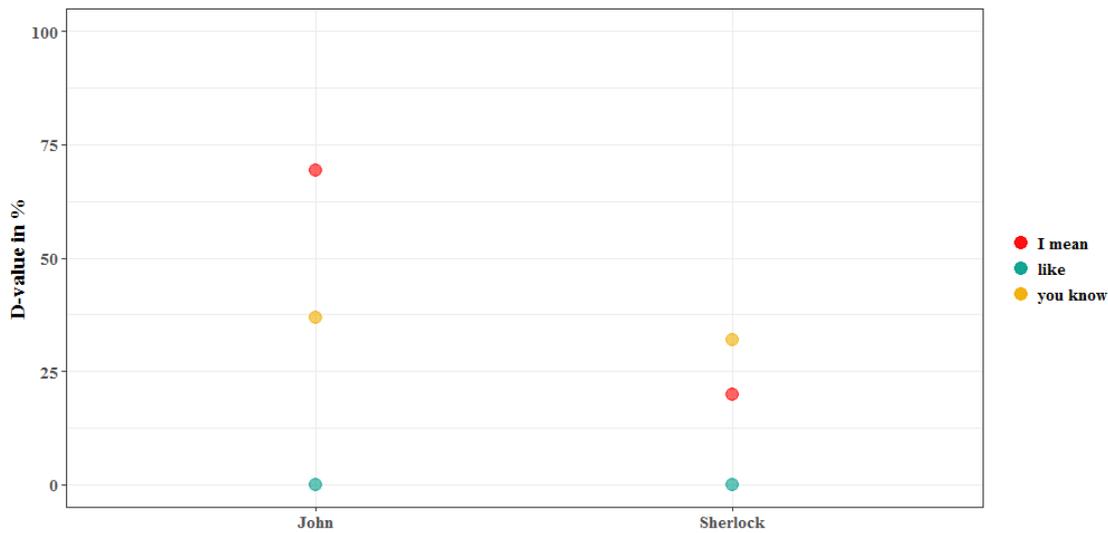


Figure 50: D-Value across characters: Sherlock

Both series show much lower d-values in comparison to their American counterparts, as discussed above. The overall pattern of *I mean* leading before *you know* and *like* in pragmatic functions can also be found here.

A notable distribution of pragmatic marker d-values can be seen with Sherlock. His distribution of pragmatic functions across all three markers is the lowest of the Television Dialogue Corpus and, similarly to Max from *Parenthood* and, in some ways, Emily and Richard from *Gilmore Girls*, a certain lack of pragmatic competence could be a cause here. While never officially confirmed, many particularities of his characterization point to social ineptness (with discussions commonly ranging from autism spectrum disorders to anxious personality disorders) which the previous analysis has shown correlated with a lessened use of discourse managing devices<sup>59</sup>.

---

<sup>59</sup> Bednarek investigated language patterns of *The Big Bang Theory*'s Sheldon, a character that is very similarly characterized as both Sherlock and Max. While she focused on other features, it is interesting to see how language is used to portray social otherness across different genres. While Sheldon's character uniqueness is predominantly used for humour and release, Sherlock (as part of the crime genre) is supposed to awe the audience with his exceptional talents and Max (within the genre of dramedy) is part of a theme seeking empathy from the audience. Linguistically, these different aims of characterization

Finally, the d-value distribution for the series *Angel* and *Buffy the Vampire Slayer* are displayed below.

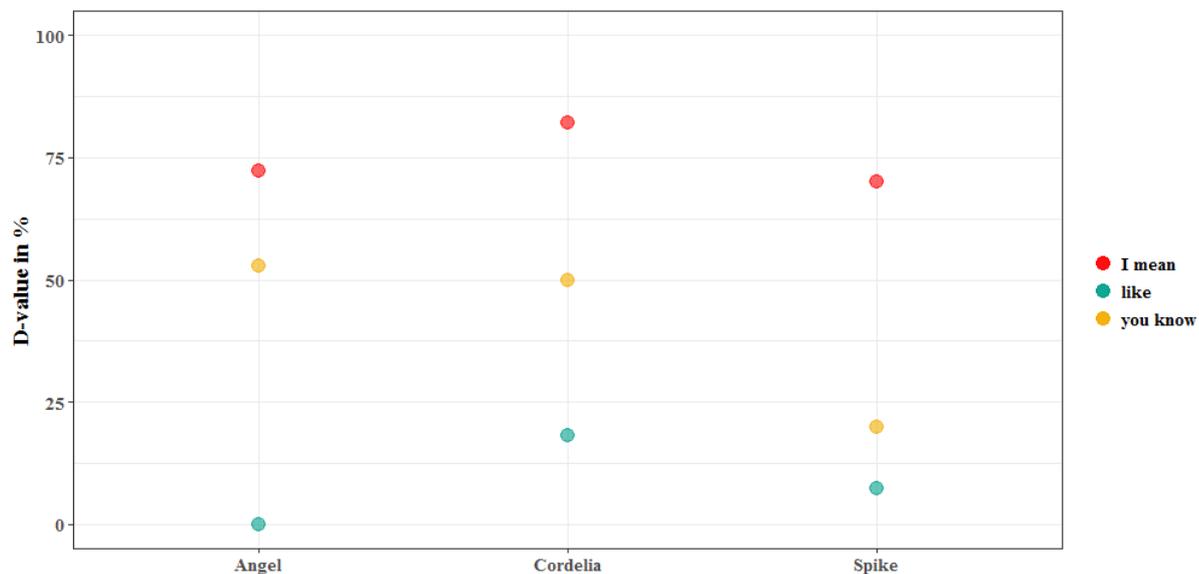


Figure 51: D-Value across characters: *Angel*

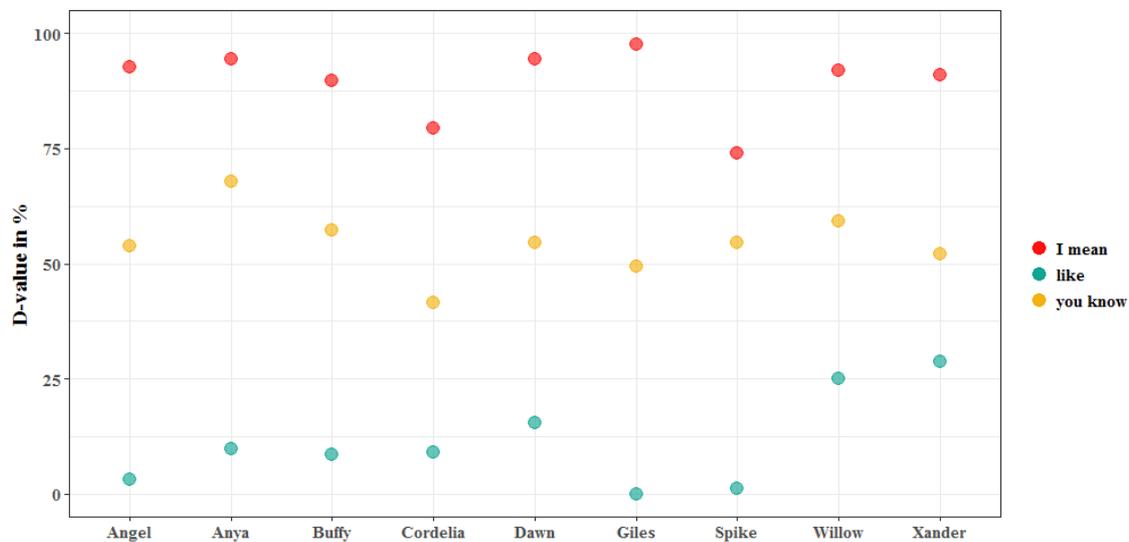


Figure 52: D-Value across characters: *Buffy the Vampire Slayer*

---

might differentiate how they are portrayed, while at the same time, common linguistic representations in terms of pragmatic competence might appear.

Similar to the other series discussed above, we find the pattern of *I mean* with the highest d-value, followed by *you know*, and *like*. Giles and Spike (followed by Angel in *Buffy* and *Angel*) in *Buffy* stand out for their low (or non-existent) use of pragmatic marker *like*. This highlights, again, the categorization of British English as being different from American English (and thus showing that *Torchwood*'s characterization of Jack is series-dependent and possibly linked to the writers' perception of the features). Interestingly, we can see that Xander uses *like* in its pragmatic function relative to propositional *like* the most. In that he defies the stereotype of the *Valley Girl*, or indeed, outs himself as one. With flipped character schemes a central theme of the series, it might be that we can see an instance here where linguistic patterns are consciously exaggerated one way (against expectations) to emphasize that stereotypes are in fact turned upside down. Dawn and Cordelia, arguably part of the *Valley Girl* category, also exhibit high uses of pragmatic marker *like* and, together with Xander, are the innovators of the series. It is unclear whether linguistic patterning is consciously employed for all characters at equal measures at all times or if certain themes are picked up and changed where and when appropriate. It might be that some characters show a higher degree of linguistic characterization with reference to series/season plotlines than others (Xander in terms of stereotype defiance). Nevertheless, even if we must assume low conscious input for a number of characters, linguistic patterns do betray some character categorization.

The overall pattern of pragmatic markers' d-values in the Television Dialogue Corpus is similar for all series in that *I mean* has the highest d-value, followed by *you know* and then *like*.

The next section repeats the analysis of d-values with hedges *sort of* and *kind of*.

### 6.1.2 *D-Value of hedges*

Similarly to the analysis of pragmatic markers above, hedges are multifunctional in that they can appear with propositional and pragmatic functions. As previously mentioned (see section 5.2), propositional functions of *sort of* and *kind of* appear in noun phrases only and serve typification or categorization rather than hedging uses. As the analysis of d-values for pragmatic markers has shown, the d-value can give indications as to how established certain features are, as well as if genres of scriptedness allow for similar distributions as naturally occurring speech genres.

With both hedge forms (*sort of* and *kind of*) claimed at being functionally identical, their respective d-value can be directly compared to one another. Table 23 shows overall counts of tokens found in the Television Dialogue Corpus, specifying those that are propositional (used as typification) and those that are pragmatic (used as hedges), including also the d-value and the total occurrence.

Table 23: *Sort of / kind of* distribution in TV corpus

<b>Function</b>	<b><i>sort of</i> (incl. <i>sorta</i>)</b>	<b><i>kind of</i> (incl. <i>kinda</i>)</b>
Hedge	359	1091
typification	111	476
d-value	76%	70%
<b>Total</b>	<b>470</b>	<b>1567</b>

Beeching (2016) included only *sort of* in her analysis of pragmatic markers in British English, which makes comparisons across the two variants difficult. However, her summary of d-values from different genres enables me to position fictional television dialogue within the spectrum of discourse patterns (see Table 24).

Table 24: D-value (in %) by genre, adapted from Beeching (2016:161)

<b>d-value by genre</b>	<b>sort of</b>
Leisure	74%
UWE Role-Play	86%
Spoken demographic	90%
Education	72%
Business	70%
Institutional	60%
<b>Television Dialogue Corpus</b>	<b>76%</b>

As previously mentioned, the written-to-be spoken language in television series would be closest related to the spoken BNC subcorpus of Leisure, here showing a d-value of 74%. With 76% in the Television Dialogue Corpus, the two corpora are surprisingly similar.

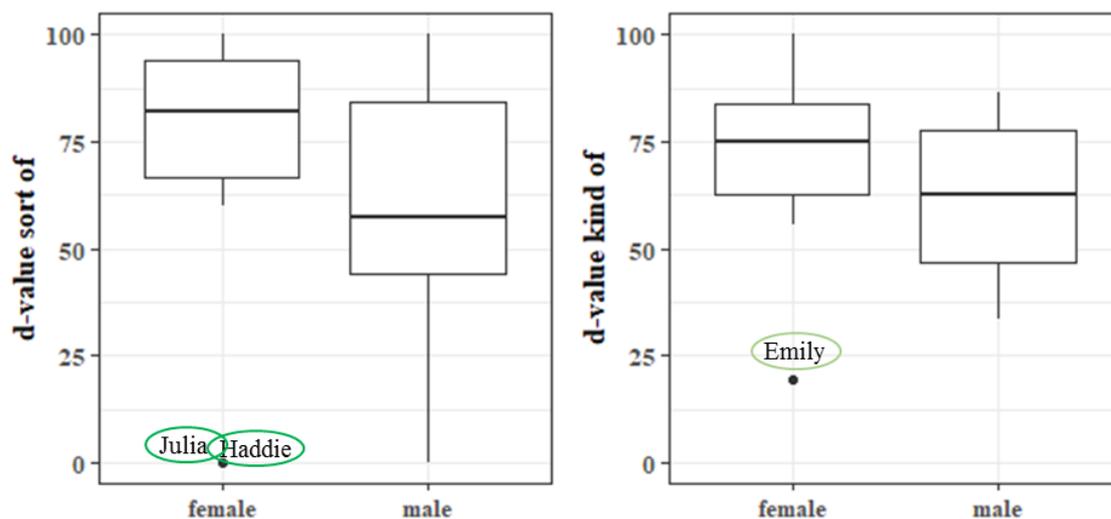


Figure 53: D-Value hedges across gender

Figure 53 above shows the d-value distribution across gender for all series. Both variants show a similar pattern whereby female characters use the pragmatic hedges at higher rates than male characters. This is statistically significant for *sort of* at  $< .05$ . The previous analysis has already shown that *sort of* is preferred in British English which directly

relates to the negative outliers in these Figures. The d-value of 0 here of Julia and Haddie (and other *Parenthood* characters in the following Figures) is no indication of the actual discourse value of the feature, so much as it is an indication of the low use of the form overall.

Figure 54 shows the distribution across age and similarly to pragmatic marker *like*, we can see a link between higher rates by female characters and an increase of use by younger characters (i.e. indication of language change).

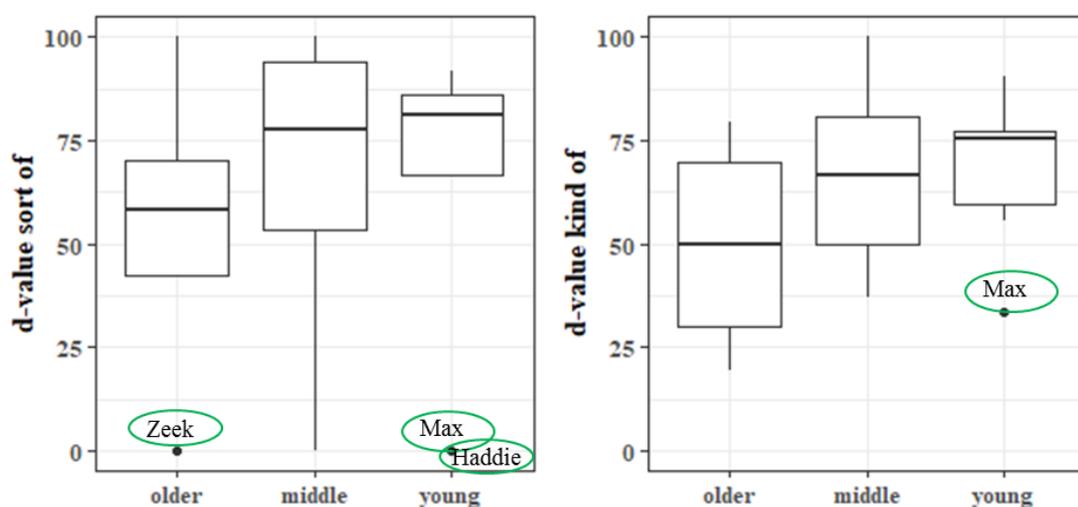


Figure 54: D-Value hedges across age

In contrast with pragmatic marker *like* however, the stratification across age does not just affect the younger characters against all others, but can be seen as a steady decrease with increasing age. This might point to a change in progress that has been happening for a longer time, or, across more generations. While not statistically significant for the d-values, the trend is visible in the figure above (as well as in normalised per 10,000 words distributions of *kind of* as seen in section 5.2).

Max's low d-value for *kind of* is comparable to that of Richard from *Gilmore Girls* (not marked as an outlier and part of the older character group). It is not clear whether Max

lags behind the pragmaticalization process (i.e. the increase of pragmatic uses of *kind of*) or whether he merely uses propositional *kind of* at a higher frequency than what most other characters do. This is explored in a little more detail in the character-specific analysis below.

In terms of national distinctiveness the patterns of the variants' d-value are slightly different from the normalized frequency analysis in chapter 5.2.

In the frequency per 10,000 words analysis I found that each main variety of English had a clear preferent of variant (*kind of* was used most frequently by American characters, *sort of* was used most frequently by British characters). The clear preference of *sort of* by British characters might imply then that *sort of*'s discourse value would also be highest for British characters. Figure 55 below however shows that the pragmatic function is more established or used by American characters, for both variants. More intriguing, despite the fact that *sort of* is seemingly more established as a pragmatic marker, American characters prefer to use *kind of* (see table 6).

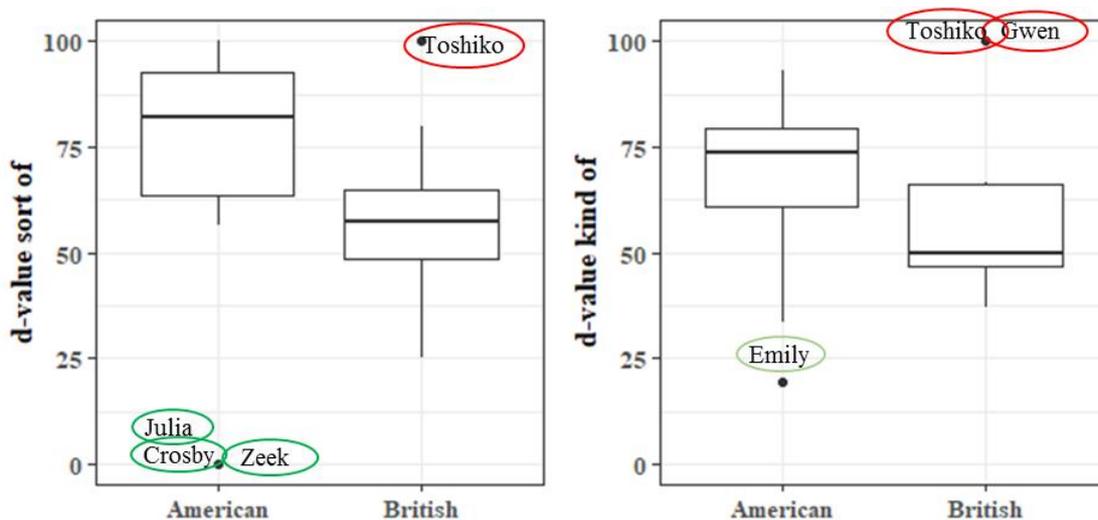


Figure 55: D-Value hedges across nationality

A possible interpretation is that while *sort of* is already more established, *kind of* is used increasingly (corroborated by the significant preference by the youngest speaker group) and is catching up in terms of pragmatic functions. Another interpretation is that there is a distinctive use of the propositional function of *kind of* which makes it the more likely variant which ultimately pushes down the d-value for *kind of* in comparison to *sort of*. Unfortunately, the actual uses of the forms in the Television Dialogue Corpus are too low to conduct a more detailed analysis. Notwithstanding, the d-value distribution across nationality confirmed that hedges are used in distinct ways by British and American characters. The following analysis on individual characters explores these distinctions further.

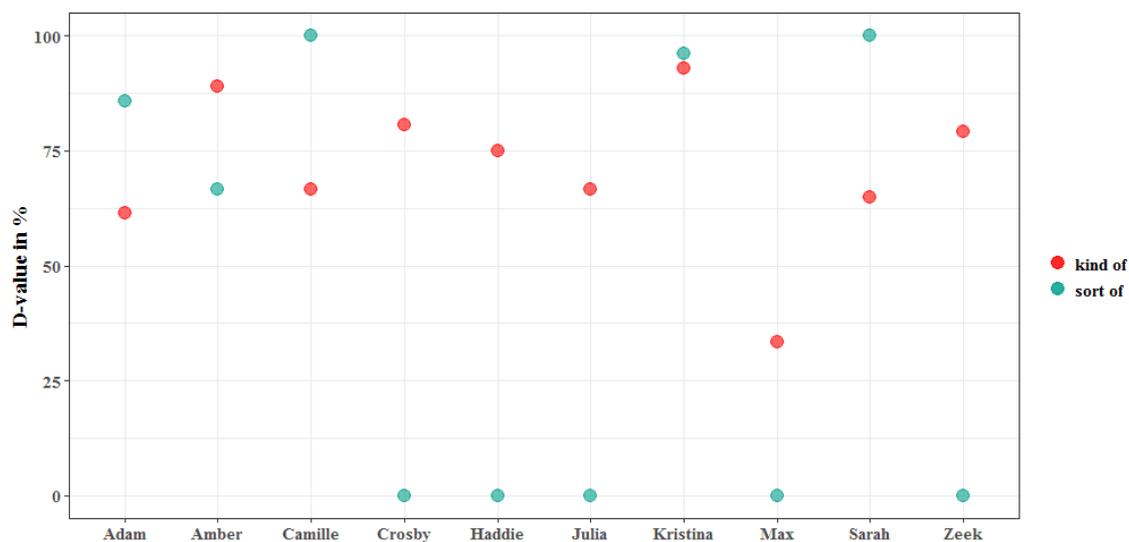


Figure 56: D-Values of hedges across characters, *Parenthood*

The d-value in *Parenthood* does not appear to follow a clear pattern. The character that stands out with his use of both hedge variants is Max. Similarly to the previous discussion of the representation of a lack in pragmatic competence in possible combination with his ASD diagnosis, we might see a diminished use of pragmatic function for the same effect.

Additionally, it might be the case that the overwhelming uses of *kind of* by Max are propositional, not because the pragmatic function is something he does not (know how to) use, but because he uses disproportionately many propositional *kind ofs*. The semantic context of propositional *kind of* is linked to typification (as defined in chapter 5.2) and defining contexts. In that, the propositional function of *kind of* is in some ways the exact opposite of hedge *kind of*: while one makes the utterance vaguer, the other provides a distinct definition in form of specification. Max, who is struggling with interpersonal or indirect communication and is characterized as being more comfortable with clear language, would be likely to also use higher rates of concrete and/or formal language. While the overall counts of *kind of* for Max are too low to confidently identify this as a reason behind these patterns, other characters with similar pragmatic competency backgrounds (i.e. Sherlock, Paris) might support this interpretation.

Hedging functions in *Gilmore Girls* show patterns of the variants across all speakers with a consistently higher d-value for *sort of* over *kind of*, further supporting the claim that *sort of* might be more established as a hedging discourse feature than *kind of*.

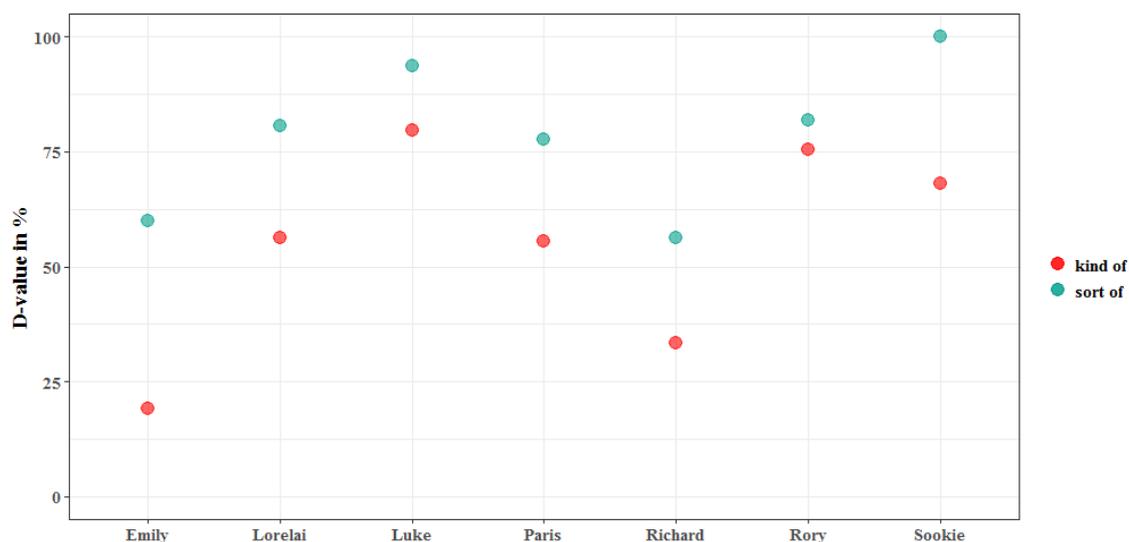


Figure 57: D-Values of hedges across characters: *Gilmore Girls*

Characters with particularly low d-values for both variants are, again, the grandparents Emily and Richard. The fact that they don't seem to use these features as much as other characters in hedging functions (in relation to propositional functions) can be attributed to their status of being more formal and possibly upper class. This last point could be further supported by Paris, also upper class, being next in line with a lower d-value for both *kind of* and *sort of*. This also ties in with the above-made point on Max, although it is worth pointing out here that Paris' d-value distributions are not markedly different from that of other characters.

For *Buffy* and *Angel* patterns seen in the other series are also visible in that *sort of* has higher d-values throughout. Exceptional here is the hedge use of *Buffy's* Cordelia and Dawn. Both use *kind of* with higher d-values. This also reflects the use of hedges by Amber from *Parenthood*, thus possibly an indicator that *kind of*, in hedge functions, is indexing particular young, female characters. Interestingly, the pattern for Cordelia is not

maintained as she goes from *Buffy* to *Angel*, where her characters shifts from a stereotypical popular high school girl to core character.

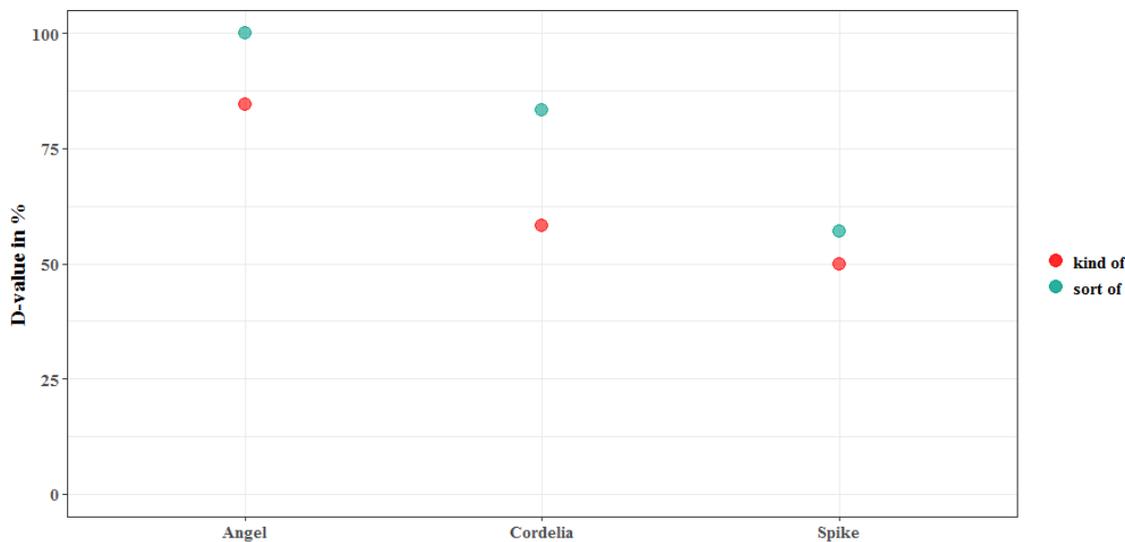


Figure 58: D-Values of hedges across characters: *Angel*

Giles and Spike (*Buffy* and to a lesser degree *Angel*) both show a relative low d-value for both variants. Similarly to other characters from the Television Dialogue Corpus (Max from *Parenthood*, Emily and Richard from *Gilmore Girls*), this might point towards the creation of outgroups either through a lack of pragmatic competence (and thus a decreased use of discourse managing devices highlighting social disorders) or increased formality for class distinctions. Giles and Spike are outgroups in the *Buffy/Angel* universe because they are both characterized through distinctive English stereotypes. Their particular use of hedges (which might be a relative low use of the pragmatic variant OR a relative increase use of the typification) might thus index Britishness. As we have seen in the previous analysis, formality and upper-class style in American characters corresponds with British English patterns, and this is seemingly the case for pragmatic functions of hedges as well.

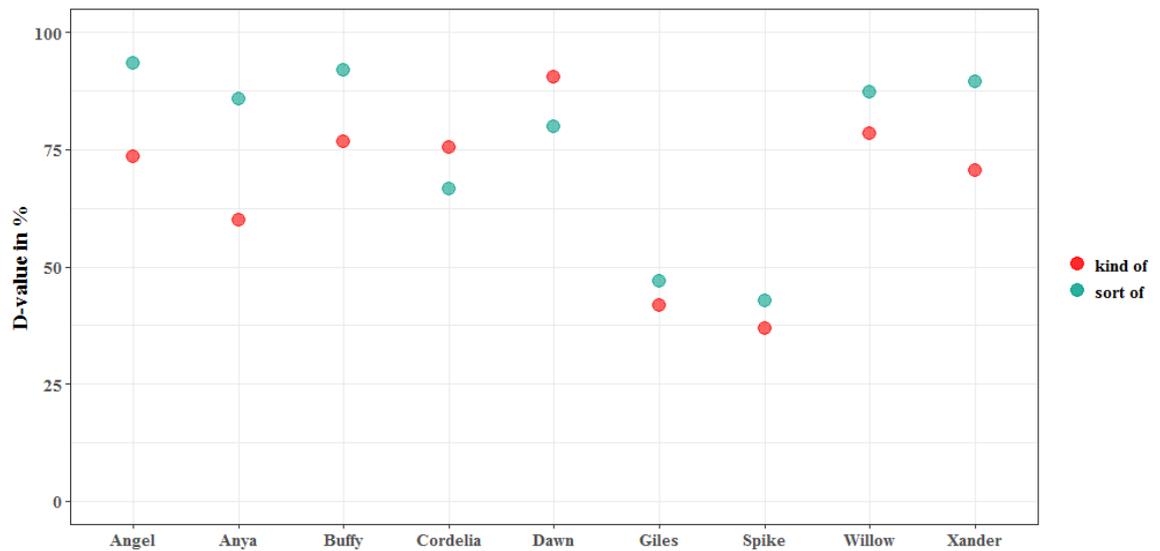


Figure 59: D-Values of hedges across characters: *Buffy the Vampire Slayer*

Finally, the British series show a varied d-value distribution across both series and all characters. The American/British distinction that was apparent in *Buffy* is, again, not visible with *Torchwood's* Jack. For *Torchwood* overall, there do not seem to be any patterns that were discernible in other series (young female with higher *kind of* d-value or formality indexed through low general d-values). A possibility here might be that the sub-corpus of *Torchwood* is, especially compared to the other sub-corpora, relatively small and does not offer enough tokens for patterning.

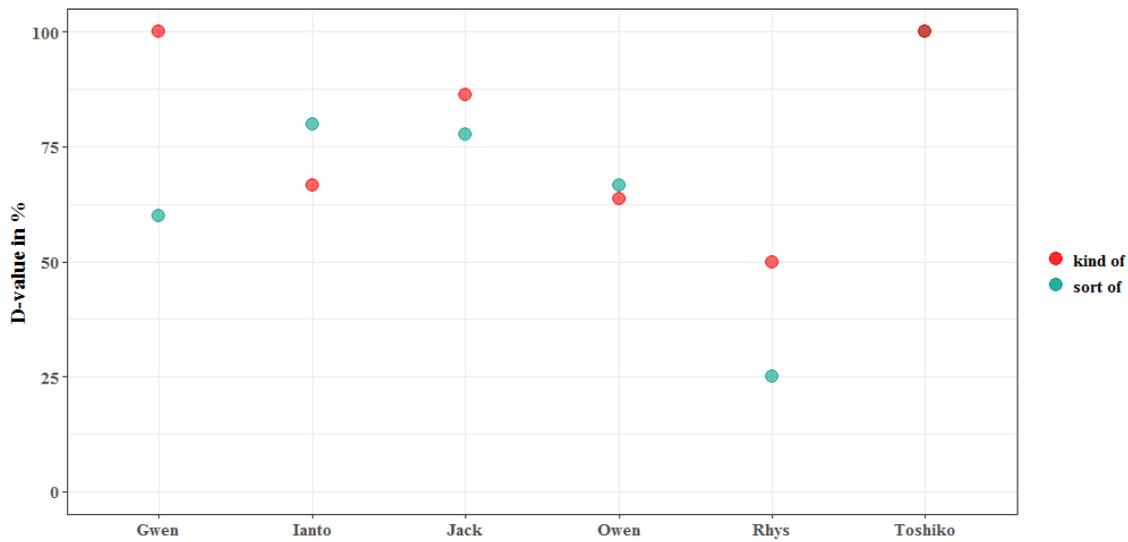


Figure 60: D-Values of hedges across characters: *Torchwood*

Sherlock and John's (*Sherlock*) relative use of pragmatic vs. propositional use of the variants are similar to the values we find with Giles and Spike (*Buffy, Angel*), and Richard (*Gilmore Girls*), again, supporting the claim of British patterns that map onto formal, upper class American characters.

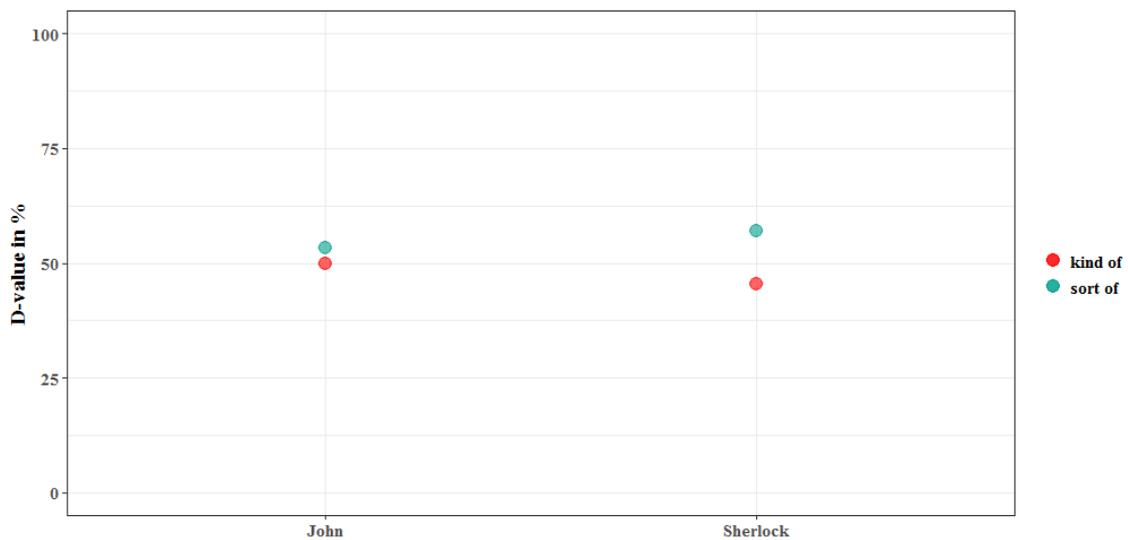


Figure 61: D-Value of hedges across characters: *Sherlock*

### **6.1.3 *D-Values as indicators of change***

In summary, the analysis of discourse values has added further evidence, if not completely new findings, to previously discussed characterization patterns. In terms of linguistic change, the d-value distribution is a promising methodological addition to tracking innovations and possible indexicality of various features. The previous frequency analysis already indicated that some of the variables under investigation were currently undergoing change (pragmatic marker *like*, hedge *kind of*, a general shift in intensifier preference, etc.). The analysis of functional shifts of either increasing or decreasing d-values across character style or type adds a second dimension to our knowledge of change. Pragmatic marker *like* for instance is shown to expand considerably in pragmatic function in the youngest character group, particularly for the female characters. This clear stratification not only points towards language innovation, but also towards marked indexes and saliency of linguistic forms. This is further explored in the discussion, summarizing findings from both analysis chapters (for discussion, see chapter 7).

## **6.2 Changing characters, changing patterns**

This section very briefly explores the shift of characters within fictional stories and whether their linguistic patterns shift accordingly. Additionally, I analyse character types and their language patterns with reference to production difference. Findings from the previous analyses have shown that depending on how a series is produced, overall frequencies of linguistic features might shift also. Here, I present a brief account of that notion, highlighting that meta-information of television series production values need to be taken into consideration for any interpretation of language use.

With long-running television series, linguists have the opportunity to study language use in real-time studies with possible foci on systematic language change, as well as concepts of life-span changes, etc. (see chapter 2 for a brief introduction to sociolinguistic theory on age as a social variable).

The longest running series in the Television Dialogue Corpus are *Gilmore Girls* and *Buffy the Vampire Slayer* with seven seasons each. In terms of linguistic change, that is a relatively short amount of time, as most language features evolve over several generations (as was shown with various apparent-time analyses for the individual features). The following analyses illustrate differences in real time (to track changes of how features evolve) as well as differences across two differently produced series (to track changes across genres). My aim is to show how linguistic patterns shift alongside characters or character types. For that reason, I have chosen to look at characters that undergo considerable change from the first appearance to their last

### **6.2.1 Character change: Cordelia, from *Buffy* to *Angel***

As introduced in chapter 3, Cordelia is a secondary character on the first three seasons of *Buffy* before becoming part of the main ensemble in the spin-off series *Angel*. When first introduced she encompasses the then current stereotype of popular high school cheerleader. Coming from an upper-class background and surrounded by many friends in season one of *Buffy*, Cordelia continuously undergoes slight shifts in character. Initially introduced as Buffy's antagonist, she ultimately joins the tight-knit friendship group and helps fighting vampires alongside Buffy, Willow, and Xander.

We have explored this shift with reference to intensifier usage in Reichelt and Durham (2017) and found that her patterns of intensifier preference, over the course of three

seasons, converge with those of Willow, exemplifying their establishing trust and friendship.

Here, I want to add to those findings by investigating how Cordelia of *Buffy* differs from Cordelia of *Angel*. The latter has Cordelia become a character that is altogether more rounded in terms of participating story lines (in *Buffy*, many of her contributions were based on antagonistic commentary or comedic relief).

The Figure below (62) shows some of the features from the main analysis and how they differentiate between the two versions of the character (values represent normalizations of per 10,000 words).

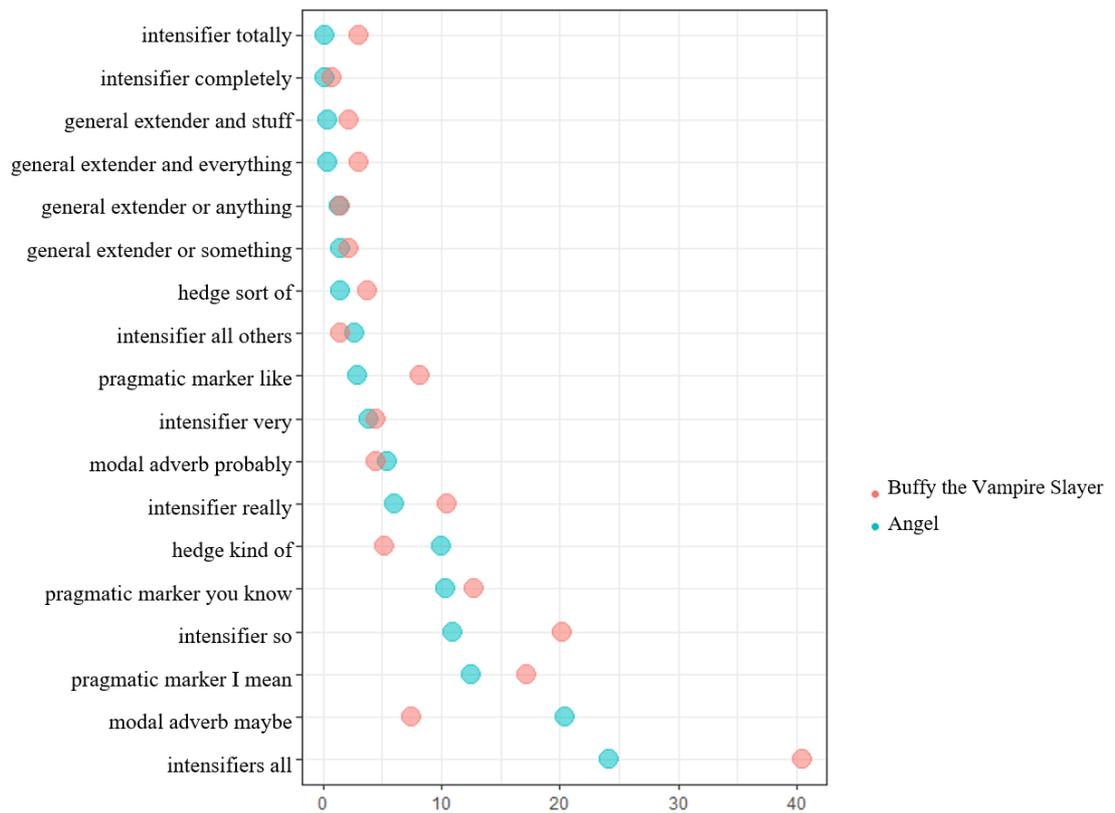


Figure 62: Changing patterns in Cordelia's linguistic profile

There are few features that overlap for both series, giving the impression that Cordelia considerably changed her linguistic profile from *Buffy* to *Angel*. It is worth noting that the features that do overlap, are not particularly indexical for her character type, which might mean that they were not used as characterization cues in either of the series.

Going to the features that the previous analysis has found to be indexical for young, female, American speakers (in particular those features that index the *Valley Girl* persona), we can see that Cordelia shifted in between the two series.

Intensifiers *totally*, *really*, and *so* (as well as the overall count of intensification) decreased from *Buffy* to *Angel*. This goes alongside her transitional shift from a character based on stereotypical language patterns to a more layered character that is shown to grow up and move away from her former self. A similar pattern is also seen with pragmatic markers *like*, and to a smaller degree *you know*, and *I mean*. Notably however, hedges *sort of* and *kind of* are shifting in opposite ways. While *Buffy*'s Cordelia uses more *sort of* than Cordelia in *Angel*, her use of *kind of* actually increases. In *Buffy*, the hedge use was not markedly variable between the two variants (comparing the two red markers for *sort of* and *kind of* in the above figure). In *Angel* however, the distinction between the two variants is clearly marked, with *kind of* being favoured. This further adds to the above discussion on how *kind of* is currently changing in terms of saliency and it might be that within the years of filming the two series, it has gained indexical meaning. This would explain *kind of*'s use as a characterization cue in *Angel*, but not in *Buffy*. Again, further insight into the development of this feature is needed to fully capture its use as an identifying linguistic marker.

This brief exploration has shown that a character shift is indeed visible in changing linguistic patterns, further contributing to the importance of language variation for the identification and individualization functions of dialogue.

The final analysis investigates the different linguistic patterns for similar character types to see whether a change in series and production method might have an effect on how a character is presented.

### **6.2.2 Character types: from Lorelai to Sarah**

The data used here is of two characters in a similar role (30-something struggling single mom), portrayed by the same actress (Lauren Graham) from the first season of *Gilmore Girls* (2000) to the fourth season of *Parenthood* (2013). In having the same actress portray both roles, possible interference from the actor can be minimized and any differences can be directly linked to the medium.

As mentioned in the series descriptions (chapter 6) both series fall into the genre of dramedy (a hybrid form of drama and comedy) and focus largely on family dynamics. The characters of Lorelai and Sarah both start their respective character arcs by returning to their parents' doorstep (Lorelai in asking for tuition for her daughter's school, Sarah in asking to stay with her family after moving back into the area). Both characters find success in their professional fields (Lorelai as the owner of an inn, Sarah as a photographer) and storylines are mostly based on their relationships with their children as well as various romantic relationships with male characters. While their characters differ in some ways, the overall character type is comparable and would suggest that linguistic patterns are equally overlapping.

Where the series differ however, as mentioned at various points, is the way it is directed. *Gilmore Girls* is a series that is completely scripted, while *Parenthood* is much more lenient towards ad-libbed performances and impromptu changes to the scripts. It is this difference in production that results in *Gilmore Girls* being much closer to typical scripted language (with fewer vague moments, false starts, hesitations, etc.) and *Parenthood* much closer to naturally occurring language.

The features I include here are the discourse values of pragmatic markers *you know*, *I mean*, and *like*. By splitting up the dialogue across the individual seasons, I wanted to make sure to have meaningful counts. Further, as the analysis above has shown, the pragmaticalization of pragmatic markers differs from variant to variant. *You know* and *I mean* have both appeared to be stable in their development as pragmatic devices (see beginning of this chapter). *Like*, on the other hand, is increasing, particularly with female characters.

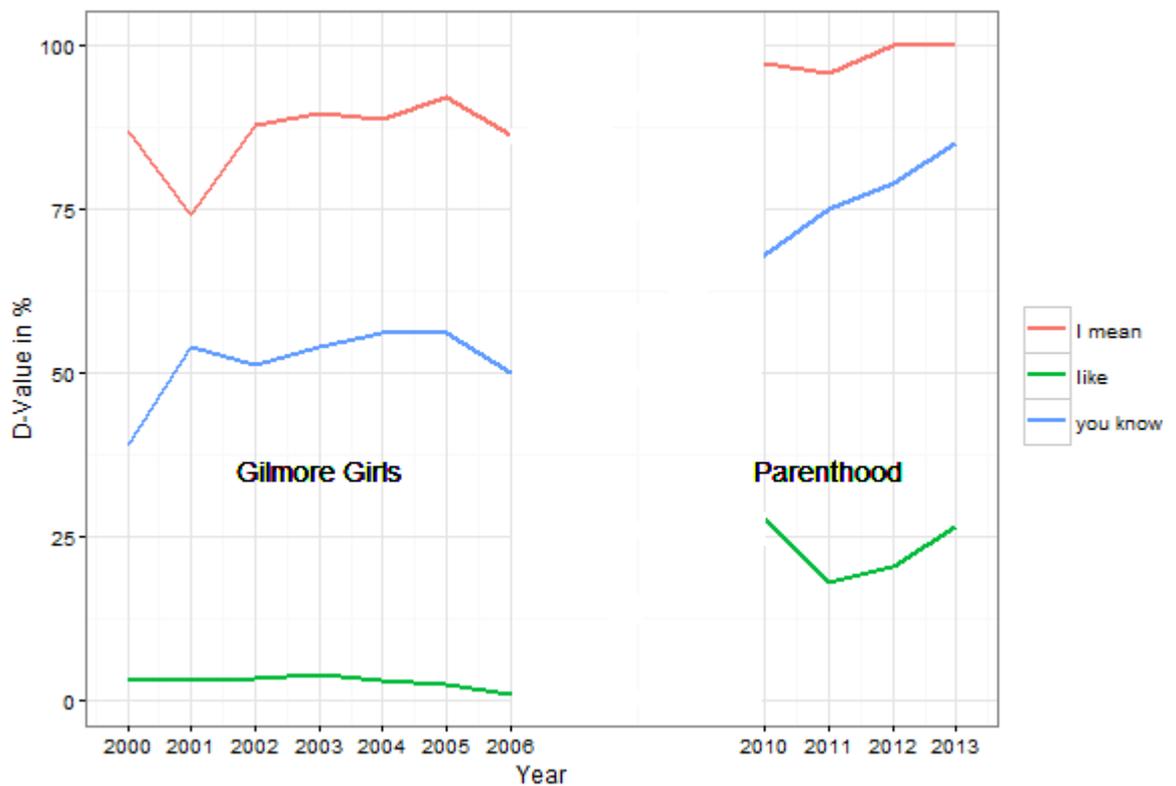


Figure 63: D-value Lorelai Gilmore & Sarah Braverman across the series' running time (in %)

All three markers show an increase over time, possibly reflecting the increasing establishment of the markers in language use in general. Alternatively, the increase might account for the presumed increase in shared knowledge between characters and audiences. As mentioned previously, markers that imply shared knowledge are used to establish close interpersonal links between speakers. In television dialogue contexts, the shared knowledge has to be accounted for not only between the social interactants on screen, but also within the parasocial interaction off screen. With continuous story lines, dramedies in particular enjoy audiences that follow up on the series from start to finish. That means that the more episodes a series has, the more interpersonal relationship the characters have with the audience: shared knowledge, at that point, can be assumed and used as a cue to make the audience feel as part of the group: forming almost a type of

community of practice. I suggest that an increase in vagueness and features that imply shared knowledge is a reoccurring pattern in many long-running drama series.

Both series show a remarkable difference from season one (2000 to 2001 for *Gilmore Girls* and 2010 to 2011 for *Parenthood*) in the features' d-values. This is a typical trend for television series, as characters are usually established with much more exaggerated tones in initial episodes to address character schemes that audiences can recognize in order to invest quickly into the storyline. Both series are more consistent from season two onwards respectively, yet show clearly that the series are produced under differing circumstances. This translates into higher d-values in *Parenthood* with the use of pragmatic markers increasing impressions of vagueness and thus being omitted from *Gilmore Girls* much more.

Across the running time of *Gilmore Girls*, all three markers seem to be used relatively consistently (bar season one for the above-mentioned reason) with a notable drop in the final season. This points towards a possible change of script writing style, as this is the season where the head-writers and creators of the series left the production. It would be worthwhile to follow up on this possible trend in terms of other features and general characterization patterns, though any further investigation lies regrettably outside the scope of the present study.

In summary, this last analysis has shown that character categorizations might suggest certain patterns in language use, but that other factors will necessarily have to be taken into consideration as well. Lorelai and Sarah, despite representing very similar personas, show that frequencies and grammatical establishment of linguistic features differ considerably. This should undoubtedly have an effect on how character types are

interpreted in terms of language use and what information needs to be provided to fully account for linguistic patterns in these scripted contexts.

## 7 Discussion

The first half of the discussion chapter aims at returning to the research questions I posed at the beginning of this thesis. Here, I highlight relevant findings from the analyses that contribute to our understanding of how linguistic variation adds to the process of creating fictional characters.

The second half of the discussion frames the current study within the wider context of sociolinguistics and shows what scripted television in particular adds to our knowledge of language.

Any interpretations drawn obviously derive from the specific findings I gained here and are in no way a complete portrayal of the linguistic portfolio fictional television series inhabit. Different foci will likely add new nuances and might even refute some of the here presented results. However, that should not undermine any interpretation that is following, but rather spur on future research into this expanding field. For that reason, I will also emphasize areas where I feel the current research could potentially be revisited in the future.

### 7.1 Summary

The starting point of this investigation was a relatively simple query which ultimately led to my overall research questions:

*Why does character A speak differently from character B?*

At the core of this question lies the idea of fictional identity and how linguistic identity in this case is shaped through commonality as well as diversity. This in itself is not yet a puzzle worth solving; after all, the fact that characters will differ in the way they speak is

not something any audience would dispute. The more persuuable aspect lies with *how* and *why* these characters differ in their speech.

In summary, on screen, every fictional character is unique and their use of language is uniquely theirs. A character's background (as exemplified in chapter 2) consists of multiple layers of information, some might overlap with other characters' backgrounds (such as broad social categories), whereas others are very specifically defined for a single character (such as Max's Asperger's syndrome).

Going back to the initial question then (*Why does character A speak differently from character B?*), my approach was to single out instances where characters show differences in language use and investigate whether their preferences are constrained by their individual background.

As a reminder, the study is grounded in sociolinguistics and investigates systematic and socially meaningful variation in language use (see chapter 2). The features that I chose to investigate for this study are all, more or less, related to the notion of stance (explored in more detail in chapter 2). Landert (2017: 489) says that on the character level, i.e. from the fictolinguistic perspective, "stance expressions provide a resource for characterization and character alignment". The way characters position themselves within the story and alongside other characters is thereby crucial for character exposition. Further, as Bucholtz (2009: 147) writes, "interactional stance taking may come to be ideologically tied to larger social categories" and "linguistic forms that have become linked to particular categories may variously exploit or set aside such associations as speakers deploy these forms for their own interactional purposes". As part of a feature group of implicit characterization cues (see discussion on Culpeper (2001) and Bednarek (2011)'s frameworks in chapter 2), linguistic forms associated with speaker stance thus seemed

promising for my aims. The design of the study, in brief, included a wide range of stance markers that were previously found to contribute to a variety of pragmatic meanings in discourse (see individual analysis chapters).

Within the analyses I explored the different features, their functions, as well as how they interacted with individual characters and their unique backgrounds.

Each of the analyses consists of a brief summary that highlights the particular findings in terms of characterization. Here are some of the common patterns I that I found throughout:

### **7.1.1 *Outliers***

I want to distinguish here between outliers in terms of frequencies of feature use (empirical outliers), and outliers in terms of fictional story (character outliers).

Empirical outliers were markedly often from *Parenthood*, providing concrete evidence to the distinctiveness of scripted and unscripted television dialogue. In consideration of that it might be worth investigating different types of directing and resulting genre classifications. In particular the direct comparison between characters Lorelai and Sarah in chapter 6 has shown that differences in processes of pragmaticalization have an effect on how results need to be interpreted.

Character outliers, again, can be split into two sub-groups. Characters that are outliers within their respective series context (e.g. Giles in *Buffy*, or Jack in *Torchwood*), as well as characters that are outliers across television dialogue as a whole (e.g. Max from *Parenthood*, or Sherlock from *Sherlock*). The first one is a type of outlier that can point towards patterns of communities of practice and how in-and outgroups are established in contexts where the audience itself needs to be established as either in or outside the main

ensemble. The second one ultimately stands out the most by diverging from common linguistic patterns (Max for instance shows unusually low frequencies of most interpersonal pragmatic markers). Here, the analysis enabled me to see that scripted television dialogue offers sociolinguists the opportunity to study in- and outgroups and how language diversity can index differences in speaker identities. Further, the choice of variation (what difference triggers what type of variation and to what degree), informs linguists (as well as audiences) of the specific meaning linguistic features carry.

### **7.1.2 *Common patterns***

In line with the above, certain patterns are repetitive across a variety of features: Max (as well as Sherlock) for instance lacks pragmatic competence and expresses that through low frequencies of pragmatic markers, hedges, general extenders, etc.

Other patterns that repeated across features are the relative similarities between Giles and Richard Gilmore (and Emily Gilmore to a degree). The repetition of variance across all features indicates particular stereotypes. Fictional television dialogue thus mirrors typification of personas and might help us understand how perception of linguistic features will heavily inform how we view people based on their language. This links into Coupland's framework of stylization (see chapter 2) and how stereotypes come to be.

### **7.1.3 *Change***

Linguistic as well as character change was explored in several ways. Apparent time analyses of the three character age groups yielded general patterns of linguistic change and reflected current findings from naturally occurring language. Most remarkable here is the increase of pragmatic marker *like*, the general shift of preference between adjective intensifiers, and the pragmatic development of hedges *kind of* and *sort of*.

As previously mentioned, tracing real-time developments in language through media is highly contested in sociolinguistic research. While caution is indeed appropriate in providing evidence of how language will be changing, television dialogue does give an impression of how far along a certain change is, and who is adopting such a change. The relative spread (or decline) of linguistic features within dialogue can be used as an indicator of the particular stage of change.

With these broad findings in mind, the next section will return to the research questions posed in the beginning of the thesis, answering each of them in detail.

## **7.2 How are linguistic variables used for the purpose of characterization?**

An emerging theme for how linguistic variables are used to style characters or character groups was the respective saliency of that linguistic form. In the feature group of pragmatic markers for instance, I was able to see that *like* was highly stratified across characters. The form is very salient in meta-linguistic commentary as a stereotype for American speech, particularly indexing young female speakers. The Television Dialogue Corpus reflects that clearly with characters like Haddie, Willow, or Amber leading in *like* use. Similar distributions are found with other features that are salient in terms of perceived new-ness and popularity with certain speaker groups, such as intensifiers *so* and *totally*. Where linguistic features carry such strong associations with stereotypical speakers, we can possibly argue that the styling is, at least in part, consciously guided by the saliency. This does not only mean that characters fitting into these perceived stereotypical speaker groups (e.g. Haddie as a *Valley Girl*) will use them extensively, but also that characters who are furthest away from these groups will almost never use these features (e.g. Giles or Sherlock). I would argue that for salient features the distribution of use is very much informed by its function as a characterization device, rather than pragmatic device. This is somewhat diminished with features that are more established in

language use and forms that are less strong in indexing particular speaker attributes. Pragmatic markers *you know* or *I mean* for instance are more established than *like* and are less likely to index one particular type of character. Here, the pragmatic device function will be stronger. Characters who use these markers in high frequencies are not stereotyped as a character type, but are associated with more general pragmatic meanings, such as vagueness, hesitation, or mitigation. The distinction between characters is less stark but ultimately more nuanced.

Linguistic characterization, as argued above, is achieved through varying frequencies of features. We can distinguish between forms that style through their stereotypical functions, and forms that style through their pragmatic function and associated meaning (categorized in table 25 below).

Table 25: Categorization of characterization cues

	<b>stereotype, high saliency</b>		<b>pragmatic function, low saliency</b>
Intensifiers	<i>so, totally</i>	<i>very, really, extremely</i>	<i>completely</i>
pragmatic markers	<i>like</i>	<i>you know, I mean</i>	
Hedges		<i>kind of sort of</i>	
modal adverbs		<i>perhaps</i>	<i>maybe, probably</i>
general extenders			<i>and stuff or something, and everything, or anything</i>
	<b>characterization cue</b>	<b>characterization cue</b>	<b>no characterization cue</b>

This distinction depends highly on the saliency of forms and naturally shifts as language use in general, as well as perceptions of language use changes over time. Thus, it might be useful to revisit characterization patterns to investigate this distinction when forms that are strong stereotypes now are more established, or what stereotypical features were used in past dialogue that have lost their saliency today.

The analysis has also provided insight into characterization that is not only governed by frequency, but by choice. Linguistic variables that share pragmatic functions across variants but differ in social meaning are used to differentiate between characters. The analysis of hedges for instance illustrated clearly that not just the frequency, but the choice is highly relevant in characterization. Emily and Richard from *Gilmore Girls*, throughout the analyses, were prime examples of distinguishing themselves from the series' other characters by showing preference for distinct variants. The choice context, in terms of characterization, often highlighted character outliers by associating them with non-preferential variants. Giles, outlier in *Buffy* in terms of age and nationality, exhibited distinct choices in almost all feature categories. Some of these lined up with Spike, who shared his outlier status in terms of nationality.

In terms of characterizing social categories, the analysis has shown that what are defining categories for some characters, are not defining for others. As discussed in chapter 2, each character carries with them a particular set of identifying information. Across the corpus, the importance of this information is balanced to support the overall stories the series are telling which means that in some series gender for instance is a much more important distinctive marker than in others. In *Parenthood* for example, we have seen that Crosby is linguistically aligned with many of the female characters and is contrasted with Adam. This supports the characterization of both characters within the series. Adam is more serious and conservative than Crosby, who often clashes with his older brother in terms of life choices and approaches to conflict. The identifying information here is not their gender, but their attitude to life in more general terms. Depending on the series, identifying characters through social categories is thus not sufficient for a full interpretation of characterization processes.

In summary, the distribution of linguistic features across characters is able to guide characterization in different ways. Distribution here combines notions of frequency and choice of features. It is important to note that the features rarely support one reading of characterization and that any interpretation should always be linked to the context of the character and the series. As we have seen in the analysis of intensifiers for instance, the frequency of intensifiers can point to general emotional engagement of a character. In contrast, Max, a usually emotionally withdrawn character uses clustered intensification as a characterizing marker for his outbursts and distress associated with Asperger's syndrome.

### **7.3 How do linguistic variables interact to create character styles?**

As mentioned above, certain characterizing patterns recurred across different linguistic features. In particular, character outliers such as Giles in *Buffy*, Emily and Richard in *Gilmore Girls*, or Max in *Parenthood*, would prefer variants that other characters in the respective series would be less likely to use. This shows that characterization is linguistically consistent across a character's dialogue and thus supports the overall construction of a character identity. As mentioned above, linguistic characterization can and should be distinguished in terms of a feature's saliency. With regards to character styles, features that align with a particular character type (such as *Valley Girl*) will often coincide with other features that carry similar social meaning (such as *like* and *totally*). Here, a clear interaction of linguistic variables can be attested. Features that do not index character types, but rather pragmatic function, are more difficult to interpret as interacting. Throughout the analyses I considered how the interaction across language features could be measured in a way that directly compares characters and their relatedness to each other. As the individual analysis chapters have shown, there is some overlap between characters from different series, often related to one or two common

pieces of background information (social class, pragmatic competence, etc.). It is at this point, however, where the multifaceted nature of pragmatic features becomes problematic from an analytical standpoint. Testing the interaction of features through appropriate statistical tools is an endeavour that should be further explored where there are resources available to account for the individual linguistic particularities. A way of finding similarities in linguistic styles across genres and series offers the opportunity to explore character types more thoroughly and to test whether certain linguistic patterns are indicative of language situations in the real world. I will further investigate this last point in the second half of the discussion.

In general, an investigation of character styles often shows the interaction of linguistic variables. For instance, the linguistic patterns found for Emily and Richard consistently mark them as different from other characters on the series: the combination of choices and frequency across most linguistic features pointed to their defining character background of upper class, older and more formal. Paris, who shares some of these patterns, also shares some of the character background (i.e. upper class). However, linguistic features used for characterization do not all combine into one reading of character style. Paris' other uses of linguistic features will define her character as a young person who is oftentimes incredibly rude (diminished pragmatic competence). *Torchwood's* Jack uses some features that define him as an outsider in terms of nationality (being American within a British ensemble), while other features will not mark this character background information.

In season three of *Buffy*, Cordelia remarks to her friends' bewilderment that she likes standardized tests.

- (66) Cordelia: Actually, I'm looking forward to it. I do well on standardized tests. ... What? I can't have layers? (*BVS*)

This goes against her established character type of popular cheerleader *Valley Girl* who does not care about academic merit. Her remark about layers is a good metaphor for character styles: not every character will be defined by one kind of (linguistic) styling.

I would argue that the notion of character style can be helpful in some cases, but might also gloss over the fact that characters can be multifaceted and diverse in their reflection of self.

#### **7.4 Are patterns congruent across characters and series?**

The analysis has shown that some patterns overlap across series or series groupings. Once again, the salient characterization patterns of the *Valley Girl* persona for instance can be traced across series. Surprisingly, linguistic patterns indexing British nationality are consistent across the whole corpus, despite the fact that not all British characters are part of a British ensemble or UK-produced series<sup>60</sup>. Of note is the alignment of the British linguistic pattern with formal American English. Richard and Emily Gilmore, with little exception, exhibit linguistic patterns that are comparable to Giles' or Sherlock's uses of language. British English (and British English inspired) distributions of language features is one of the most consistent patterns I found in the Television Dialogue Corpus. Other patterns are much less marked, due to the character relevancy of that particular social category (as discussed above in terms of gender in *Parenthood*) or characteristics being uniquely defined within a series (e.g. Sherlock's pragmatic competence). The analysis of the individual features has shown, in almost all cases, that distributions need to be interpreted with reference to the particular character's background information or the

---

<sup>60</sup> And despite the fact that James Marsters for instance plays a British character while being American himself.

series' overall theme. With that in mind, I would reject the notion of defining fictional television dialogue as a genre in and of itself. This follows Bednarek's conclusion to an analysis of keywords and trigrams in fictional television series: "each series will have its own linguistic profile, being a cultural artefact in its own right — characterised by its own characters, settings, relationships, actions, events and language" (2012b:60). In addition to the notion of likeness of patterns across series, it is important to keep in mind that external factors influencing the fictional aspects of a series play a role in its linguistic profile. A series that allows for unscripted dialogue, as seen with *Parenthood*, will most likely exhibit varying frequencies from a series that strictly sticks to script.<sup>61</sup>

To summarize this section, linguistic patterns that guide characterization may resemble each other across different series. This is particularly so with established and/or salient characterization patterns, as we have seen above. However, this is not to mean that characterization patterns are universally applicable across television series, in particular when it comes to comparing frequencies within linguistic patterns.

## **7.5 A fictolinguistic approach: an evaluation**

The fictolinguistic approach followed throughout this study allowed me to sidestep the initial question of authorship and 'realness' of scripted language. The focus on the characters as the enactors of language meant that I could consider *their* fictional background as the influence of their patterned variation rather than the social background of scriptwriters or actors.

The brief exploration of character shifts in the cases of Lorelai and Sarah in chapter 6 suggests that linguistic analyses comparing similar character roles across series might be

---

<sup>61</sup> While not part of the current study, external influence will also affect distributions of swear words, as highly regulated by traditional network companies.

a promising next step. Both characters have similarities in their social background (independent single mothers, returning to their family and struggling in romantic relationships) and are incidentally portrayed by the same actor (Lauren Graham). A closer analysis would allow more insights into the effects of television production and a possible unpacking of the idea of an author's voice within dialogue.

The creative input in providing fictional dialogue is undoubtedly complex as many voices work together to create fictional characters. By only focusing on the end product here, I was able to assign authorship and authenticity values to the characters themselves. This approach put me as a researcher on the side of the consumer rather than the producer of language. The findings I gained through this approach showed that scripted language is as varied as unscripted language and that characters are using language as a marker of identity for themselves and those around them. Challenging the dialogue itself by calling its resemblance to naturally occurring language into question would diminish its profound function within the fictional world. Thus, the fictolinguistic approach helped me “towards the prioritising of literary function over linguistic accuracy” (Hodson, 2017: 20).

## 8 Conclusion

The focus of this study was driven by the sociolinguistics of identity construction which necessitates, in my view, the central role of the speaker or character in the analysis. With that, my findings tie back in with chapter 2 of this thesis and how we understand fictional characters. Linguistic variation I found across all 37 characters contributes to each individual's recognisability and individualization. Further, linguistic variation enables the characters to distinguish themselves from each other in systematic and comprehensible ways. Staying within the fictolinguistic approach allowed me to pinpoint particular character differences that are implied through linguistic variation.

The wider and arguably more impactful picture for most sociolinguists is to find out what this can tell us beyond fiction. Chapter 6 provided an analysis that investigates how language is shifting over time and how, in this instance, pragmatic markers and hedges are developing in terms of delexicalization and establishment. The findings in the chapter were related to issues of characterization, but are also relevant when investigating language developments over time more generally. By analysing language use of fictional characters, I ultimately also analyse how language itself and meaning in language is perceived by the creators of television dialogue. In this concrete instance, I found that discourse values (d-values) can reflect the shifting functions of language features, such as *like*. This offers insights for sociolinguists into how languages are developing over time and where language features shift semantically and/or pragmatically. While the argument might hold true that scripted language is not an accurate reflection of naturally occurring language, it is worth pointing out that even if it is not 'real', it is likewise not fake or improbable. Thus, fictional television dialogue can provide a resource when tracing syntactic expansions of linguistic features. Further, we might be able to judge

attitudes towards language by monitoring how many characters (and which characters) use certain features. Remaining with the example of pragmatic marker *like*, current findings suggest that it is still somewhat stigmatized and associated with a particular speaker group. A future investigation might find that the feature is adopted by more characters and characters that do not fall into the stereotype of young, female American. Tying these observations together, I argue that fictional television series (or comparable audio-visual media) can be a meaningful resource for perceptual sociolinguistics.

Finally, the study has provided novel insight into how fictional television characters are created and maintained through their linguistic profile. I highlighted the important role dialogue plays within audio-visual media, challenging the notion that “actions speak louder than dialogue” (Mittell, 2015:135). Fictional television dialogue serves a great number of functions within the context it is created in (first and foremost as a means for characters to communicate), as well as contexts it can be appropriated for. This research thus adds to an increasing number of studies that prove that television as a cultural artefact deserves to be front and centre when we wonder how our language is used and how it is changing.

## List of Media Sources

- Angel Synopsis. (2017) *Angel synopsis IMDb*. Available at: [http://www.imdb.com/title/tt0162065/?ref\\_=nv\\_sr\\_1](http://www.imdb.com/title/tt0162065/?ref_=nv_sr_1).
- Angel*. (1999-2004). creators: Joss Whedon & David Greenwalt. The WB.
- Best, C. (2016) *Parenthood Transcripts Seasons One to Four*. Available at: <https://crazy-internet-people.com/site/parenthood/> [Last accessed May 2016].
- Buffy Synopsis. (2017) *Buffy the Vampire Slayer synopsis IMDb*. Available at: [http://www.imdb.com/title/tt0118276/?ref\\_=nv\\_sr\\_1](http://www.imdb.com/title/tt0118276/?ref_=nv_sr_1).
- Buffy the Vampire Slayer*. (1997-2003). creator: Joss Whedon. The WB (1997–2001), UPN (2001–2003).
- Chakoteya.net. (2015). *The Torchwood Transcripts - Episode listing, Season One to Four*. [online] Available at: <http://www.chakoteya.net/Torchwood/episodes.htm> [Last accessed July 2015].
- George Bischel, H. (2015). *BuffyWorld — "Angel" — Season One*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/angel/season1.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Angel" — Season Two*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/angel/season2.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Angel" — Season Three*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/angel/season3.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Angel" — Season Four*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/angel/season4.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Angel" — Season Five*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/angel/season5.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Buffy the Vampire Slayer" — Season One*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season1.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Buffy the Vampire Slayer" — Season Two*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season2.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Buffy the Vampire Slayer" — Season Three*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season3.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Buffy the Vampire Slayer" — Season Four*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season4.php> [Last accessed July 2015].
- (2015). *BuffyWorld — "Buffy the Vampire Slayer" — Season Five*. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season5.php> [Last accessed July 2015].

- (2015). *BuffyWorld* – "Buffy the Vampire Slayer" – Season Six. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season6.php> [Last accessed July 2015].
- (2015). *BuffyWorld* – "Buffy the Vampire Slayer" – Season Seven. [online] Buffyworld.com. Available at: <http://www.buffyworld.com/buffy/season7.php> [Last accessed July 2015].
- Crazy-internet-people.com. (2015). *Gilmore Girls - Season 1 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s1/s1s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls - Season 2 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s2/s2s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls - Season 3 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s3/s3s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls - Season 4 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s4/s4s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls - Season 5 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s5/s5s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls - Season 6 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s6/s6s.html> [Last accessed July 2015].
- (2015). *Gilmore Girls – Season 7 Transcripts*. [online] Available at: <https://crazy-internet-people.com/site/gilmoregirls/pages/s7/s7s.html> [Last accessed July 2015].
- Gilmore Girls*. (2000-2007). creator: Amy Sherman-Palladino. The WB (2000–2006), The CW (2006–2007).
- Parenthood*. (2010-2015). creator: Jason Katims. NBC
- Sherlock*. (2010-2017). creators: Mark Gatiss & Steven Mofatt. BBC One.
- Torchwood*. (2006-2011). creator: Russel T. Davies. BBC Three (2006), BBC Two (2008), BBC One (2009), Starz HD (2011).
- Transcripts.foreverdreaming.org. (2016). *Sherlock Transcripts*. [online] Available at: <http://transcripts.foreverdreaming.org/viewforum.php?f=51> [Last accessed May 2016].

## References

- Adams M. (2003) *Slayer Slang - A Buffy the Vampire Slayer Lexicon*: Oxford University Press USA.
- . (2011) *From Elvish to Klingon*: Oxford University Press.
- . (2013) Vignette 13b - Working with Scripted Data: Variations among Scripts, Texts, and Performances. In: Mallinson C, Childs B and Van Herk G (eds) *Data Collection in Sociolinguistics - Methods and Applications*. New York: Routledge, 232-235.
- Adolphs S. (2006) *Introducing electronic text analysis: A practical guide for language and literary studies*: Routledge.
- Aijmer K. (1984) 'Sort of' and 'Kind of' in English conversation. *Studica Linguistica* 38: 118-128.
- . (2002) *English Discourse Particles - Evidence from a corpus*, Amsterdam, Philadelphia: John Benjamins.
- Androutsopoulos J. (2012a) Introduction: Language and society in cinematic discourse. *Multilingua-Journal of Cross-Cultural and \dots* 31: 139--154.
- . (2012b) Repertoires, characters and scenes: Sociolinguistic difference in Turkish German comedy. *Multilingua* 31: 301--326.
- . (2014) Mediatization and sociolinguistic change. Key concepts, research traditions, open issues. *Mediatization and sociolinguistic change*: 3-48.
- Angel. (2017) *Angel synopsis IMDb*. Available at: [http://www.imdb.com/title/tt0162065/?ref =nv\\_sr\\_1](http://www.imdb.com/title/tt0162065/?ref =nv_sr_1).
- Anthony L. (2014) AntConc (Version 3.4. 3)[Computer Software]. Tokyo, Japan: Waseda University.
- Bailey G. (2008) Real and Apparent Time. In: Chambers JK, Trudgill P and Schilling-Estes N (eds) *The Handbook of Language Variation and Change*. Blackwell Publishing Ltd, 312-332.
- Barchas J. (2007) Very Austen: Accounting for the language of Emma. *Nineteenth-Century Literature* 62: 303-338.
- Barnfield K and Buchstaller I. (2010) Intensifiers in Tyneside: Longitudinal developments and new trends. *English World-Wide* 31: 252-287.
- Bauer L and Bauer W. (2002) Adjective Boosters in the English of Young New Zealanders. *Journal of English Linguistics* 30: 244-257.
- Bayley R and Regan V. (2004) Introduction: The acquisition of sociolinguistic competence. *Journal of Sociolinguistics* 8: 323-338.
- Bednarek M. (2010) *The Language of Fictional Television: Drama and Identity*, London: Continuum.
- . (2011a) Expressivity and televisual characterization. *Language and Literature* 20: 3-21.
- . (2011b) The stability of the televisual character - a corpus stylistic case study. In: Piazza R, Bednarek M and Rossi F (eds) *Telecinematic Discourse: Approaches to the language of films and television series*. Amsterdam/Philadelphia: John Benjamins Publishing, 185-204.
- . (2012a) Constructing 'nerdiness': Characterisation in The Big Bang Theory. *Multilingua* 31: 199-229.
- . (2012b) "Get us the hell out of here": Key words and trigrams in fictional television series. *International Journal of Corpus Linguistics* 17: 35-63.

- Bednarek M and Zago R. (2017) *Bibliography of linguistic research on fictional (narrative, scripted) television series and films/movies, version 1 (January 2017)*. Available at: <http://unipv.academia.edu/RaffaeleZago>.
- Beeching K. (2016) *Pragmatic Markers in British English - Meaning in Social Interaction*, Cambridge: Cambridge University Press.
- Beers Fägersten K. (2016) *Watching TV with a linguist*, Syracuse: Syracuse University Press.
- Bell A. (1992) Hit and miss: Referee design in the dialects of New Zealand television advertisements. *Language & Communication* 12: 327-340.
- Bell A and Gibson A. (2011) Staging language: An introduction to the sociolinguistics of performance. *Journal of Sociolinguistics* 15: 555-572.
- Berman AS. (2014) *The Gilmore Girls Companion*, Albany, Georgia: BearManor Media.
- Biber D, Johansson S, Leech G, et al. (1999) *Longman Grammar of Spoken and Written English*, London: Pearson Education Limited.
- Bleichenbacher L. (2012) Linguicism in Hollywood movies? Representations of, and audience reactions to multilingualism in mainstream movie dialogues. *Multilingua* 31: 155--176.
- Bolinger D. (1972) *Degree Words*, The Hague: Nouton.
- Brown P and Levinson SC. (1978) Universals in language usage: Politeness phenomena. *Questions and politeness: Strategies in social interaction*. Cambridge University Press, 56-311.
- Bubel C. (2006) The linguistic construction of character relations in TV drama: doing friendship in "Sex and the City". Saarbrücken: Universität des Saarlandes.
- Bubel C and Spitz A. (2006) "One of the last vestiges of gender bias": The characterization of women through the telling of dirty jokes in Ally McBeal. *Humour* 9: 71-104.
- Bucholtz M. (1999) "Why be normal?": Language and Identity Practices. *Language and Society* 28: 203-223.
- . (2005) Identity and interaction: a sociocultural linguistic approach. 585--614.
- . (2009) From stance to style: Gender, interaction, and indexicality in Mexican immigrant youth slang.
- Bucholtz M, Bermudez N, Fung V, et al. (2007) Hella nor cal or totally so cal?: The perceptual dialectology of California. *Journal of English Linguistics* 35: 325-352.
- Buchstaller I. (2009) The Quantitative Analysis of Morphosyntactic Variation: Constructing and Quantifying the Denominator. *Language and Linguistics Compass* 3: 1010–1033.
- Buffy. (2017) *Buffy the Vampire Slayer synopsis IMDb*. Available at: [http://www.imdb.com/title/tt0118276/?ref =nv\\_sr\\_1](http://www.imdb.com/title/tt0118276/?ref =nv_sr_1).
- Burrows JF. (1987) *Computation into criticism : a study of Jane Austen's novels and an experiment in method*, Oxford: Clarendon.
- Cameron D. (1996) The language-gender interface: challenging co-optation. In: Bergvall VL, Bing JM and Freed AF (eds) *Rethinking Language and Gender Research: Theory and Practice*. London: Longman, 31-53.
- Chambers JK. (1995) *Sociolinguistic Theory*, Oxford, Cambridge: Blackwell.
- Channell J. (1994) *Vague language*, Oxford: Oxford University Press.
- Cheshire J. (1982) *Variation in an English dialect: A sociolinguistic study*, Cambridge: Cambridge University Press.
- . (2002) Sex and gender in variationist research. In: Chambers JK, Trudgill P and Schilling-Estes N (eds) *Handbook of Language Variation and Change*. Oxford: Blackwell, 423-443.

- Clark HH and Fox Tree JE. (2002) Using *uh* and *um* in spontaneous speaking. *Cognition* 84: 73-111.
- Coates J. (2003) The role of epistemic modality in women's talk. In: Facchinetti R, Krug MG and Palmer FR (eds) *Modality in Contemporary English*. Berlin/ New York: Mouton de Gruyter, 331-348.
- . (2013) *Women, Men and Everyday Talk*, Basingstoke: Palgrave Macmillan.
- . (2015) *Women, men and language: A sociolinguistic account of gender differences in language*: Routledge.
- Cornillie B. (2009) Evidentiality and epistemic modality: On the close relationship between two different categories. *Functions of Language* 16: 44-62.
- Coupland N. (2004) Age in social and sociolinguistic theory. In: Nussbaum JF and Coupland J (eds) *Handbook of communication and aging research*. Mahwah/ London: Lawrence Erlbaum Associates, 69-90.
- . (2009) The mediated performance of vernaculars. *Journal of English Linguistics* 37: 284-300.
- . (2011) Voice, place and genre in popular song performance. *Journal of Sociolinguistics* 15: 573-602.
- . (2014) Language change, social change, sociolinguistic change: A meta-commentary. *Journal of Sociolinguistics* 18: 277-286.
- Culpeper J. (2001) *Language and characterisation : people in plays and other texts*, Harlow: Longman.
- Davies R. (2006) Torchwood [Television Series]. Cardiff, UK: BBC Wales.
- Denis D and Tagliamonte S. (2017) Language change and fiction. In: Locher MA and Jucker AH (eds) *Pragmatics of Fiction*. Berlin/Boston: De Gruyter Mouton, 553-584.
- Dhaenens F. (2013) The fantastic queer: Reading gay representations in Torchwood and True Blood as articulations of queer resistance. *Critical Studies in Media Communication* 30: 102-116.
- Dines ER. (1980) Variation in discourse—"and stuff like that". *Language in Society* 9: 13-31.
- Djonov E and Zhao S. (2014) *Critical multimodal studies of popular discourse*, New York: Routledge.
- Doležel L. (1988) Mimesis and Possible Worlds. *Poetics Today* 9: 475-496.
- Dose S. (2013) Flipping the script: A Corpus of American Television Series (CATS) for corpus-based language learning and teaching. *Studies in Variation, Contacts and Change in English - Corpus Linguistics and Variation in English: Focus on Non-Native Englishes* 2013: n.P.
- Dragojevic M, Mastro D, Giles H, et al. (2016) Silencing nonstandard speakers: A content analysis of accent portrayals on American primetime television. *Language in Society* 45: 59-85.
- Dumas NW. (2016) "This guy says I should talk like that all the time": Challenging intersecting ideologies of language and gender in an American Stuttering English comedienne's stand-up routine. *Language in Society* 45: 353-374.
- Eckert P. (1984) Age and linguistic change. In: Kertzer DI and Keith J (eds) *Age and Anthropological Theory*. Ithaca, NY: Cornell University Press, 219-233.
- . (1997) Age as a sociolinguistic variable. In: Coulmas F (ed) *The Handbook of Sociolinguistics*. Oxford: Blackwell, 151-167.
- . (2000) *Language variation as social practice: The linguistic construction of identity in Belten High*, Malden/ Oxford: Wiley-Blackwell.
- . (2002) Constructing meaning in sociolinguistic variation. *Annual Meeting of the American Anthropological Association, New Orleans*.

- . (2014) The Problem with Binaries: Coding for Gender and Sexuality. *Language and Linguistics Compass* 8: 529-535.
- Eckert P and McConnell-Ginet S. (1992) Think practically and look locally: Language and gender as community-based practice. *Annual review of Anthropology* 21: 461-488.
- . (1999) New generalizations and explanations in language and gender research. *Language in Society* 28: 185-201.
- Eder J. (2008) *Die Figur im Film: Grundlagen der Figurenanalyse*, Marburg: Schüren.
- Ehrlich S, Meyerhoff M and Holmes J. (2017) *The handbook of language, gender, and sexuality*: John Wiley & Sons.
- Ellcessor E. (2012) Tweeting @feliaday: Online Social Media, Convergence, and Subcultural Stardom. *Cinema Journal* 51: 46-66.
- Erickson G. (2002) "Sometimes you need a story": American Christianity, vampires, and Buffy. In: Wilcox R and Lavery D (eds) *Fighting the forces : what's at stake in Buffy the Vampire Slayer*. Lanham, Md.: Rowman & Littlefield, 108-119.
- Erman B. (1987) Pragmatic Expressions in English. A Study of You know, You see and I mean in Face-to-Face Conversation. *Stockholm studies in English* 69: 1-238.
- . (1995) Grammaticalization in progress; The case of or something. In: Moen I, Gram Simonsen H and Ladrup H (eds) *XVth Scandinavian Conference of Linguistics*. Oslo, Norway: Department of Linguistics University of Oslo, 136-147.
- . (2001) Pragmatic markers revisited with a focus on you know in adult and adolescent talk. *Journal of Pragmatics* 33: 1337-1359.
- Ferguson SL. (1998) Drawing fictional lines: dialect and narrative in the Victorian novel. *Style-Fayetteville-* 32: 1-17.
- Fetzer A. (2010) Hedges in Context: Form and Function of *Sort of* and *Kind of*. In: Kaltenböck G, Mihatsch W and Schneider S (eds) *New Approaches to Hedging*. Bingley: Emerald Group Publishing, 49-72.
- Field S. (2003) *The Definitive Guide to Screenwriting*, London: Ebury Press.
- Fox Tree JE. (2007) Folk notions of *um*, and *uh*, *you know*, and *like*. *Text & Talk* 27: 297-314.
- Fraser B. (2010) Pragmatic Competence: The Case of Hedging. In: Kaltenböck G, Mihatsch W and Schneider S (eds) *New Approaches to Hedging*. 15-34.
- Giles DC. (2010) Parasocial Relationships. In: Eder J, Jannidis F and Schneider R (eds) *Characters in Fictional Worlds: Understanding Imaginary Beings in Literature, Film, and Other Media*. Berlin, New York: de Gruyter, 442-458.
- Giroux HA and Simon RI. (1989) *Popular culture, schooling, and everyday life*: JF Bergin & Garvey.
- Gonzalez-Diaz V. (2014) 'I quite detest the man': Degree adverbs, female language and Jane Austen. *Language and Literature* 23: 310-330.
- Graham L. (2016) *Talking as Fast as I Can*, New York City: Ballantine Books.
- Gries S and David C. (2007) This is kind of / sort of interesting: variation in hedging in English. In: Päivi Pahta IT, Terttu Nevalainen & Jukka Tyrkkö (ed) *Studies in Variation, Contacts and Change in English: Towards Multimedia in Corpus Studies*. Helsinki: Research Unit for Variation, Contacts and Change in English (VARIENG).
- Harwood J and Giles H. (1992) 'Don't make me laugh': age representations in a humorous context. *Discourse and Society* 3: 403-436.
- Helford ER. (2002) "My emotions give me power": The Containment of Girls' Anger in Buffy. In: Wilcox R and Lavery D (eds) *Fighting the forces : what's at stake in Buffy the Vampire Slayer*. Lanham, Md.: Rowman & Littlefield, 18-34.

- Herbst T. (1994) Linguistische Aspekte der Synchronisation von Fernsehserien Phonetik, Textlinguistik, Übersetzungstheorie.
- Higgins CaFG. (2012) Styling Hawai 'i in Haole wood: White protagonists on a voyage of self discovery. *Multilingua* 31: 177--198.
- Hodson J. (2014) *Dialect in film and literature*: Basingstoke, Hampshire : Palgrave Macmillan.
- . (2017) *Dialect and Literature in the Long Nineteenth Century*, London, New York: Taylor & Francis.
- Holmes J. (1990) Hedges and Boosters in women's and men's speech. *Language and Communication* 10: 185-205.
- . (2008) Language and gender. *Language Teaching* 24: 207-220.
- Ilhem M-S. (2013) Rhetorical Use of Literary Dialect in English Literature: From Chaucer to Shaw. *International Journal of English Language and Literature Studies* 2: 102-123.
- Ito R and Tagliamonte S. (2003) Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers. *Language in Society* 32: 257-279.
- Johnson DE. (2009) Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3: 359-383.
- Johnson SA and Milani TM. (2010) Language ideologies and media discourse : texts, practices, politics. continuum: London/New York.
- Johnstone B. (2009) Stance, style, and the linguistic individual. *Stance: sociolinguistic perspectives*: 29-52.
- . (2010) Locating Language in Identity. In: Llamas C and Watt D (eds) *Language and Identities*. Edinburgh: Edinburgh University Press, 29-36.
- Kay P. (1984) The Kind of/ Sort of Construction. *Tenth Annual Meeting of the Berkeley Linguistics Society*. 157-171.
- Kirchner JS. (2006) And in some language that's English? Slayer Slang and Artificial Computer Generation. *Slayage - The online international journal of Buffy studies* 20.
- Kozloff S. (2000) *Overhearing Film Dialogue*, Berkeley/Los Angeles: University of California Press.
- Kranich S. (2011) To hedge or not to hedge: the use of epistemic modal expressions in popular science in English texts, English–German translations, and German original texts. *Text & Talk-An Interdisciplinary Journal of Language, Discourse & Communication Studies* 31: 77-99.
- Labov W. (1963) The Social Motivation of a Sound Change. *Word* 19: 273-309.
- . (1972a) *Sociolinguistic Patterns*, Philadelphia: University of Pennsylvania Press.
- . (1972b) Some principles of linguistic methodology. *Language in Society* 1: 97-120.
- . (1978) Where does the linguistic variable stop? A response to Beatrix Lavandera. *Sociolinguistic Working Paper* 44: 4-22.
- . (1984) Intensity. In: Schiffrin D (ed) *Meaning, form, and use in context: Linguistic applications (GURT)*. Washington, D.C.: Georgetown University Press, 43-70.
- . (1990) The intersection of Sex and Social Class in the Course of Linguistic Change. *Language Variation and Change* 2: 205-254.
- . (2001) *Principles of Linguistic Change, Vol. 2: Social Factors*, Oxford: Blackwell.
- . (2009) The Transmission Problem in Linguistic Change. In: Coupland N and Jaworski A (eds) *The New Sociolinguistics Reader*. Basingstoke: Palgrave Macmillan, 66-74.
- Lakoff G. (1973) Hedges: A Study of Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2: 458 - 508.

- Landert D. (2017) Stance in Fiction. In: Locher MA and Jucker AH (eds) *Pragmatics of Fiction*. Berlin/Boston: De Gruyter Mouton, 489-514.
- Leibniz GW. (2000) *Theodicy: Essays on the Goodness of God, the Freedom of Man and the Origin of Evil*. Wipf and Stock Publishers.
- Lorenzo-Dus N and Garcés-Conejos Blitvich P. (2013) *Real talk : reality television and discourse analysis in action*, Basingstoke: Palgrave Macmillan.
- Macaulay R. (2002) You know, it depends. *Journal of Pragmatics* 34: 749-767.
- . (2006) Pure grammaticalization: The development of a teenage intensifier. *Language Variation and Change* 18: 267-283.
- Mandala S. (2007) Solidarity and the Scoobies: An analysis of the -y suffix in the television series Buffy the Vampire Slayer. *Language and Literature* 16: 53-73.
- . (2008) Representing the Future: Chinese and Codeswitching in Firefly. 31-40.
- . (2010) *The Language in Science Fiction and Fantasy: The Question of Style*, London/New York: Continuum.
- Martinez IMP and Pertejo PN. (2012) He's absolutely massive. It's a super day. Madonna, she's a wicked singer. Youth Language and Intensification: A Corpus-Based Study. *Text &Talk* 32: 773-796.
- Masson C. (2006) "Is that just a comforting way of not answering the question?": Willow, Questions, and Affective Response in *Buffy the Vampire Slayer*. *Slayage - The online international journal of Buffy studies* 20.
- McKee R. (2016) *Dialogue: The Art of Verbal Action for Page, Stage, Screen*, New York/Boston: Twelve.
- Milroy L. (1980) *Language and Social Networks*, Oxford: Blackwell.
- Mittell J. (2015) *Complex TV - The Poetics of Contemporary Television Storytelling*, New York, London: New York University Press.
- Mittmann B. (2006) With a little help from Friends (and others): Lexico-pragmatic characteristics of original and dubbed film dialogue. In: Houswitschka C, Knappe G and Müller A (eds) *Anglistentag 2005*. Bamberg: Trier - WVT, 573-585.
- Moore E. (2003) Learning style and identity: A sociolinguistic analysis of a Bolton high school. University of Manchester.
- . (2004) Sociolinguistic style: A multidimensional resource for shared identity creation. *Canadian Journal of Linguistics/Revue canadienne de linguistique* 49: 375-396.
- . (2006) 'You tell all the stories': Using narrative to explore hierarchy within a Community of Practice1. *Journal of Sociolinguistics* 10: 611-640.
- . (2010) Interaction between social category and social practice: explaining was/were variation. *Language Variation and Change* 22: 347-371.
- Moss G. (2001) From the Valley to the Hellmouth: "Buffy"'s Transition from Film to Television. *Slayage* 1.
- Murphy B. (2010a) *Corpus and Sociolinguistics: Investigating Age and Gender in Female Talk*, Amsterdam, Netherlands: Benjamins.
- Murphy Bn. (2010b) *Corpus and sociolinguistics investigating age and gender in female talk*, Amsterdam: Amsterdam : John Benjamins.
- Nussbaum D. (2016) Gilmore Girls: See the first page of the reboot script. *Entertainment Weekly*. New York: Time Inc.
- Ochs E. (1992) 14 Indexing gender. *Rethinking context: Language as an interactive phenomenon* 11: 335.
- Paradis C. (1997) *Degree modifiers of adjectives in spoken British English*, Lund: Lund University Press.
- . (2000) It's well weird: Degree modifiers of adjectives revisited: the nineties. In: Kirk J (ed) *Corpora galore: analyses and techniques in describing English*. Amsterdam: Rodopi, 147-160.

- Petrucci P. (2012) The translation of cinematic discourse and the question of character equivalence in *Talk to me*. *Multilingua* 31.
- Piazza R, Bednarek M and Rossi F. (2011a) Introduction: Analysing telecinematic discourse. In: Piazza R, Bednarek M and Rossi F (eds) *Telecinematic discourse: Approaches to the language of films and television series*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-17.
- . (2011b) *Telecinematic Discourse: Approaches to the language of films and television series*. Amsterdam, Philadelphia: John Benjamins.
- Pichler H. (2010) Methods in discourse variation analysis: Reflections on the way forward. *Journal of Sociolinguistics* 14: 581-608.
- Planchenault G. (2012) Accented French in films: Performing and evaluating in-group stylisations. *Multilingua* 31: 253--275.
- Podesva RJ. (2011) The California vowel shift and gay identity. *American Speech* 86: 32-51.
- Powers T. (2016) *Gender and the Quest in British Science Fiction Television: An Analysis of Doctor Who, Blake's 7, Red Dwarf and Torchwood*: McFarland.
- Prince E, Frader J and Bosk C. (1982) On hedging in physician-physician discourse. In: di Pietro RJ (ed) *Linguistics and the Professions*. New Jersey: Ablex, 83-97.
- Quaglio P. (2009) *Television Dialogue The sitcom friends vs. natural conversation*, Amsterdam: John Benjamins Publishing Company.
- Queen R. (2004) 'Du hast jar keene Ahnung': African American English dubbed into German. *Journal of Sociolinguistics* 8: 515--537.
- . (2015a) Keynote: Indexical Authenticity and Linguistic Variation in Fictional Audiovisual Media. *Language in the Media* 6. Hamburg.
- . (2015b) *Vox Popular - The Surprising Life of Language in the Media*, Chichester: John Wiley & Sons.
- Quirk R, Greenbaum S, Leech G, et al. (1980) *A Grammar of Contemporary English*, London: Longman.
- Rawlins JO. (2010) Your Guide to the Girls: Gilmore-isms, Cultural Capital, and a Different Kind of Quality TV. In: Diffrient DS and Lavery D (eds) *Screwball Television: Critical Perspectives on Gilmore Girls*. Syracuse: Syracuse University Press, 36-56.
- Reich N. (2013) Between Admiration and Administration: Code-Switching, Style-Shifting, and Sociolinguistic Crossing in Philip Meadows Taylor's *Confessions of a Thug*. *The English Languages: History, Diaspora, Culture* 4: 1-19.
- Reichelt S and Durham M. (2017) Adjective Intensification as a Means of Characterization: Portraying In-Group Membership and Britishness in *Buffy the Vampire Slayer*. *Journal of English Linguistics* 45: 60-87.
- Rey JM. (1996) To Boldly Speak : Changes in the Linguistic Characterization of Women in the Televised *Star Treks*. *Deseret Language and Linguistic Society Symposium* 22: 65-72.
- Richardson K. (1999) *Worlds in common? : television discourse in a changing Europe*, New York: New York : Routledge.
- . (2010) *Television Dramatic Dialogue - A Sociolinguistic Study*, Oxford: Oxford University Press.
- Richardson K and Queen R. (2012) Describing, analysing and judging language codes in cinematic discourse. *Multilingua* 31: 327--336.
- Rickford JR, Wasow T, Zwicky A, et al. (2007) Intensive and Quotative all: Something Old, Something New. *American Speech* 82: 3-31.
- Robin L. (1975) Language and woman's place. *Journal of Women s Health*.

- Rodgers MPH and Webb S. (2011) Narrow Viewing: The Vocabulary in Related Television Programs. *Tesol Q.* 45: 689-717.
- Ruddell C. (2006) "I am the Law" "I am the Magics": Speech, Power and the Split Identity of Willow in Buffy the Vampire Slayer. *Slayage - The online international journal of Buffy studies* 20.
- Sayers D. (2014) The Mediated Innovation Model: A Framework for Researching Media Influence in Language Change. *Journal of Sociolinguistics* 18: 185-212.
- Scannell P. (1991) *Broadcast talk*, London: Sage Publications.
- Schiffrin D. (1987) *Discourse Markers*, Cambridge: Cambridge University Press.
- Schudson M. (1987a) The New Validation of popular Culture: Sense and Sentimentality in Academia. *Critical Studies in Mass Communication* 4: 51-68.
- . (1987b) The new validation of popular culture: sense and sentimentality in academia. *Critical Studies in Media Communication* 4: 51-68.
- Searle JR. (1975) The logical status of fictional discourse. *New Literary History: on Narrative and Narratives* 6: 319-332.
- Semino E. (2014) *Language and World Creation in Poems and Other Texts*, London: Taylor & Francis.
- Silverstein M. (2003) Indexical order and the dialectics of sociolinguistic life. *Language and Communication* 23: 193-229.
- Smith J, Durham M and Fortune L. (2007) "Mam, my trousers is fa'in doon!": Community, caregiver, and child in the acquisition of variation in a Scottish dialect. *Language Variation and Change* 19: 63-99.
- Stamou AG. (2011) Speech style and the construction of social division: Evidence from Greek television. *Language & Communication* 31: 329-344.
- . (2012) Representations of linguistic variation in children's books: register stylisation as a resource for (critical) language awareness. *Language Awareness* 21: 313-329.
- . (2014) A literature review on the mediation of sociolinguistic style in television and cinematic fiction: Sustaining the ideology of authenticity. *Language and Literature* 23: 118-140.
- Stamou AG, Agrafioti A and Dinas KD. (2012) Representations of youth (language) in Greek TV commercials. *Journal of Youth Studies* 15: 909-928.
- Stenström A-B. (1990) Lexical items peculiar to spoken discourse. In: Svartvik J (ed) *The London-Lund Corpus of Spoken English: Description and Research*. Lund Lund University Press 137-175.
- . (1999) He Was Really Gormless-She's Bloody Crap. Girls, Boys and Intensifiers. In: Hasselgard H and Oksefjell S (eds) *Out of Corpora: Studies in Honour of Stig Johansson*. Amsterdam: Rodopi, 69-78.
- Stenström A-B, Andersen G and Hasund IK. (2002) *Trends in Teenage Talk - Corpus Compilation, Analysis and Findings.*, Amsterdam/Philadelphia: John Benjamins.
- Stuart-Smith J. (2014) No longer an elephant in the room. *Journal of Sociolinguistics* 18: 250-261.
- Suzuki D. (2015) Form and function of the modal adverbs: recent linguistic change and constancy in British English. *Linguistics* 53: 1365-1389.
- Tagliamonte S. (2005) *So who? Like how? Just what?* Discourse markers in the conversations of Young Canadians. *Journal of pragmatics : an interdisciplinary journal of language studies* 37.
- . (2006) *Analysing sociolinguistic variation*, Cambridge: Cambridge University Press.
- . (2008) So different and pretty cool! Recycling intensifiers in Toronto, Canada. *English Language and Linguistics* 12: 361-394.
- Tagliamonte S and Denis D. (2010) The Stuff of Change: General Extenders in Toronto, Canada. *Journal of English Linguistics* 38: 335-368.

- Tagliamonte S and Roberts C. (2005) So weird; so cool; so innovative: The use of intensifiers in the television series Friends. *American Speech* 80: 280-300.
- Tagliamonte SA. (2014) Situating media influence in sociolinguistic context. *Journal of Sociolinguistics* 18: 223-232.
- Thornham S and Purvis T. (2005) Television Drama - Theories and Identities. Basingstoke: Palgrave Macmillan.
- Trudgill P. (1972) Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society* 1: 179-195.
- . (1974) *The Social Differentiation of English in Norwich*, Cambridge: Cambridge University Press.
- . (2014) Diffusion, drift, and the irrelevance of media influence. *Journal of Sociolinguistics* 18: 213-222.
- Ventola E and Moya Guijarro AJs. (2009) The world told and the world shown : multisemiotic issues. *World told and the world shown : multisemiotic issues*. Basingstoke/ New York: Palgrave Macmillan.
- Wagner SE. (2012) Real-time evidence for age grad (ing) in late adolescence. *Language Variation and Change* 24: 179-202.
- Weinreich U, Labov W and Herzog M. (1968) Empirical Foundations for a theory of language change. In: Lehmann W and Malkiel Y (eds) *Directions for Historical Linguistics*. Austin: University of Texas Press, 97-195.
- Wilde J. (2015) Gay, Queer, or Dimensional? Modes of Reading Bisexuality on Torchwood. *Journal of Bisexuality* 15: 414-434.
- Xiao R and Tao H. (2007) A corpus-based sociolinguistic study of amplifiers in British English. *Sociolinguistic Studies* 1: 241-273.



## Appendices

## Appendix A: Distribution tables

### *Pragmatic markers*

Character	series	<i>like</i>	<i>I mean</i>	<i>you know</i>	
Angel	<i>Angel</i>	1.00	12.68	19.6	
Cordelia		2.89	12.53	10.4	
Spike		1.46	5.12	2.9	
Xander	<i>Buffy the Vampire Slayer</i>	3.76	18.97	14.4	
Dawn		8.02	20.35	18.5	
Cordelia		8.25	17.25	12.7	
Willow		4.79	30.42	18.7	
Angel		0.00	11.79	17.2	
Buffy		2.68	15.71	23.2	
Anya		1.76	36.10	24.2	
Spike		0.00	6.21	16.8	
Giles		0.00	8.45	6.8	
Rory		<i>Gilmore Girls</i>	1.31	16.62	9.9
Sookie			0.72	19.37	14.8
Luke	0.47		18.25	21.4	
Lorelai	1.20		12.29	15.3	
Paris	0.50		12.99	8.0	
Emily	0.00		5.06	5.9	
Richard	0.00		1.12	5.6	
Julia	4.51		14.67	48.0	
Crosby	<i>Parenthood</i>	11.49	20.98	51.7	
Amber		47.88	27.50	55.0	
Zeek		5.10	58.66	57.1	
Haddie		77.71	29.64	20.0	
Sarah		14.36	29.61	74.0	
Camille		2.24	14.55	88.4	
Max		4.46	4.46	14.3	
Adam		2.58	10.31	35.6	
Kristina		14.20	23.92	31.0	
Sherlock		<i>Sherlock</i>	0.00	0.59	4.4
John			0.00	5.51	8.6
Jack	<i>Torchwood</i>	0.00	2.53	3.4	
Owen		1.31	4.58	17.0	
Toshiko		1.06	3.19	1.1	
Rhys		3.98	9.95	21.9	
Ianto		0.00	2.86	4.3	
Gwen		0.97	7.77	3.9	
mean overall			6.23	14.93	21.78
mean by series	<i>Angel</i>	1.78	10.11	10.9	
	<i>Buffy the Vampire Slayer</i>	3.25	18.36	16.9	
	<i>Gilmore Girls</i>	0.60	12.24	11.56	
	<i>Parenthood</i>	18.45	23.43	47.51	
	<i>Sherlock</i>	0.00	3.05	6.51	
	<i>Torchwood</i>	1.22	5.15	8.58	
mean by gender	female	10.27	18.40	25.42	
	male	1.98	11.28	17.94	
mean by age	young	13.67	20.21	19.11	
	middle	2.82	10.98	19.99	
	older	1.84	19.85	39.26	
mean by nationality	American	8.22	18.46	26.61	
	British	0.88	5.42	0.00	

## Hedges

character	series	sort of	kind of
Angel	<i>Angel</i>	1.56	8.9
Cordelia		1.54	10.03
Spike		2.93	1.46
Angel	<i>Buffy the Vampire Slayer</i>	1.81	9.98
Anya		2.64	5.28
Buffy		3.55	10.58
Cordelia		3.75	5.25
Dawn		2.47	11.72
Giles		3.09	1.03
Spike		0.93	2.17
Willow		5.62	13.72
Xander		2.91	12.31
Emily		<i>Gilmore Girls</i>	1.6
Lorelai	1.45		4.19
Luke	1.77		5.53
Paris	1.75		2.5
Richard	2.52		0.56
Rory	0.98		7.58
Sookie	2.15		7.65
Adam	1.03		2.75
Amber	<i>Parenthood</i>	0.95	11.38
Camille		1.12	4.48
Crosby		0	12.49
Haddie		0	9.61
Julia		0	13.54
Kristina		6.57	6.84
Max		0	1.78
Sarah		2.24	5.83
Zeek		0	9.69
John		<i>Sherlock</i>	4.9
Sherlock	1.19		1.48
Gwen	<i>Torchwood</i>	5.83	1.46
Ianto		5.71	2.86
Jack		2.96	8.02
Owen		2.62	4.58
Rhys		1.99	1.99
Toshiko		2.13	3.19
mean overall			2.28
mean by series	<i>Angel</i>	2.01	6.79
	<i>Buffy the Vampire Slayer</i>	2.97	8.00
	<i>Gilmore Girls</i>	1.75	4.1
	<i>Parenthood</i>	1.19	7.84
	<i>Sherlock</i>	3.04	1.66
	<i>Torchwood</i>	3.54	3.68
mean by gender	female	2.44	7.13
	male	2.11	4.97
mean by age	young	2.18	8.48
	middle	2.37	8.82
	older	2.52	5.13
mean by nationality	American	1.96	7.51
	British	3.13	2.21

## General extenders

character	series	<i>or something</i>	<i>or anything</i>	<i>and stuff</i>	<i>and everything</i>	GE general
Angel	<i>Angel</i>	1.11	0.33	0	0	1.4
Cordelia		1.54	1.35	0.4	0.4	3.7
Spike		1.46	0	0	0	1.5
Angel	<i>Buffy the Vampire Slayer</i>	0.91	0	0	0	0.9
Anya		3.52	0.44	0	0.4	4.4
Buffy		1.89	0.39	0.6	0.5	3.4
Cordelia		2.25	1.5	2.2	3	9
Dawn		3.08	1.23	3.7	0.6	8.6
Giles		0.62	0	0	0.2	0.8
Spike		0	0	0.3	0	0.3
Willow		3.31	0.5	2.5	1.5	7.8
Xander		1.37	0.34	0.3	0.5	2.6
Emily		0.53	0	0	0.1	0.7
Lorelai		2.24	0.63	0.3	0.4	3.6
Luke	2.94	0.47	0.1	0.5	4	
Paris	1.25	0.75	0	0	2	
Richard	0.56	0	0	0.3	0.8	
Rory	2.02	0.55	0.2	0.8	3.5	
Sookie	1.43	0.24	0.5	0	2.2	
Adam	1.55	0	0	0	1.5	
Amber	6.64	0	3.3	0	10	
Camille	0	2.24	0	0	2.2	
Crosby	5	0.25	1.2	0	6.5	
Haddie	6.41	3.2	0.8	0	10.4	
Julia	2.26	0.56	0	0.6	3.4	
Kristina	1.05	0.26	0	0	1.3	
Max	1.78	0	0	0	1.8	
Sarah	2.47	0.67	1.3	1.1	5.6	
Zeek	2.04	0.51	0	1.5	4.1	
John	1.22	0	0	0	1.2	
Sherlock	0.3	0	0	0	0.3	
Gwen	0	0.49	1	0.5	1.9	
Ianto	2.86	0	0	0	2.9	
Jack	0	0	0	0.4	0.4	
Owen	0	0.65	0	0	0.7	
Rhys	1.99	0	2	0	4	
Toshiko	0	0	0	0	0	
mean overall		1.83	0.47	0.56	0.36	3.23
mean by series	<i>Angel</i>	1.37	0.56	0.13	0.13	2.19
	<i>Buffy the Vampire Slayer</i>	1.88	0.49	1.08	0.75	4.2
	<i>Gilmore Girls</i>	1.57	0.38	0.15	0.3	2.39
	<i>Parenthood</i>	2.92	0.77	0.67	0.32	4.68
	<i>Sherlock</i>	0.76	0	0	0	0.76
	<i>Torchwood</i>	0.81	0.19	0.49	0.15	1.64
mean by gender	female	2.2	0.79	0.89	0.52	4.4
	male	1.43	0.14	0.22	0.19	1.98
mean by age	young	2.92	0.85	1.18	0.64	5.59
	middle	1.4	0.22	0.32	0.18	2.11
	older	0.78	0.69	0	0.49	1.96
mean by nationality	American	2.19	0.61	0.65	0.47	3.92
	British	0.85	0.11	0.33	0.07	1.36

## Modal adverbs

character	series	probably	perhaps	maybe	stance markers overall	
Angel	<i>Angel</i>	7.01	0.33	20.4	27.7	
Cordelia		5.4	0	20.4	25.8	
Spike		4.39	0	10.2	14.6	
Angel	<i>Buffy the Vampire Slayer</i>	6.35	0	20	26.3	
Anya		5.28	0.44	19.8	25.5	
Buffy		6.95	0.16	21.6	28.7	
Cordelia		4.5	0	7.5	12	
Dawn		1.23	0	23.4	24.7	
Giles		2.47	14.64	3.5	20.6	
Spike		3.41	0.93	10.9	15.2	
Willow		6.45	0.17	27.9	34.6	
Xander		6.67	0.34	17.4	24.4	
Emily		2.53	2	6.1	10.7	
Lorelai		4.16	0.19	11.6	15.9	
Luke	<i>Gilmore Girls</i>	6.71	0	9.8	16.5	
Paris		2.5	0	8.2	10.7	
Richard		2.8	5.32	4.8	12.9	
Rory		5.78	0.38	15	21.1	
Sookie		2.87	0.24	12	15.1	
Adam		4.3	0	9.3	13.6	
Amber		6.64	0.95	12.8	20.4	
Camille	<i>Parenthood</i>	2.24	0	10.1	12.3	
Crosby		5.99	0	20.2	26.2	
Haddie		5.61	0	9.6	15.2	
Julia		5.08	0.56	17.5	23.1	
Kristina		5.78	0	11.6	17.4	
Max		11.6	0	3.6	15.2	
Sarah		3.14	0.45	11.7	15.3	
Zeek		1.53	0	9.2	10.7	
John		<i>Sherlock</i>	6.12	0	11.6	17.8
Sherlock			7.11	2.37	5.9	15.4
Gwen		<i>Torchwood</i>	0.49	2.43	10.2	13.1
Ianto	1.43		0	10	11.4	
Jack	1.27		0	12.7	13.9	
Owen	1.96		1.31	15	18.3	
Rhys	1.99		0	8	9.9	
Toshiko	3.19		0	8.5	11.7	
mean overall			4.40	0.90	12.65	17.94
mean by series	<i>Angel</i>		5.6	0.11	17.01	22.73
	<i>Buffy the Vampire Slayer</i>	4.81	1.85	16.9	23.56	
	<i>Gilmore Girls</i>	3.91	1.16	9.63	14.7	
	<i>Parenthood</i>	5.19	0.2	11.55	16.93	
	<i>Sherlock</i>	6.62	1.19	8.78	16.59	
	<i>Torchwood</i>	1.72	0.62	10.73	13.07	
mean by gender	female	4.2	0.42	13.98	18.6	
	male	4.62	1.4	11.24	17.26	
mean by age	young	5.72	0.2	15.62	21.54	
	middle	4.06	1.12	11.92	17.1	
	older	2.28	1.83	7.54	11.64	
mean by nationality	American	4.83	0.43	13.85	19.11	
	British	3.26	2.17	9.39	14.81	

## Intensifiers

Character	Series	very	really	so	completely	extremely	totally	total	
Angel	<i>Angel</i>	3.78	2.67	5.67	0	0.11	0.33	13.46	
Cordelia		3.86	5.98	10.99	0.2	0.19	0.19	24.1	
Spike		2.93	4.39	2.93	0	0	0	19.03	
Angel	<i>Buffy the Vampire Slayer</i>	5.44	6.35	5.44	0	0	0	20.87	
Anya		8.8	6.6	11.89	0.4	0.88	0.88	33.02	
Buffy		4.26	6.95	7.58	0.3	0.08	0.63	22.5	
Cordelia		4.5	10.5	20.25	0.7	0	3	40.49	
Dawn		4.93	9.87	17.88	0	0	1.23	35.77	
Giles		20.41	0.62	2.89	1.6	2.27	0	34.22	
Spike		7.14	2.48	7.45	0	0	0.31	24.82	
Willow		4.96	10.91	13.06	0	0	0.99	31.91	
Xander		6.84	4.96	8.55	0.5	0	0.68	24.1	
Emily		<i>Gilmore Girls</i>	21.99	0.53	15.46	1.3	0.93	0	42.64
Lorelai			12.92	9.39	13.27	0.9	0.38	1.17	40.28
Luke	6.48		7.77	5.18	1.1	0.12	0.24	22.37	
Paris	7.24		5.25	7.99	1.2	0.75	0.75	23.98	
Richard	32.21		1.12	5.88	0.6	1.68	0	44.81	
Rory	12.26		12.37	14.23	1.5	0.33	0.71	43.17	
Sookie	10.52		8.85	18.65	1.4	0.72	0.48	42.56	
Adam	<i>Parenthood</i>		9.28	9.45	6.01	0.5	0	1.2	28.35
Amber		16.12	20.38	32.24	0.9	0	2.37	73.48	
Camille		8.96	14.55	24.63	0	0	0	51.49	
Crosby		8.24	17.23	15.73	0.5	0	0.75	44.71	
Haddie		0.8	30.44	20.03	0.8	0	2.4	57.68	
Julia		8.46	16.92	31.03	1.7	0	2.26	64.88	
Kristina		10.52	20.51	36.02	1.3	0.26	1.84	72.04	
Max		11.6	9.81	9.81	0	0.89	1.78	34.8	
Sarah		14.36	19.74	28.71	0.4	0	0.67	65.05	
Zeek		10.71	3.57	4.08	0	0	0	19.89	
John		<i>Sherlock</i>	11.02	1.84	6.74	0	0	0.61	24.5
Sherlock	11.85		2.07	6.52	0.6	0.59	0.3	26.08	
Gwen	<i>Torchwood</i>	1.94	3.4	10.68	0.5	0	0	19.43	
Ianto		7.14	1.43	4.29	0	0	0	14.29	
Jack		3.38	2.53	3.8	0	0.42	0	10.13	
Owen		6.54	4.58	5.89	1.3	0	0	20.94	
Rhys		0	1.99	7.96	0	0	0	25.86	
Toshiko		4.25	1.06	8.51	1.1	0	1.06	17.02	
mean overall			8.83	8.08	12.38	0.58	0.29	0.73	33.91
mean by series	<i>Angel</i>	3.52	4.35	6.53	0.06	0.1	0.18	18.87	
	<i>Buffy</i>	7.48	6.58	10.55	0.41	0.36	0.86	29.74	
	<i>Gilmore Girls</i>	14.8	6.47	11.52	1.14	0.7	0.48	37.12	
	<i>Parenthood</i>	9.9	16.26	20.83	0.62	0.12	1.33	51.24	
	<i>Sherlock</i>	11.44	1.96	6.63	0.3	0.3	0.45	25.29	
	<i>Torchwood</i>	3.88	2.5	6.85	0.48	0.07	0.18	17.94	
mean by gender	female	8.51	11.27	18.06	0.78	0.24	1.09	42.18	
	male	9.17	4.72	6.38	0.37	0.34	0.34	25.18	
mean by age	young	7.18	11.17	14.54	0.56	0.26	1.3	37.08	
	middle	7.93	6.92	11.11	0.62	0.23	0.53	30.99	
	older	18.47	4.94	12.51	0.47	0.65	0	39.71	
mean by nationality	American	9.39	10.19	14.59	0.61	0.29	0.91	38.09	
	British	7.32	2.39	6.38	0.51	0.29	0.23	22.62	

## Appendix B: Frequencies, incl. raw counts

character	series	production	genre	gender	age	nationality	words
<b>Angel</b>	Angel	US	science fantasy	male	middle	American	<b>89905</b>
<b>Cordelia</b>	Angel	US	science fantasy	female	young	American	<b>51857</b>
<b>Spike</b>	Angel	US	science fantasy	male	middle	British	<b>13660</b>
<b>Xander</b>	Buffy	US	science fantasy	male	young	American	<b>58509</b>
<b>Dawn</b>	Buffy	US	science fantasy	female	young	American	<b>16215</b>
<b>Cordelia</b>	Buffy	US	science fantasy	female	young	American	<b>13335</b>
<b>Willow</b>	Buffy	US	science fantasy	female	young	American	<b>60492</b>
<b>Angel</b>	Buffy	US	science fantasy	male	middle	American	<b>11022</b>
<b>Buffy</b>	Buffy	US	science fantasy	female	young	American	<b>126682</b>
<b>Anya</b>	Buffy	US	science fantasy	female	young	American	<b>22716</b>
<b>Spike</b>	Buffy	US	science fantasy	male	middle	British	<b>32228</b>
<b>Giles</b>	Buffy	US	science fantasy	male	middle	British	<b>48505</b>
<b>Rory</b>	Gilmore Girls	US	dramedy	female	young	American	<b>183463</b>
<b>Sookie</b>	Gilmore Girls	US	dramedy	female	middle	American	<b>41820</b>
<b>Luke</b>	Gilmore Girls	US	dramedy	male	middle	American	<b>84922</b>
<b>Lorelai</b>	Gilmore Girls	US	dramedy	female	middle	American	<b>317313</b>
<b>Paris</b>	Gilmore Girls	US	dramedy	female	young	American	<b>40036</b>
<b>Emily</b>	Gilmore Girls	US	dramedy	female	older	American	<b>75047</b>
<b>Richard</b>	Gilmore Girls	US	dramedy	male	older	American	<b>35703</b>
<b>Julia</b>	Parenthood	US	dramedy	female	middle	American	<b>17726</b>
<b>Crosby</b>	Parenthood	US	dramedy	male	middle	American	<b>40039</b>
<b>Amber</b>	Parenthood	US	dramedy	female	young	American	<b>21094</b>
<b>Zeek</b>	Parenthood	US	dramedy	male	older	American	<b>19603</b>
<b>Haddie</b>	Parenthood	US	dramedy	female	young	American	<b>12483</b>
<b>Sarah</b>	Parenthood	US	dramedy	female	middle	American	<b>44579</b>
<b>Camille</b>	Parenthood	US	dramedy	female	older	American	<b>8933</b>
<b>Max</b>	Parenthood	US	dramedy	male	young	American	<b>11208</b>
<b>Adam</b>	Parenthood	US	dramedy	male	middle	American	<b>58206</b>
<b>Kristina</b>	Parenthood	US	dramedy	female	middle	American	<b>38037</b>
<b>Sherlock</b>	Sherlock	UK	crime	male	middle	British	<b>33742</b>
<b>John</b>	Sherlock	UK	crime	male	middle	British	<b>16328</b>
<b>Jack</b>	Torchwood	UK	science fantasy	male	middle	American	<b>23682</b>
<b>Owen</b>	Torchwood	UK	science fantasy	male	middle	British	<b>15285</b>
<b>Toshiko</b>	Torchwood	UK	science fantasy	female	middle	British	<b>9401</b>
<b>Rhys</b>	Torchwood	UK	science fantasy	male	middle	British	<b>5027</b>
<b>Ianto</b>	Torchwood	UK	science fantasy	male	middle	British	<b>7000</b>
<b>Gwen</b>	Torchwood	UK	science fantasy	female	middle	British	<b>20590</b>

	PM like R	Quot like R	Approx like R	All like	Dvalue like	PM like E	PM I mean R	All I mean	Dvalue I mean	PM I mean E	PM you know	All you know	Dvalue you know	PM you know E
Angel	9	1	1	354	3.1	1.0	114	123	92.7	12.7	176	327	53.8	19.6
Cordelia	15	1	13	315	9.2	2.9	65	82	79.3	12.5	54	130	41.5	10.4
Spike	2	2	2	82	7.3	1.5	7	10	70	5.1	4	20	20	2.9
Xander	22	4	6	111	28.8	3.8	111	122	91	19.0	84	161	52.2	14.4
Dawn	13	4	3	128	15.6	8.0	33	35	94.3	20.4	30	55	54.5	18.5
Cordelia	11	2	4	93	18.3	8.2	23	28	82.1	17.2	17	34	50	12.7
Willow	29	8	9	184	25	4.8	184	200	92	30.4	113	191	59.2	18.7
Angel	0	0	0	61	0	0.0	13	18	72.2	11.8	19	36	52.8	17.2
Buffy	34	11	13	674	8.6	2.7	199	222	89.6	15.7	294	514	57.2	23.2
Anya	4	2	7	132	9.8	1.8	82	87	94.3	36.1	55	81	67.9	24.2
Spike	0	2	0	166	1.2	0.0	20	27	74.1	6.2	54	99	54.5	16.8
Giles	0	0	0	117	0	0.0	41	42	97.6	8.5	33	67	49.3	6.8
Rory	24	12	37	812	9	1.3	305	333	91.6	16.6	181	381	47.5	9.9
Sookie	3	5	7	242	6.2	0.7	81	87	93.1	19.4	62	131	47.3	14.8
Luke	4	4	3	373	2.9	0.5	155	177	87.6	18.3	182	333	54.7	21.4
Lorelai	38	30	42	1415	7.8	1.2	390	454	85.9	12.3	487	951	51.2	15.3
Paris	2	1	3	139	4.3	0.5	52	62	83.9	13.0	32	75	42.7	8.0
Emily	0	0	0	362	0	0.0	38	45	84.4	5.1	44	131	33.6	5.9
Richard	0	0	0	158	0	0.0	4	6	66.7	1.1	20	61	32.8	5.6
Julia	8	0	4	70	17.1	4.5	26	31	83.9	14.7	85	134	63.4	48.0
Crosby	46	5	19	276	25.4	11.5	84	86	97.7	21.0	207	310	66.8	51.7
Amber	101	1	2	204	51	47.9	58	88	65.9	27.5	116	150	77.3	55.0
Zeek	10	1	3	86	16.3	5.1	115	120	95.8	58.7	112	153	73.2	57.1
Haddie	97	0	3	166	60.2	77.7	37	41	90.2	29.6	25	51	49	20.0
Sarah	64	4	13	288	28.1	14.4	132	135	97.8	29.6	330	418	78.9	74.0
Camille	2	0	2	38	10.5	2.2	13	13	100	14.6	79	97	81.4	88.4
Max	5	0	2	66	10.6	4.5	5	7	71.4	4.5	16	32	50	14.3
Adam	15	5	11	280	11.1	2.6	60	69	87	10.3	207	375	55.2	35.6
Kristina	54	6	10	258	27.1	14.2	91	93	97.8	23.9	118	196	60.2	31.0
Sherlock	0	0	0	66	0	0.0	2	10	20	0.6	15	47	31.9	4.4
John	0	0	0	35	0	0.0	9	13	69.2	5.5	14	38	36.8	8.6
Jack	0	0	0	88	0	0.0	6	7	85.7	2.5	8	33	24.2	3.4
Owen	2	0	1	51	5.9	1.3	7	8	87.5	4.6	26	42	61.9	17.0
Toshiko	1	1	0	39	5.1	1.1	3	6	50	3.2	1	14	7.1	1.1
Rhys	2	1	0	23	13	4.0	5	5	100	9.9	11	16	68.8	21.9
Ianto	0	0	0	35	0	0.0	2	2	100	2.9	3	7	42.9	4.3
Gwen	2	0	2	78	5.1	1.0	16	17	94.1	7.8	8	29	27.6	3.9

	<i>sort of R</i>	<i>kind of R</i>	<i>sort of E</i>	<i>kind of E</i>	<i>all sort of</i>	<i>all kind of</i>	<i>Dvalue sort of</i>	<i>Dvalue kind of</i>
Angel	14	80	1.6	8.9	15	109	93.3	73.4
Cordelia	8	52	1.5	10.0	12	69	66.7	75.4
Spike	4	2	2.9	1.5	7	4	57.1	50.0
Xander	17	72	2.9	12.3	19	102	89.5	70.6
Dawn	4	19	2.5	11.7	5	21	80.0	90.5
Cordelia	5	7	3.7	5.2	6	12	83.3	58.3
Willow	34	83	5.6	13.7	39	106	87.2	78.3
Angel	2	11	1.8	10.0	2	13	100.0	84.6
Buffy	45	134	3.6	10.6	49	175	91.8	76.6
Anya	6	12	2.6	5.3	7	20	85.7	60.0
Spike	3	7	0.9	2.2	7	19	42.9	36.8
Giles	15	5	3.1	1.0	32	12	46.9	41.7
Rory	18	139	1.0	7.6	22	184	81.8	75.5
Sookie	9	32	2.2	7.7	9	47	100.0	68.1
Luke	15	47	1.8	5.5	16	59	93.8	79.7
Lorelai	46	133	1.4	4.2	57	236	80.7	56.4
Paris	7	10	1.7	2.5	9	18	77.8	55.6
Emily	12	5	1.6	0.7	20	26	60.0	19.2
Richard	9	2	2.5	0.6	16	6	56.3	33.3
Julia	0	24	0.0	13.5	0	36	0.0	66.7
Crosby	0	50	0.0	12.5	0	62	0.0	80.6
Amber	2	24	0.9	11.4	3	27	66.7	88.9
Zeek	0	19	0.0	9.7	1	24	0.0	79.2
Haddie	0	12	0.0	9.6	0	16	0.0	75.0
Sarah	10	26	2.2	5.8	10	40	100.0	65.0
Camille	1	4	1.1	4.5	1	6	100.0	66.7
Max	0	2	0.0	1.8	0	6	0.0	33.3
Adam	6	16	1.0	2.7	7	26	85.7	61.5
Kristina	25	26	6.6	6.8	26	28	96.2	92.9
Sherlock	4	5	1.2	1.5	7	11	57.1	45.5
John	8	3	4.9	1.8	15	6	53.3	50.0
Jack	7	19	3.0	8.0	9	22	77.8	86.4
Owen	4	7	2.6	4.6	6	11	66.7	63.6
Toshiko	2	3	2.1	3.2	2	3	100.0	100.0
Rhys	1	1	2.0	2.0	4	2	25.0	50.0
Ianto	4	2	5.7	2.9	5	3	80.0	66.7
Gwen	12	3	5.8	1.5	20	3	60.0	100.0

	or something R	or anything R	and stuff R	and everything R	or something E	or anything E	and stuff E	and everything E
Angel	10	3	0	0	1.1	0.3	0.0	0.0
Cordelia	8	7	2	2	1.5	1.3	0.4	0.4
Spike	2	0	0	0	1.5	0.0	0.0	0.0
Xander	8	2	2	3	1.4	0.3	0.3	0.5
Dawn	5	2	6	1	3.1	1.2	3.7	0.6
Cordelia	3	2	3	4	2.2	1.5	2.2	3.0
Willow	20	3	15	9	3.3	0.5	2.5	1.5
Angel	1	0	0	0	0.9	0.0	0.0	0.0
Buffy	24	5	8	6	1.9	0.4	0.6	0.5
Anya	8	1	0	1	3.5	0.4	0.0	0.4
Spike	0	0	1	0	0.0	0.0	0.3	0.0
Giles	3	0	0	1	0.6	0.0	0.0	0.2
Rory	37	10	4	14	2.0	0.5	0.2	0.8
Sookie	6	1	2	0	1.4	0.2	0.5	0.0
Luke	25	4	1	4	2.9	0.5	0.1	0.5
Lorelai	71	20	8	14	2.2	0.6	0.3	0.4
Paris	5	3	0	0	1.2	0.7	0.0	0.0
Emily	4	0	0	1	0.5	0.0	0.0	0.1
Richard	2	0	0	1	0.6	0.0	0.0	0.3
Julia	4	1	0	1	2.3	0.6	0.0	0.6
Crosby	20	1	5	0	5.0	0.2	1.2	0.0
Amber	14	0	7	0	6.6	0.0	3.3	0.0
Zeek	4	1	0	3	2.0	0.5	0.0	1.5
Haddie	8	4	1	0	6.4	3.2	0.8	0.0
Sarah	11	3	6	5	2.5	0.7	1.3	1.1
Camille	0	2	0	0	0.0	2.2	0.0	0.0
Max	2	0	0	0	1.8	0.0	0.0	0.0
Adam	9	0	0	0	1.5	0.0	0.0	0.0
Kristina	4	1	0	0	1.1	0.3	0.0	0.0
Sherlock	1	0	0	0	0.3	0.0	0.0	0.0
John	2	0	0	0	1.2	0.0	0.0	0.0
Jack	0	0	0	1	0.0	0.0	0.0	0.4
Owen	0	1	0	0	0.0	0.7	0.0	0.0
Toshiko	0	0	0	0	0.0	0.0	0.0	0.0
Rhys	1	0	1	0	2.0	0.0	2.0	0.0
Ianto	2	0	0	0	2.9	0.0	0.0	0.0
Gwen	0	1	2	1	0.0	0.5	1.0	0.5

	probably R	perhaps R	maybe R	probably E	perhaps E	maybe E
Angel	63	3	183	7.0	0.3	20.4
Cordelia	28	0	106	5.4	0.0	20.4
Spike	6	0	14	4.4	0.0	10.2
Xander	39	2	102	6.7	0.3	17.4
Dawn	2	0	38	1.2	0.0	23.4
Cordelia	6	0	10	4.5	0.0	7.5
Willow	39	1	169	6.4	0.2	27.9
Angel	7	0	22	6.4	0.0	20.0
Buffy	88	2	274	6.9	0.2	21.6
Anya	12	1	45	5.3	0.4	19.8
Spike	11	3	35	3.4	0.9	10.9
Giles	12	71	17	2.5	14.6	3.5
Rory	106	7	275	5.8	0.4	15.0
Sookie	12	1	50	2.9	0.2	12.0
Luke	57	0	83	6.7	0.0	9.8
Lorelai	132	6	367	4.2	0.2	11.6
Paris	10	0	33	2.5	0.0	8.2
Emily	19	15	46	2.5	2.0	6.1
Richard	10	19	17	2.8	5.3	4.8
Julia	9	1	31	5.1	0.6	17.5
Crosby	24	0	81	6.0	0.0	20.2
Amber	14	2	27	6.6	0.9	12.8
Zeek	3	0	18	1.5	0.0	9.2
Haddie	7	0	12	5.6	0.0	9.6
Sarah	14	2	52	3.1	0.4	11.7
Camille	2	0	9	2.2	0.0	10.1
Max	13	0	4	11.6	0.0	3.6
Adam	25	0	54	4.3	0.0	9.3
Kristina	22	0	44	5.8	0.0	11.6
Sherlock	24	8	20	7.1	2.4	5.9
John	10	0	19	6.1	0.0	11.6
Jack	3	0	30	1.3	0.0	12.7
Owen	3	2	23	2.0	1.3	15.0
Toshiko	3	0	8	3.2	0.0	8.5
Rhys	1	0	4	2.0	0.0	8.0
Ianto	1	0	7	1.4	0.0	10.0
Gwen	1	5	21	0.5	2.4	10.2

	<i>completely</i>	<i>extremely</i>	<i>really</i>	<i>so</i>	<i>totally</i>	<i>Very</i>	<i>other</i>	<i>total</i>	<i>completelyE</i>	<i>extremelyE</i>	<i>reallyE</i>	<i>soE</i>	<i>totallyE</i>	<i>veryE</i>	<i>other E</i>	<i>total E</i>
Angel	0	1	24	51	3	34	8	121	0.0	0.1	2.7	5.7	0.3	3.8	0.9	13.5
Cordelia	1	1	31	57	1	20	14	125	0.2	0.2	6.0	11.0	0.2	3.9	2.7	24.1
Spike	0	0	6	4	0	4	12	6	0.0	0.0	4.4	2.9	0.0	2.9	8.8	4.4
Xander	3	0	29	50	4	40	15	141	0.5	0.0	5.0	8.5	0.7	6.8	2.6	24.1
Dawn	0	0	16	29	2	8	3	58	0.0	0.0	9.9	17.9	1.2	4.9	1.9	35.8
Cordelia	1	0	14	27	4	6	2	54	0.7	0.0	10.5	20.2	3.0	4.5	1.5	40.5
Willow	0	0	66	79	6	30	12	193	0.0	0.0	10.9	13.1	1.0	5.0	2.0	31.9
Angel	0	0	7	6	0	6	4	23	0.0	0.0	6.4	5.4	0.0	5.4	3.6	20.9
Buffy	4	1	88	96	8	54	34	285	0.3	0.1	6.9	7.6	0.6	4.3	2.7	22.5
Anya	1	2	15	27	2	20	8	75	0.4	0.9	6.6	11.9	0.9	8.8	3.5	33.0
Spike	0	0	8	24	1	23	21	80	0.0	0.0	2.5	7.4	0.3	7.1	6.5	24.8
Giles	8	11	3	14	0	99	31	166	1.6	2.3	0.6	2.9	0.0	20.4	6.4	34.2
Rory	28	6	227	261	13	225	32	792	1.5	0.3	12.4	14.2	0.7	12.3	1.7	43.2
Sookie	6	3	37	78	2	44	8	178	1.4	0.7	8.8	18.7	0.5	10.5	1.9	42.6
Luke	9	1	66	44	2	55	13	190	1.1	0.1	7.8	5.2	0.2	6.5	1.5	22.4
Lorelai	27	12	298	421	37	410	73	1278	0.9	0.4	9.4	13.3	1.2	12.9	2.3	40.3
Paris	5	3	21	32	3	29	3	96	1.2	0.7	5.2	8.0	0.7	7.2	0.7	24.0
Emily	10	7	4	116	0	165	18	320	1.3	0.9	0.5	15.5	0.0	22.0	2.4	42.6
Richard	2	6	4	21	0	115	12	160	0.6	1.7	1.1	5.9	0.0	32.2	3.4	44.8
Julia	3	0	30	55	4	15	8	115	1.7	0.0	16.9	31.0	2.3	8.5	4.5	64.9
Crosby	2	0	69	63	3	33	9	179	0.5	0.0	17.2	15.7	0.7	8.2	2.2	44.7
Amber	2	0	43	68	5	34	4	155	0.9	0.0	20.4	32.2	2.4	16.1	1.9	73.5
Zeek	0	0	7	8	0	21	3	39	0.0	0.0	3.6	4.1	0.0	10.7	1.5	19.9
Haddie	1	0	38	25	3	1	4	72	0.8	0.0	30.4	20.0	2.4	0.8	3.2	57.7
Sarah	2	0	88	128	3	64	5	290	0.4	0.0	19.7	28.7	0.7	14.4	1.1	65.1
Camille	0	0	13	22	0	8	3	46	0.0	0.0	14.6	24.6	0.0	9.0	3.4	51.5
Max	0	1	11	11	2	13	1	39	0.0	0.9	9.8	9.8	1.8	11.6	0.9	34.8
Adam	3	0	55	35	7	54	11	165	0.5	0.0	9.4	6.0	1.2	9.3	1.9	28.3
Kristina	5	1	78	137	7	40	6	274	1.3	0.3	20.5	36.0	1.8	10.5	1.6	72.0
Sherlock	2	2	7	22	1	40	14	88	0.6	0.6	2.1	6.5	0.3	11.9	4.1	26.1
John	0	0	3	11	1	18	7	40	0.0	0.0	1.8	6.7	0.6	11.0	4.3	24.5
Jack	0	1	6	9	0	8	0	24	0.0	0.4	2.5	3.8	0.0	3.4	0.0	10.1
Owen	2	0	7	9	0	10	4	32	1.3	0.0	4.6	5.9	0.0	6.5	2.6	20.9
Toshiko	1	0	1	8	1	4	1	16	1.1	0.0	1.1	8.5	1.1	4.3	1.1	17.0
Rhys	0	0	1	4	0	0	8	13	0.0	0.0	2.0	8.0	0.0	0.0	15.9	25.9
Ianto	0	0	1	3	0	5	1	10	0.0	0.0	1.4	4.3	0.0	7.1	1.4	14.3
Gwen	1	0	7	22	0	4	6	40	0.5	0.0	3.4	10.7	0.0	1.9	2.9	19.4

## Appendix C: Significance tests

feature group	feature	gender	nationality	age (young - middle)	age (middle - older)
<b>pragmatic marker</b>	<i>you know</i>	/	< .01	/	/
	<i>I mean</i>	< .01	< .001	< .01	/
	<i>like</i>	< .05	< .001	< .01	/
<b>Hedges</b>	<i>kind of</i>	/	< .001	< .05	/
	<i>sort of</i>	/	/	/	/
<b>modal adverb</b>	<i>probably</i>	/	/	/	/
	<i>perhaps</i>	/	/	/	/
	<i>possibly</i>	/	/	/	/
	<i>maybe</i>	/	/ (.056)	/	< .05
<b>general extender</b>	<i>or something</i>	/	< .05	< .01	/
	<i>or anything</i>	< .001	< .05	< .01	/
	<i>and stuff</i>	< .05	/	< .05	/
	<i>and everything</i>	< .05	< .05	< .05	/
<b>intensifier</b>	<i>very</i>	/	/	/	< .05
	<i>really</i>	< .01	< .001	< .05	/
	<i>so</i>	< .001	< .01	< .05	/
	<i>completely</i>	< .05	/	/	/
	<i>extremely</i>	/	/	/	/
	<i>totally</i>	< .01	< .05	< .01	< .05

## Appendix D: Scripts for distribution plots

### General distribution

```
> NAME <- ggplot (DATAFRAME, aes (FACTOR, DATAFRAME$`FEATURE`))
+ geom_boxplot (aes (FACTOR, DATAFRAME$`FEATURE`))
+ theme_bw()
+ labs (x=" ", y=" ")
+ ylim (0, LIMIT FOR Y-AXIS)
+ theme (text = element_text (family = "Times", face = "bold", size = 12)
# add for age plots:
+ xlim ("older", "middle", "young")
```

### Collate into plot cluster

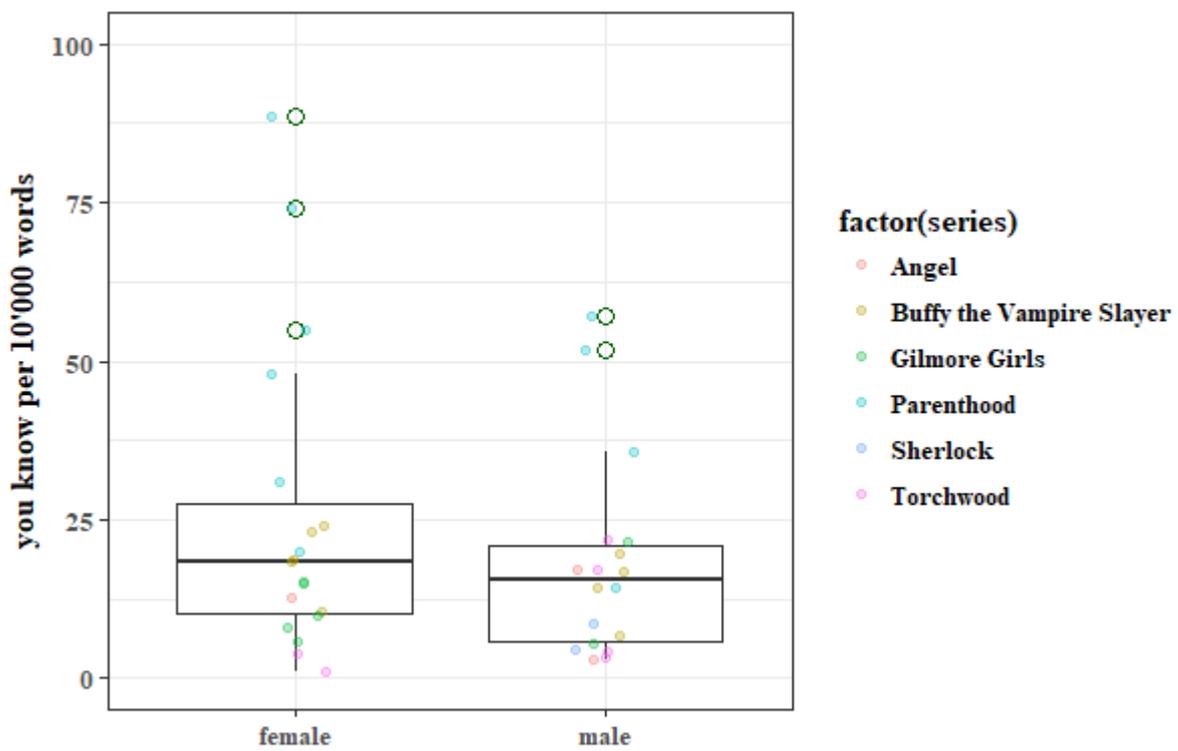
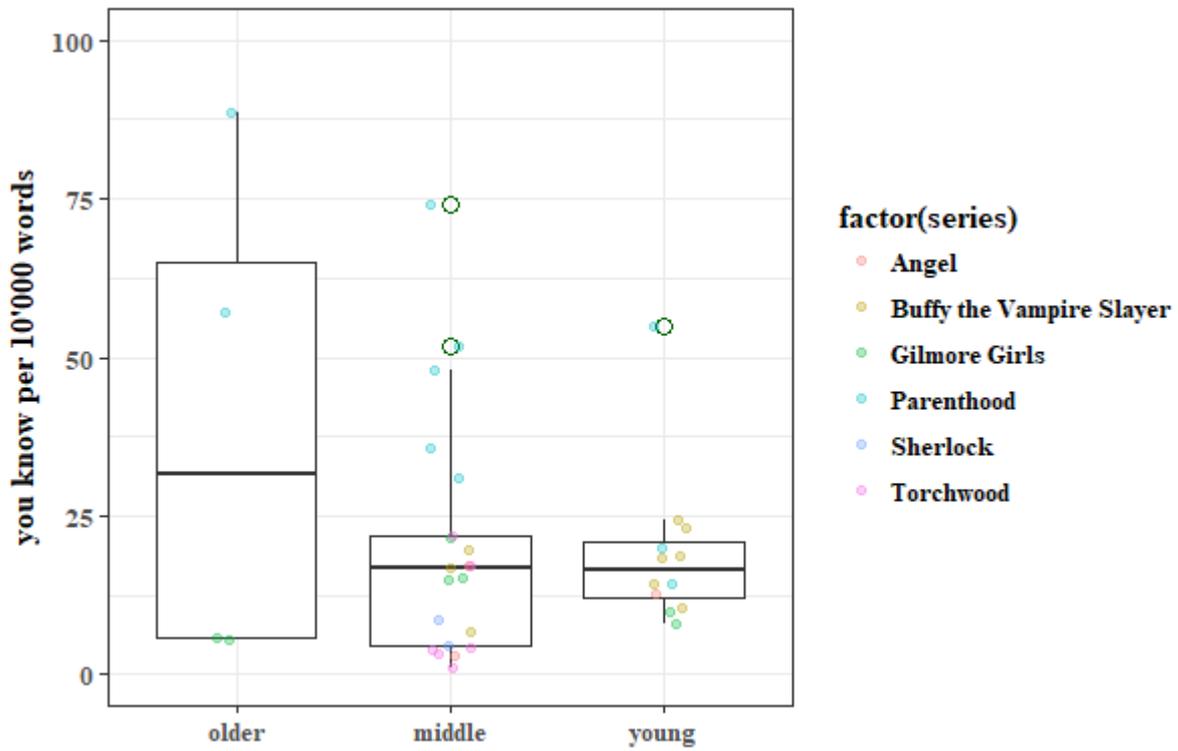
```
> NAME <- grid.arrange (NAME1, NAME2, ..., ncol = 2)
```

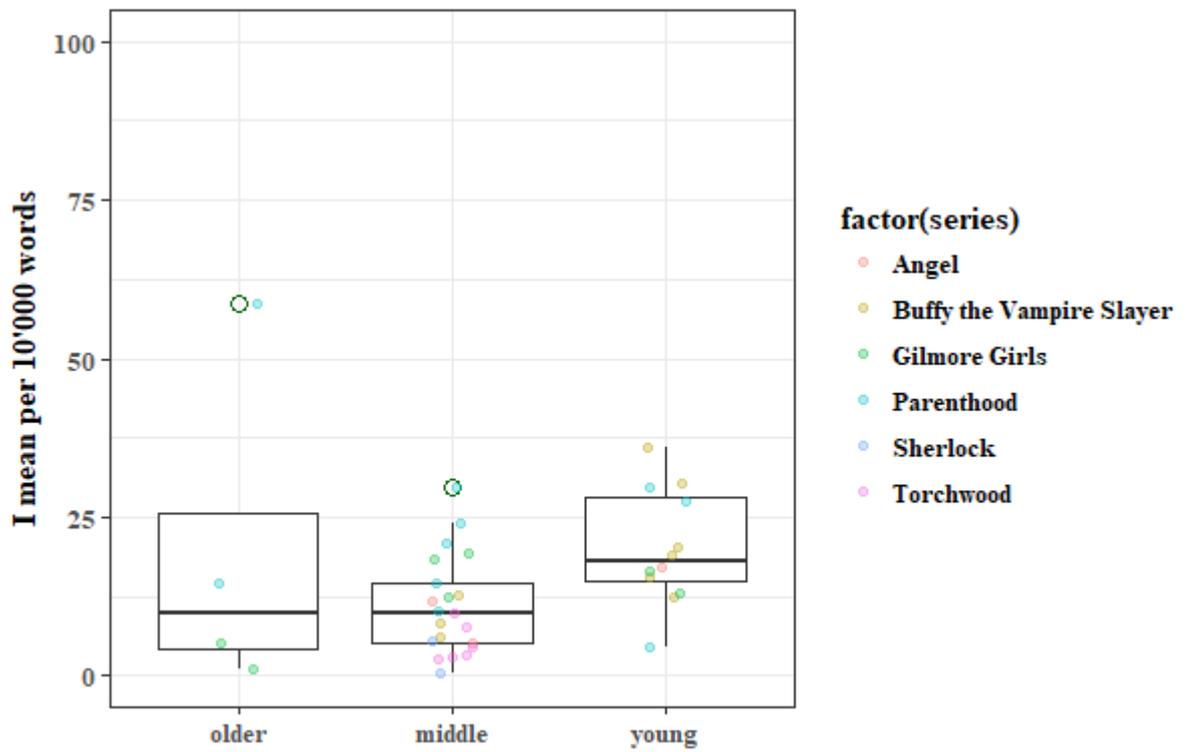
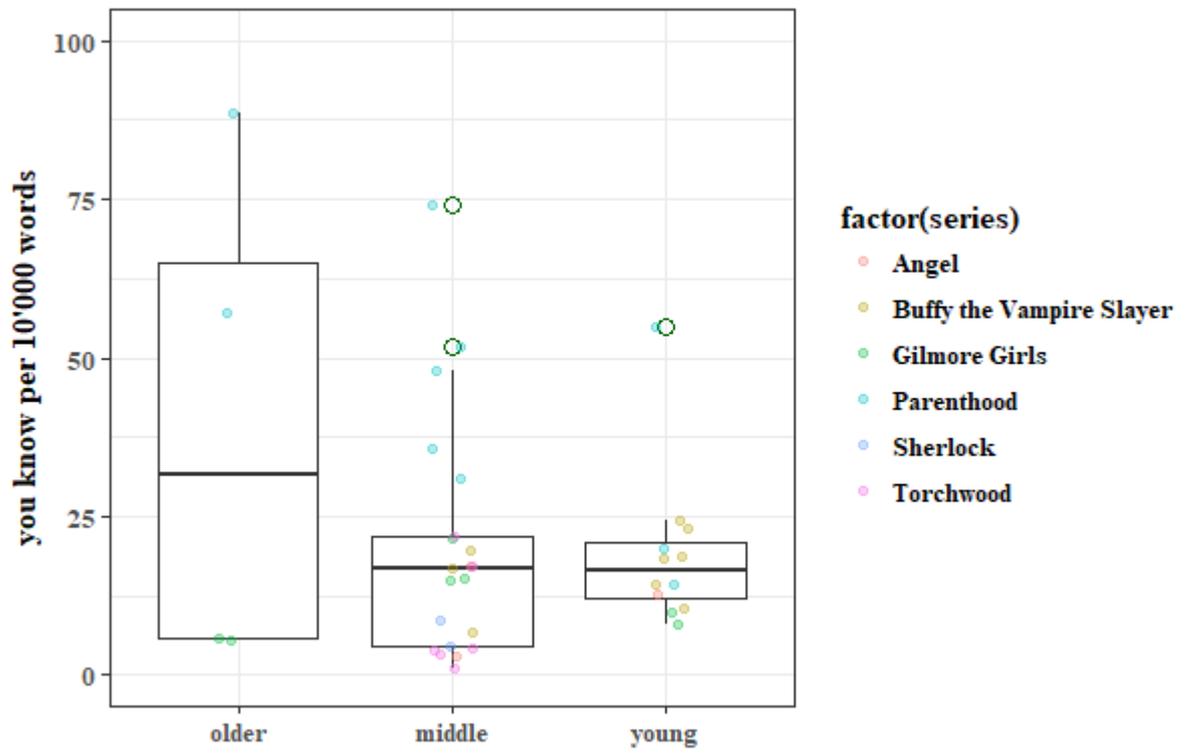
### Bar plots for character-specific feature uses

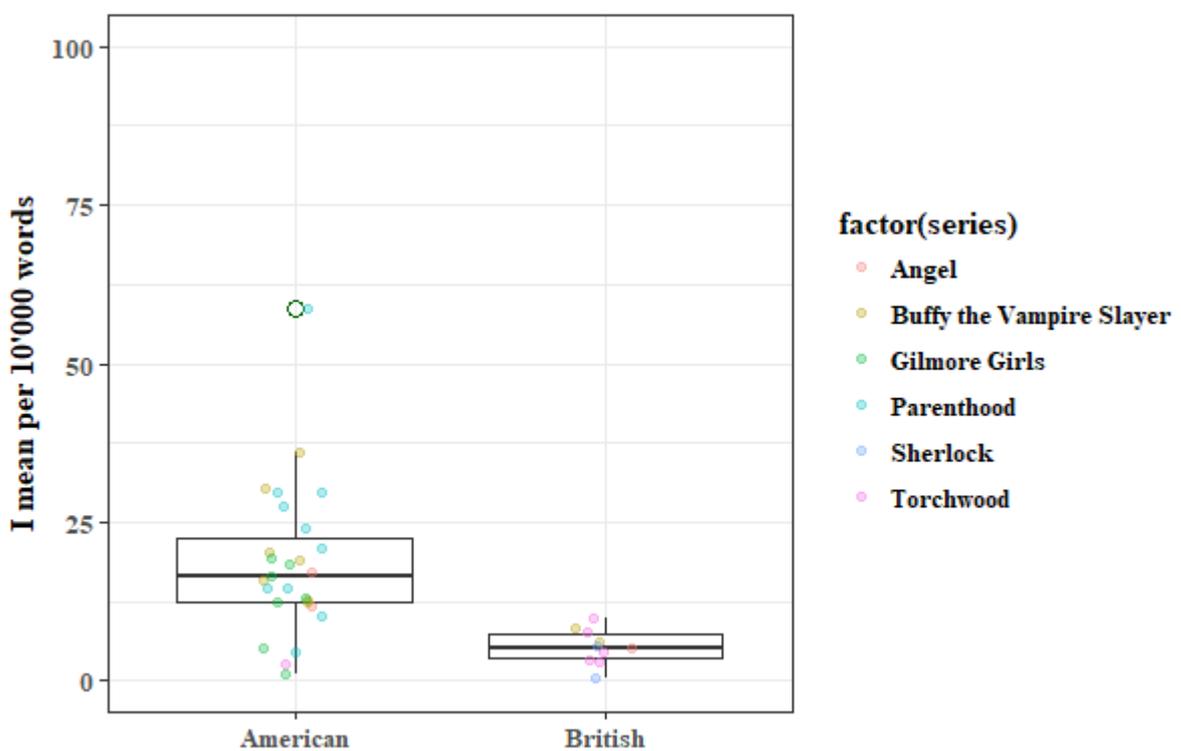
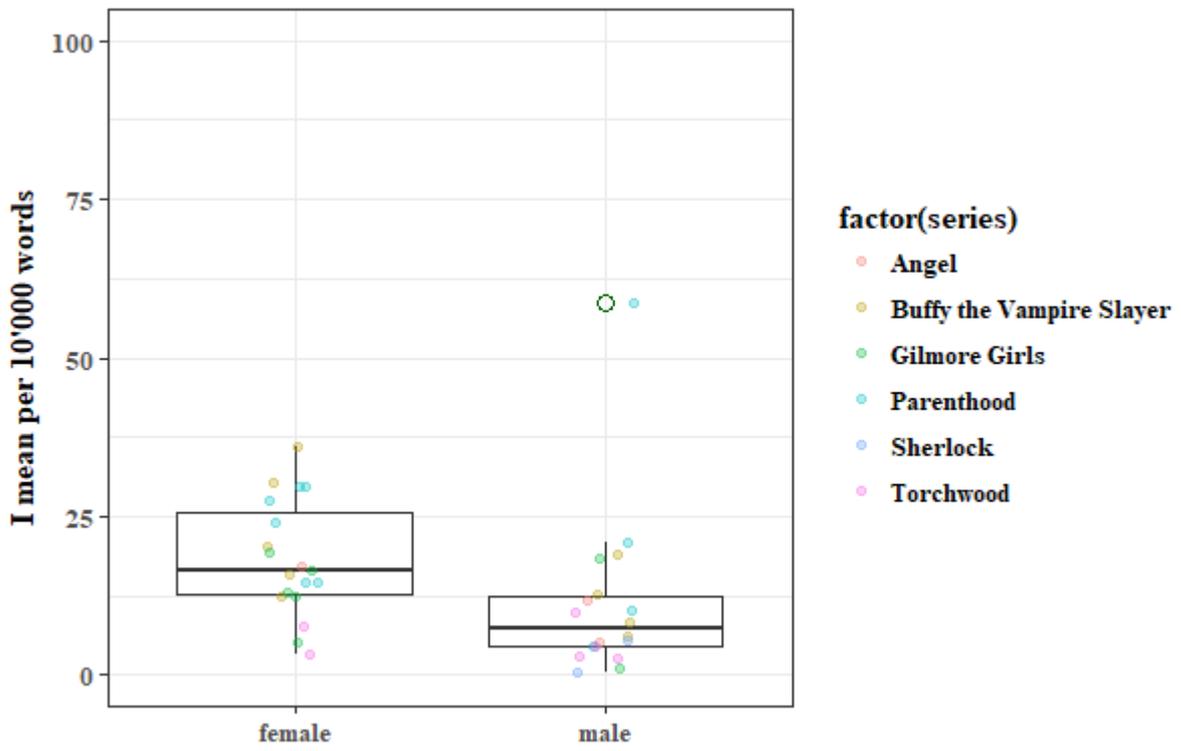
```
> NAME <- ggplot (DATAFRAME, aes (x=reorder (FACTOR,
-DATAFRAME$`FEATURE`), y=DATAFRAME$`FEATURE`))
+ geom_bar (aes (fill=FACTOR), stat="identity", colour="black", position=
position_dodge ())
+ coord_flip ()
+ ylab (" ")
+ xlab (" ")
+ theme_bw()
+ scale_fill_manual (values= c ("#fffcc", "#c7e9b4", "#7fcdbb",
"#41b6c4", "#2c7fb8", "#253494"))
+theme (text=element_text (family="Times", face="bold", size=12))
```

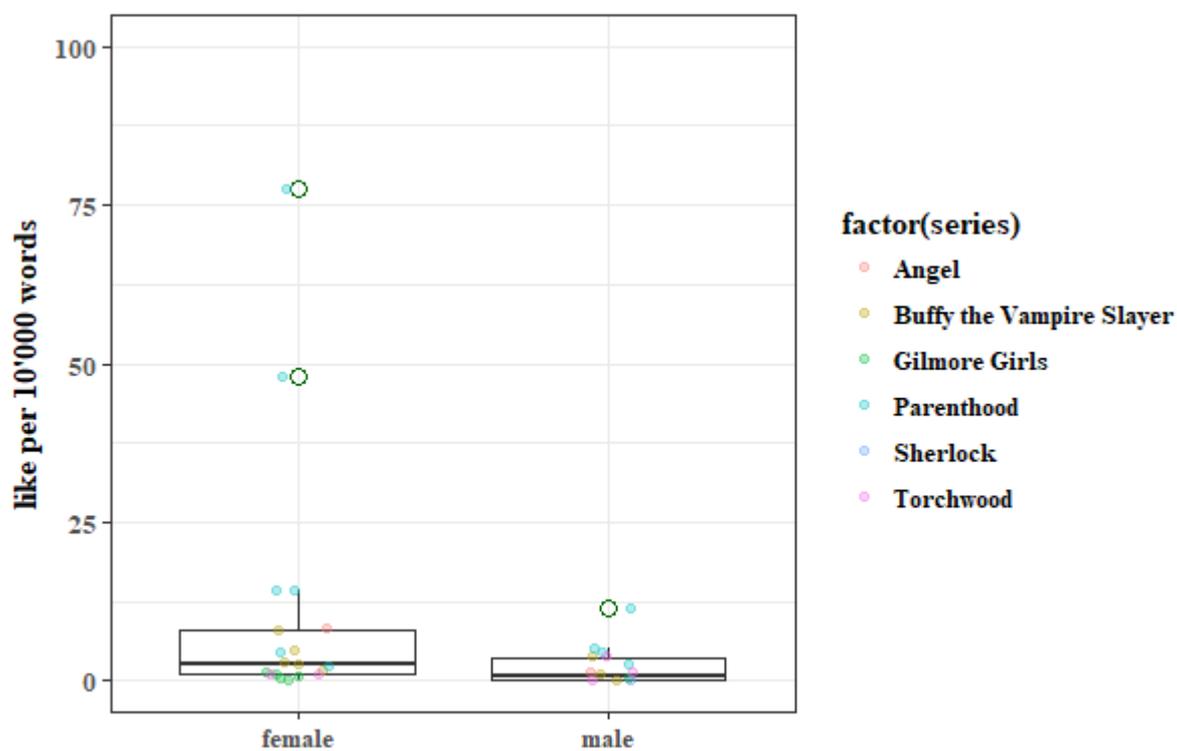
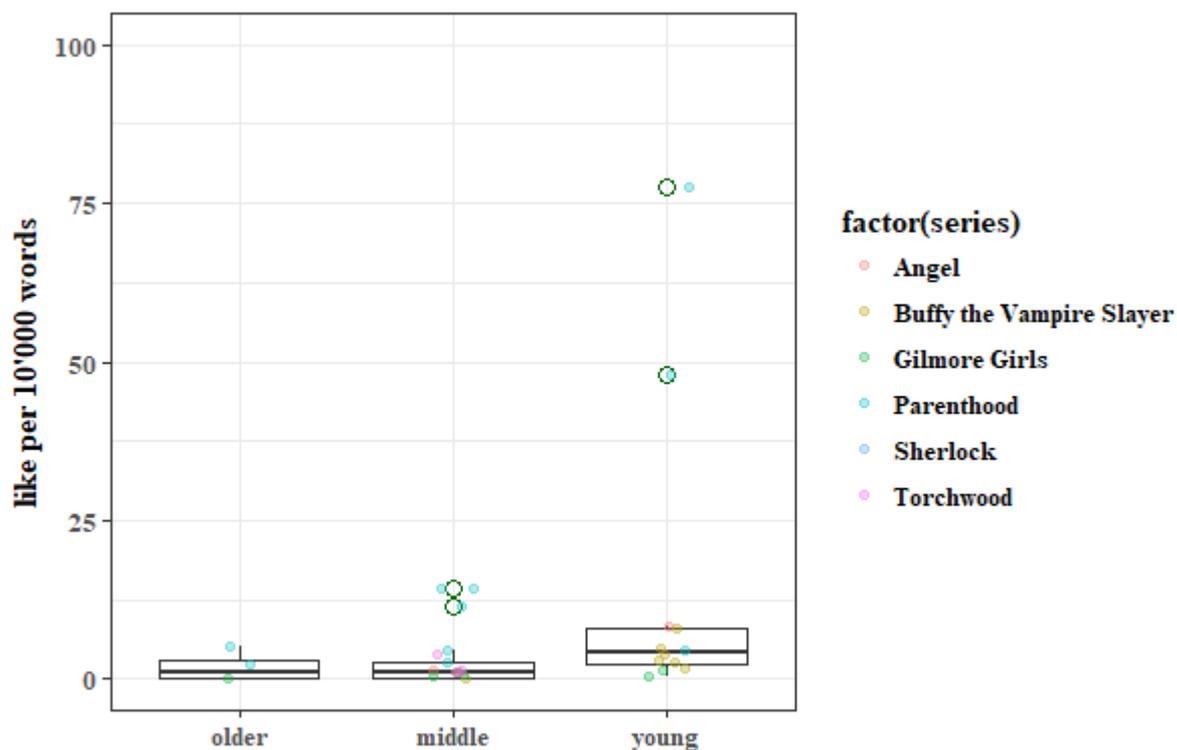
## Appendix E: Distribution plots

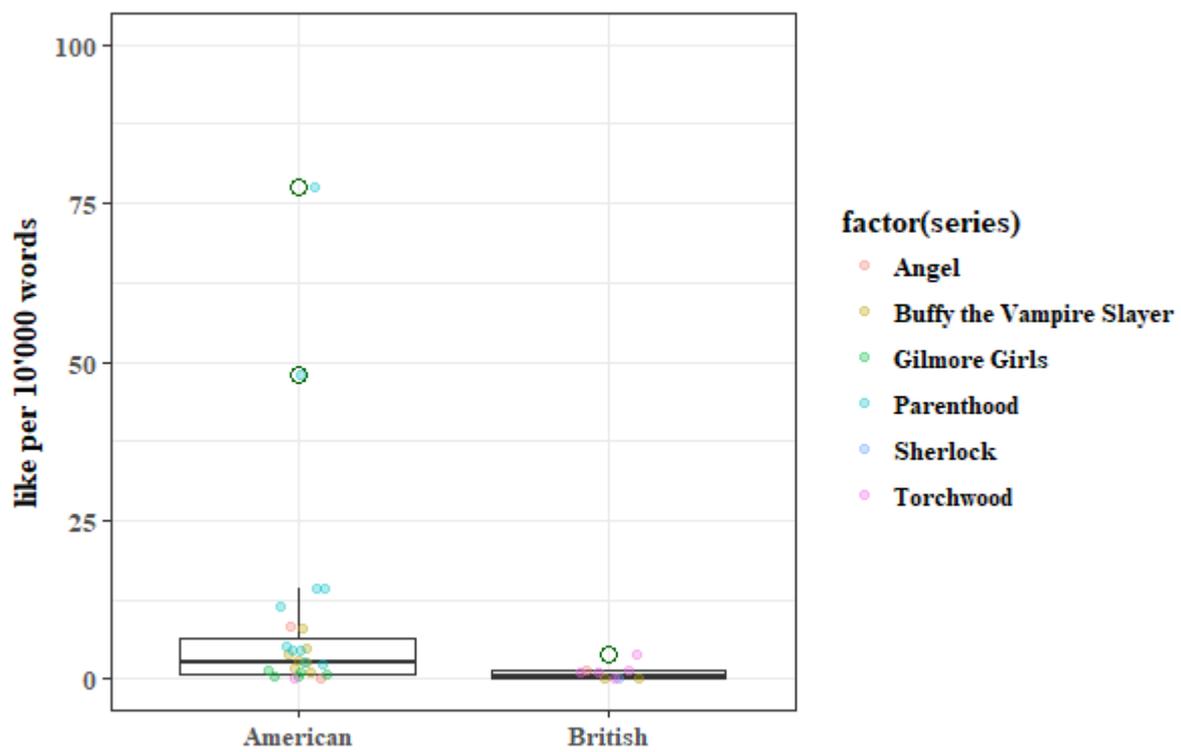
### Pragmatic markers



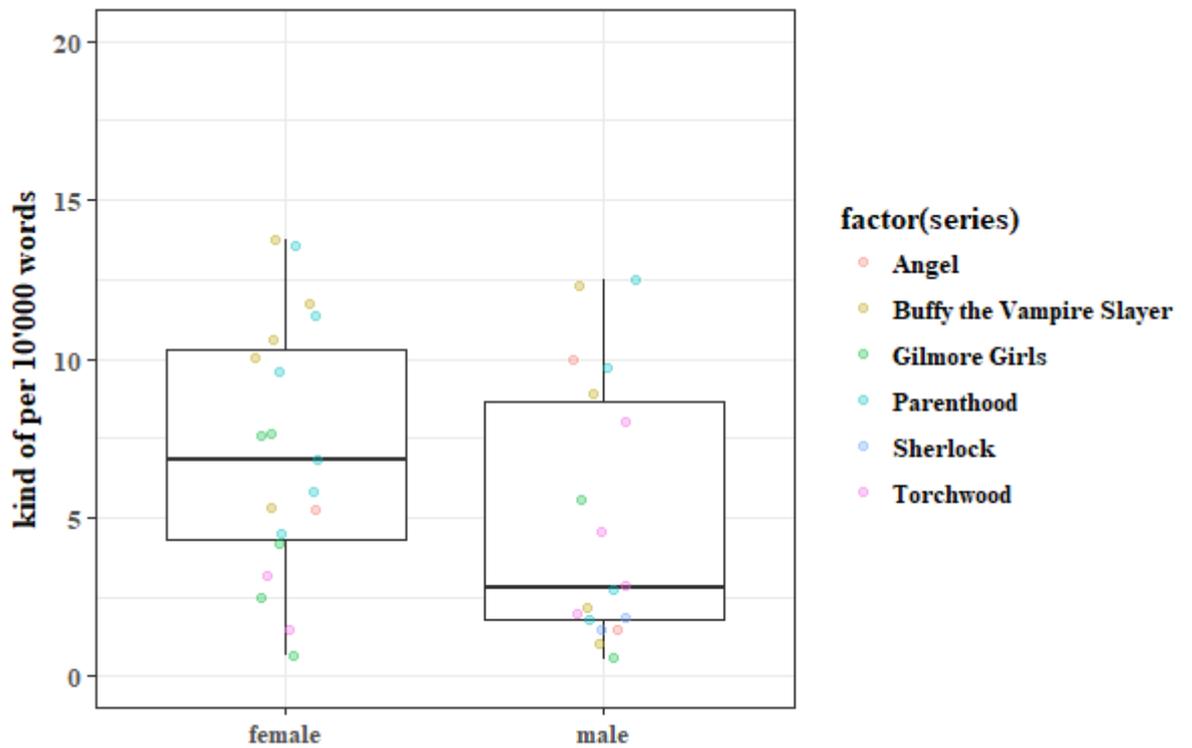
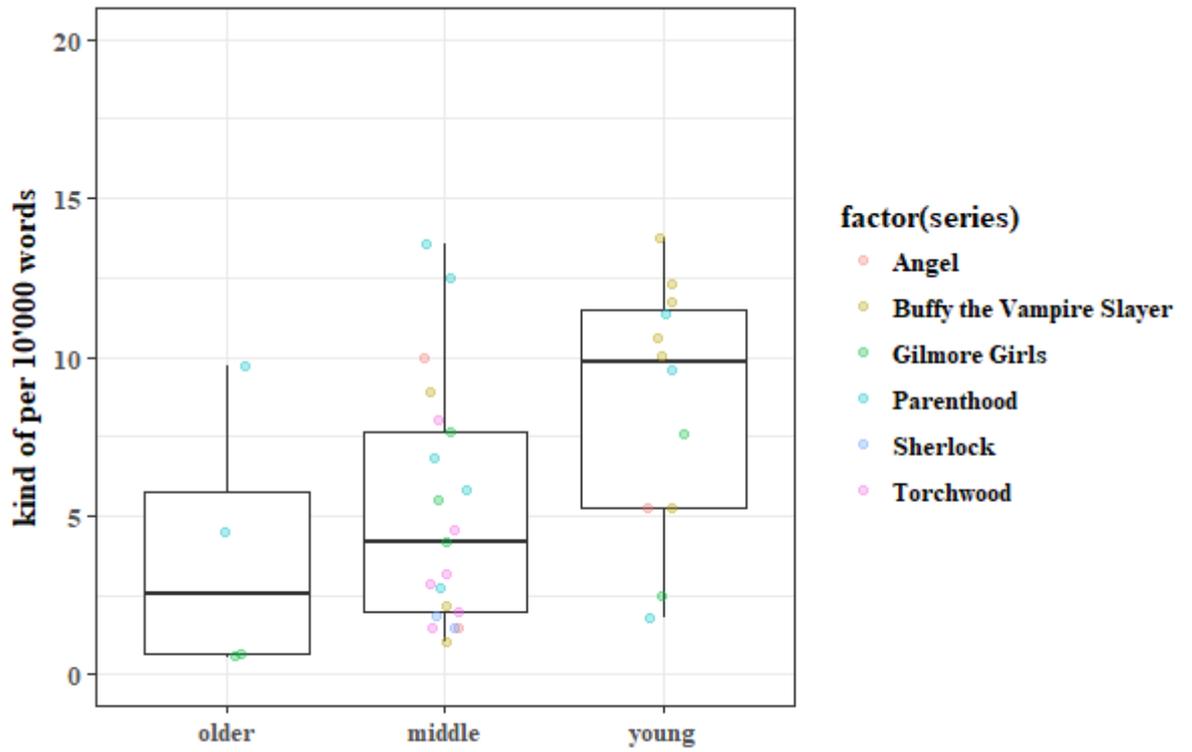


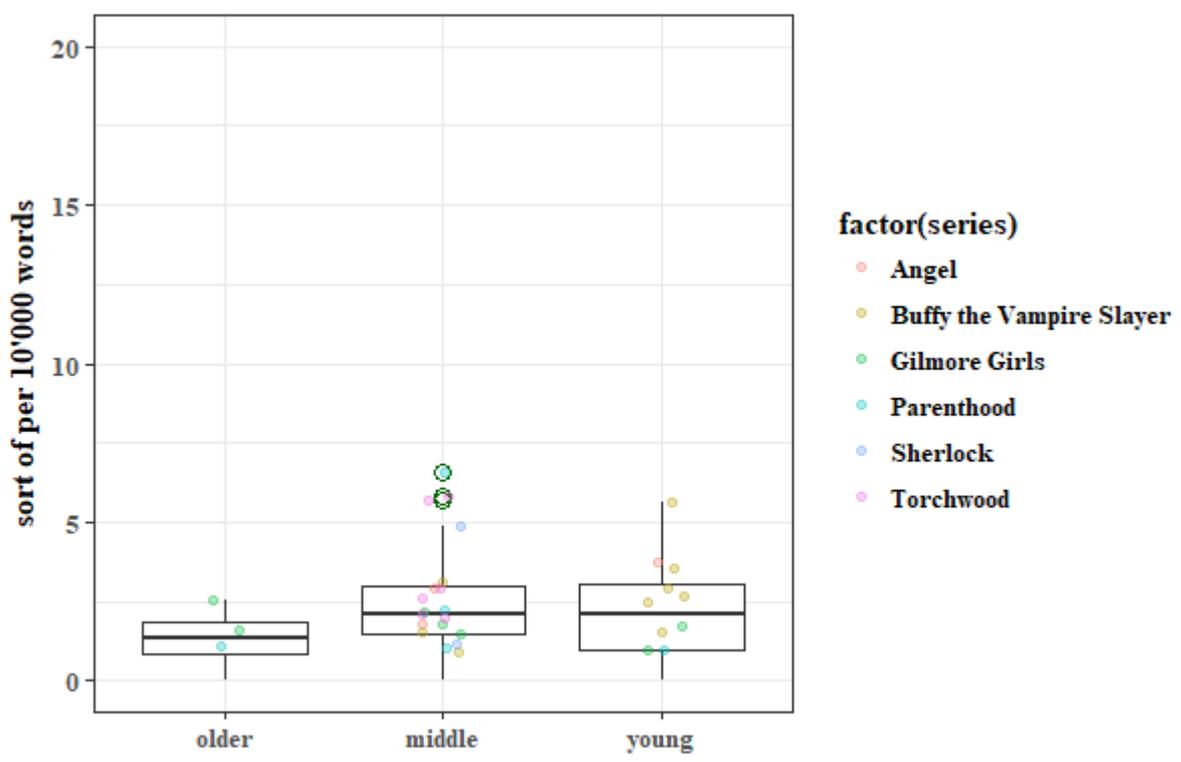
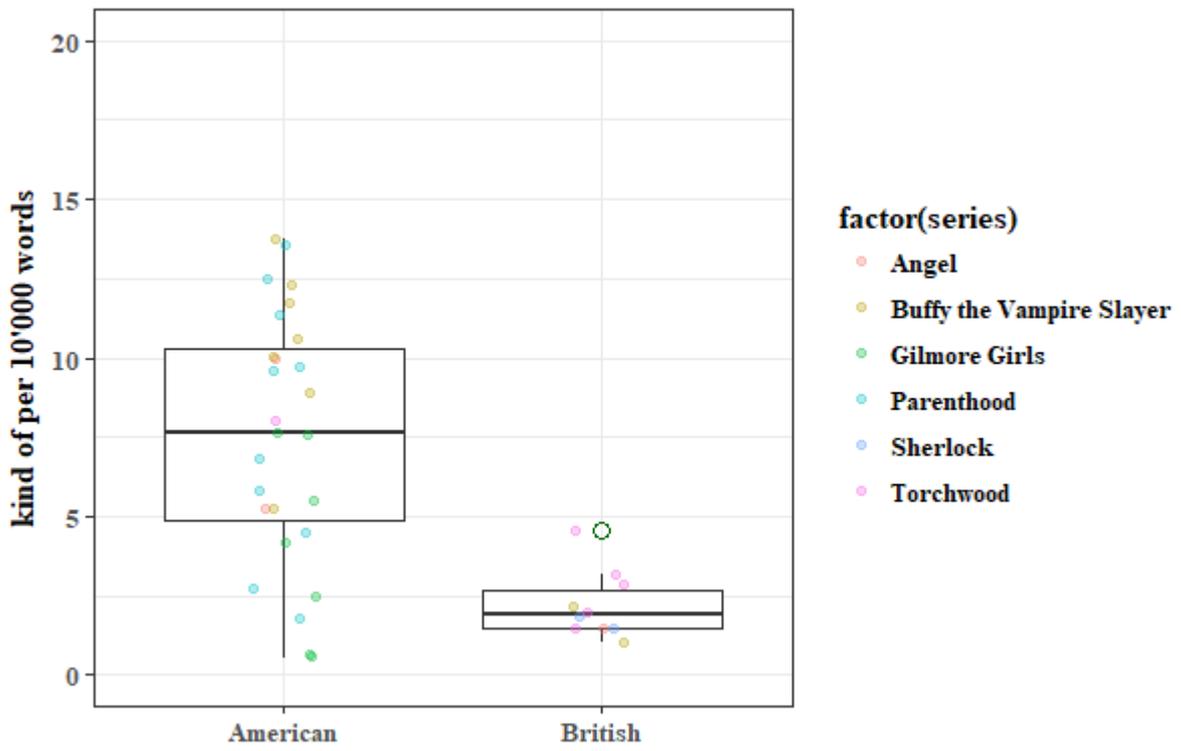


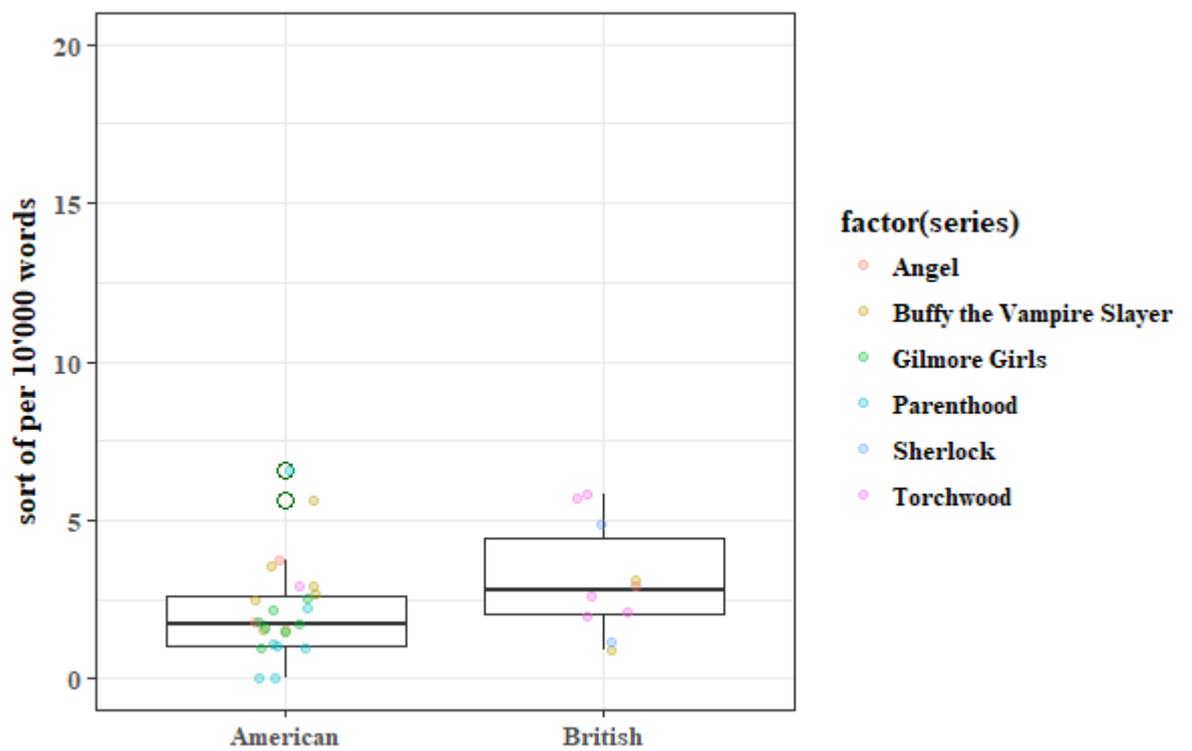
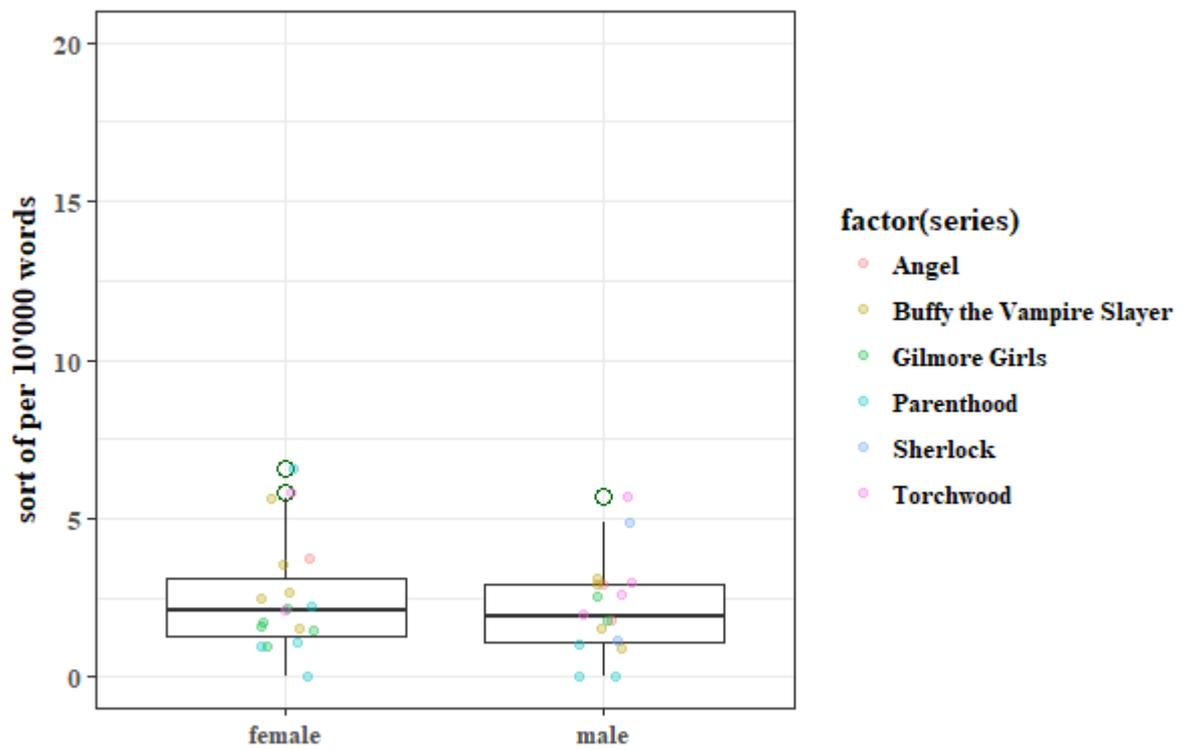




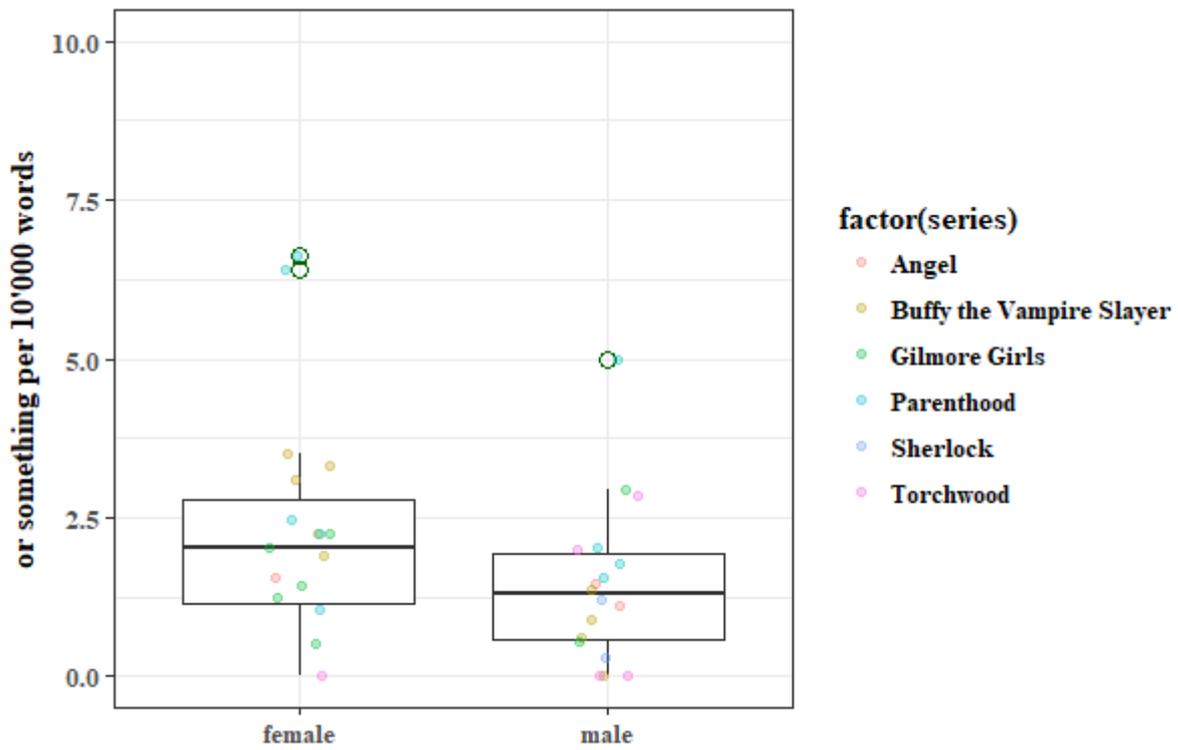
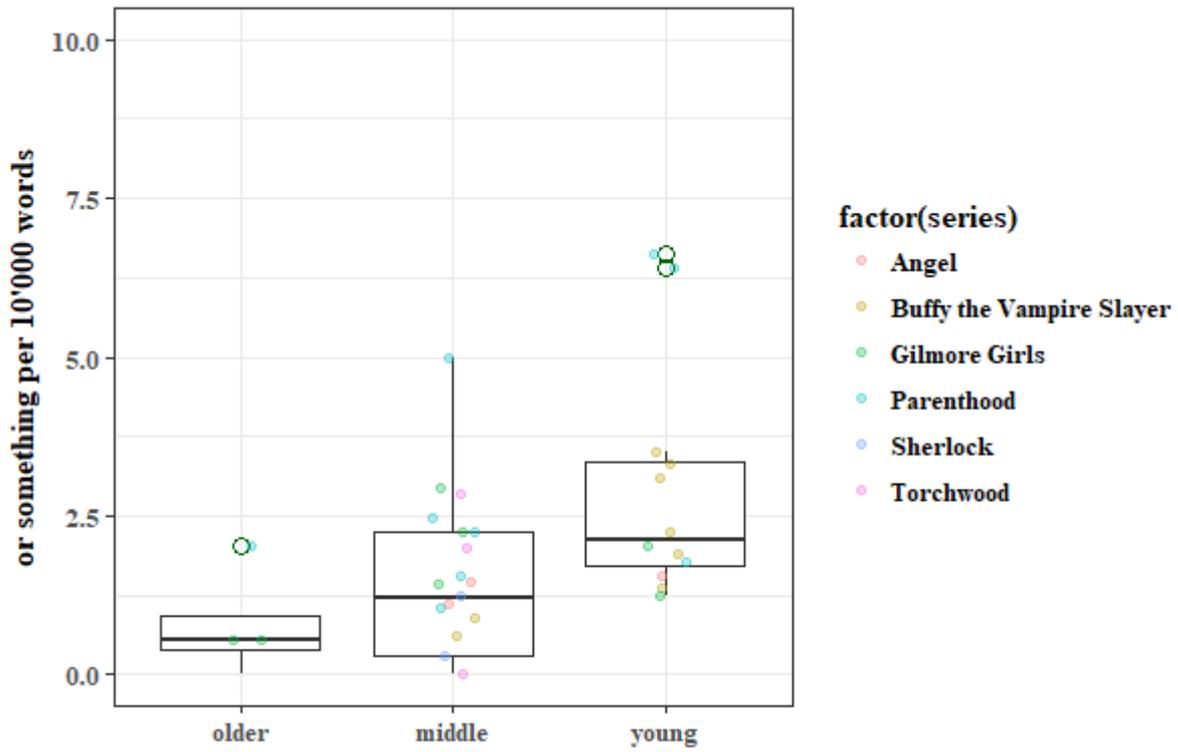
*Hedges*

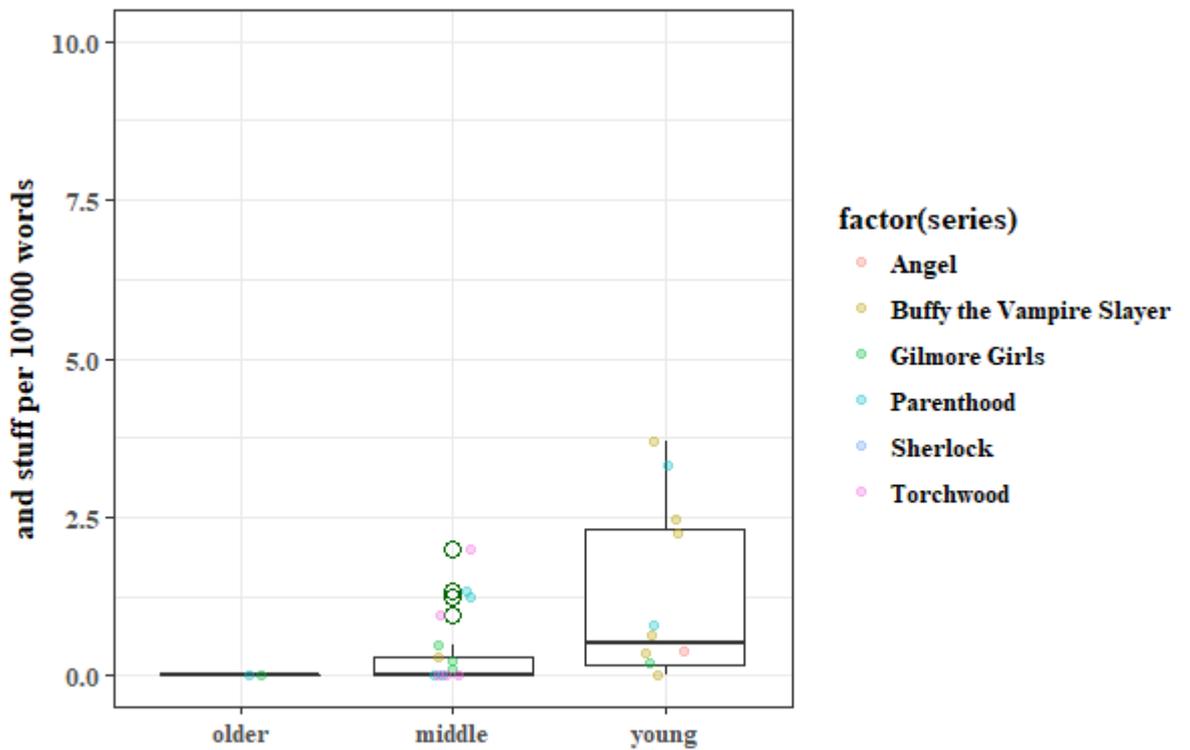
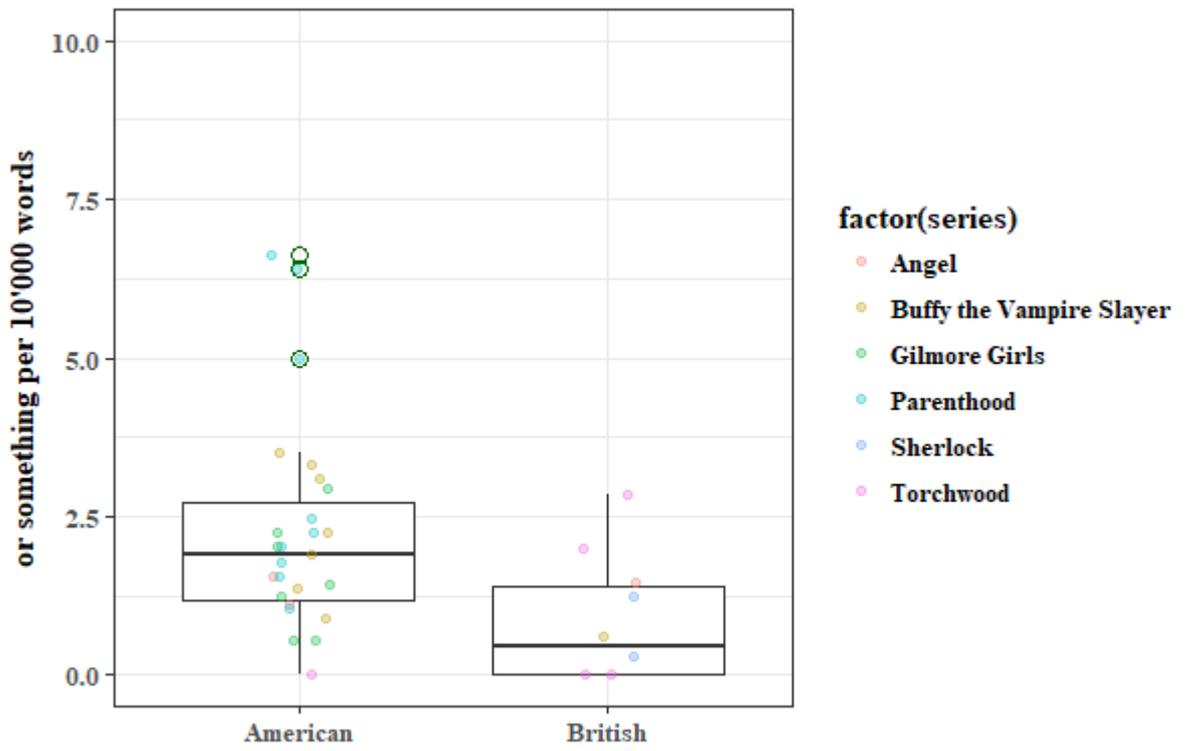


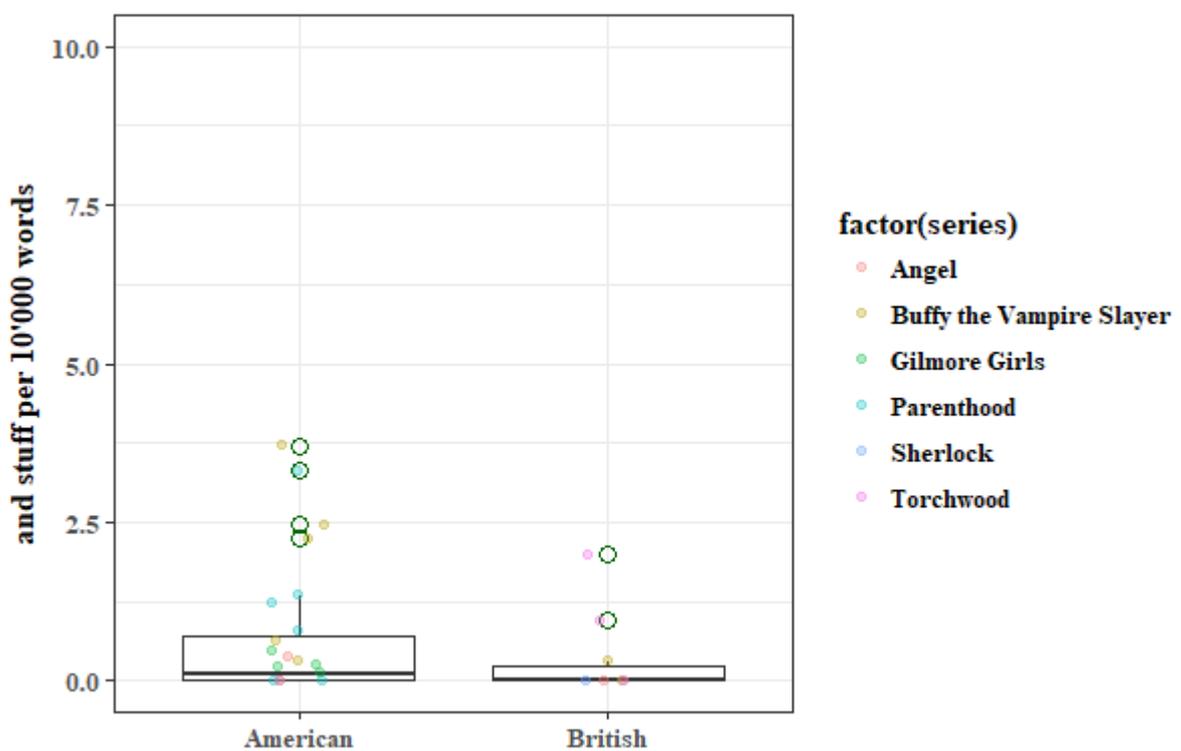
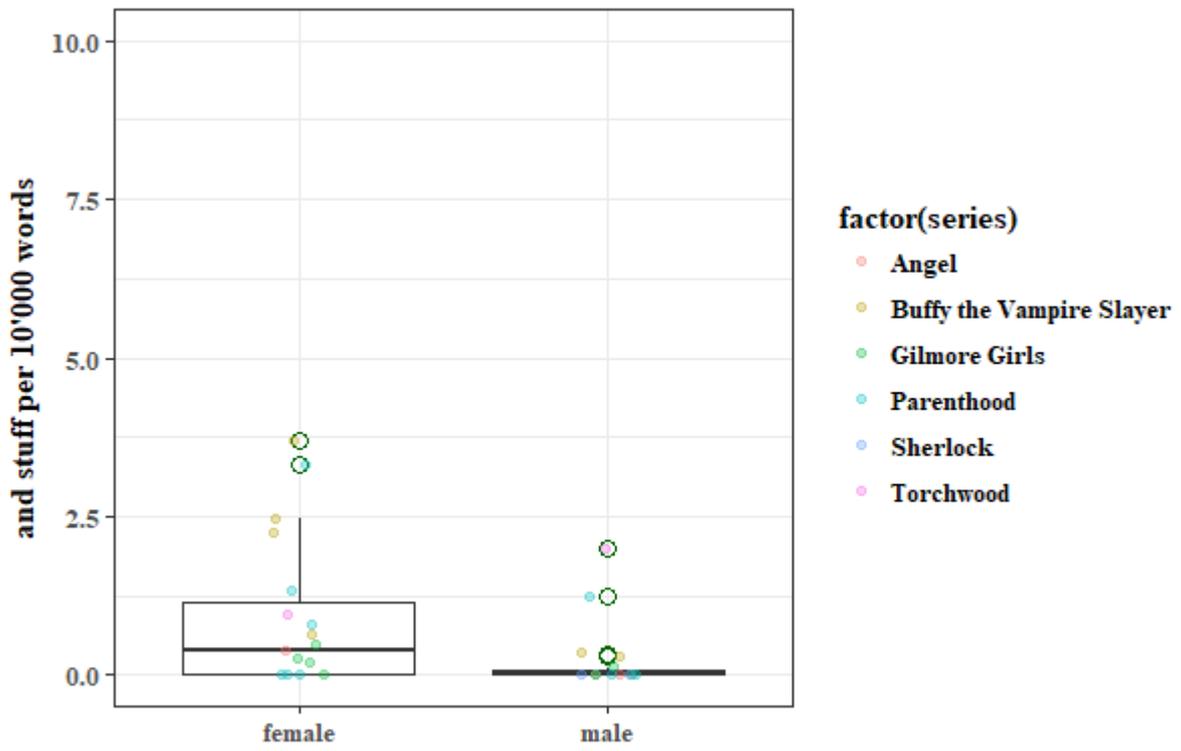


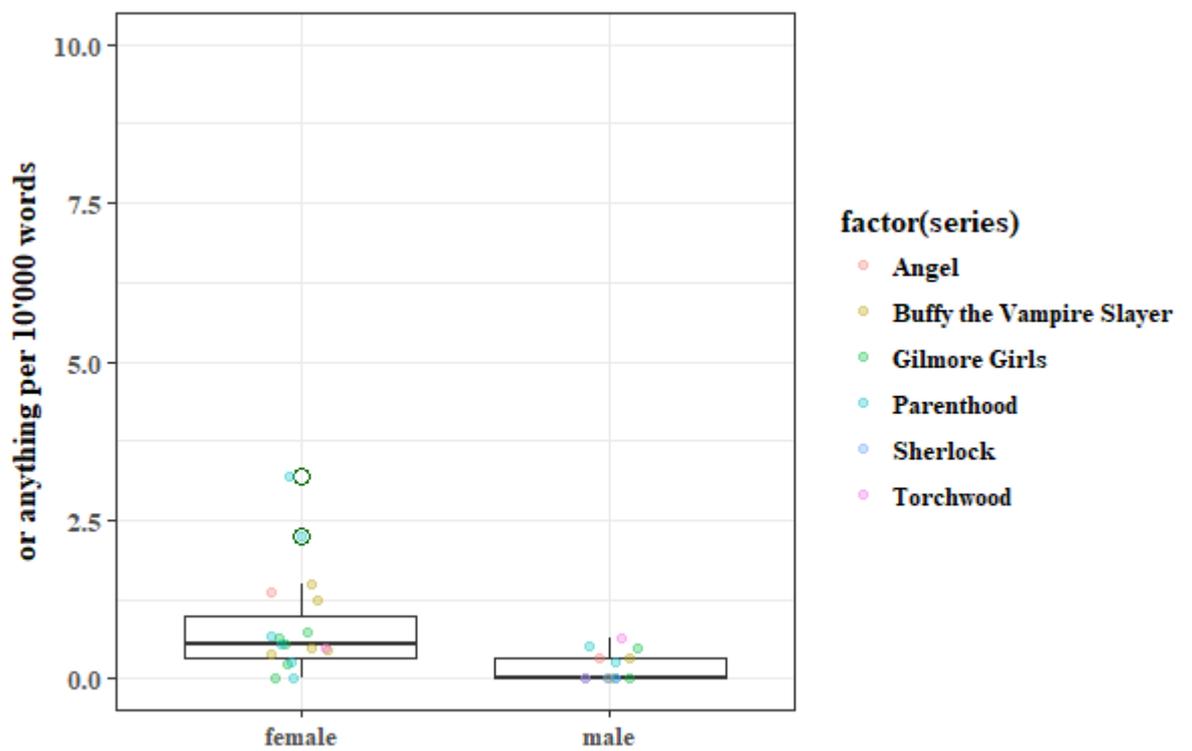
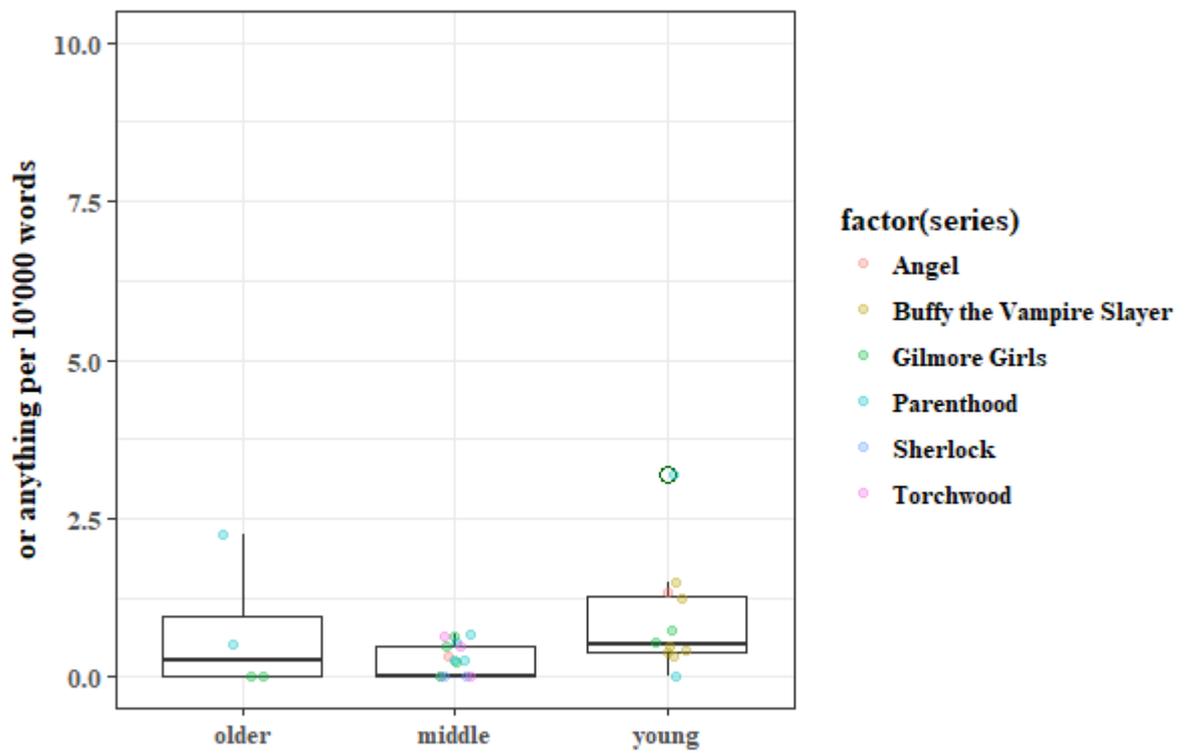


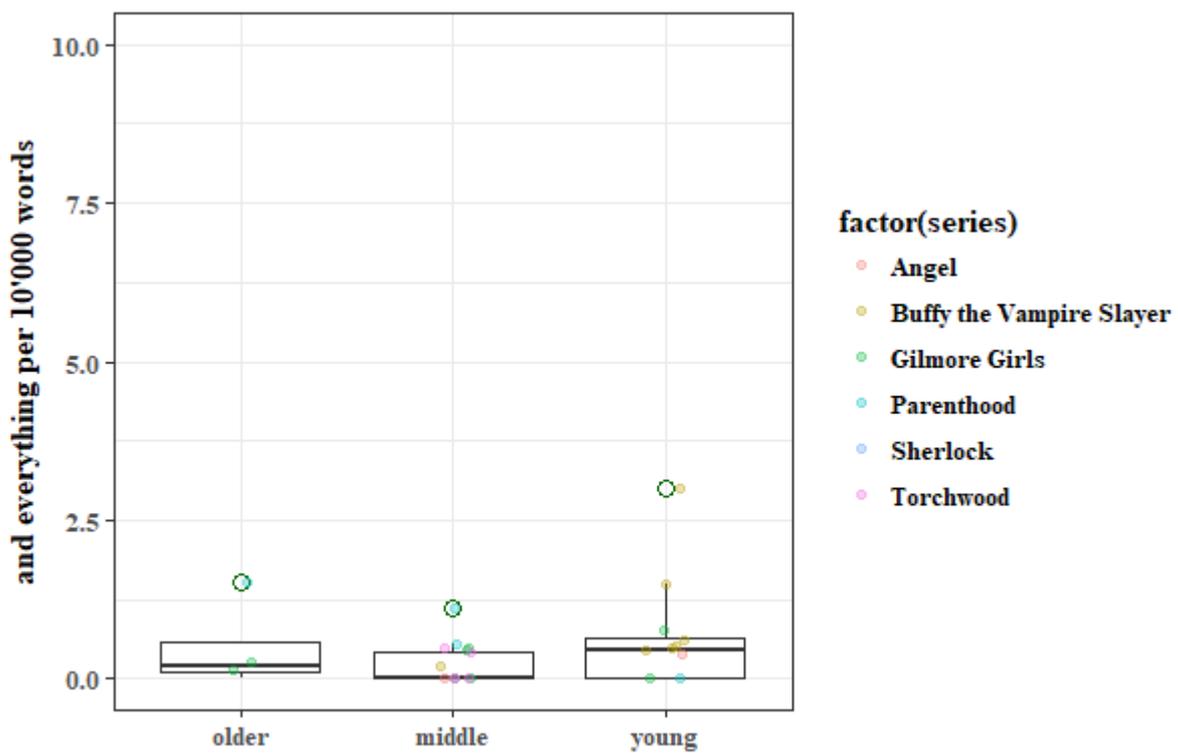
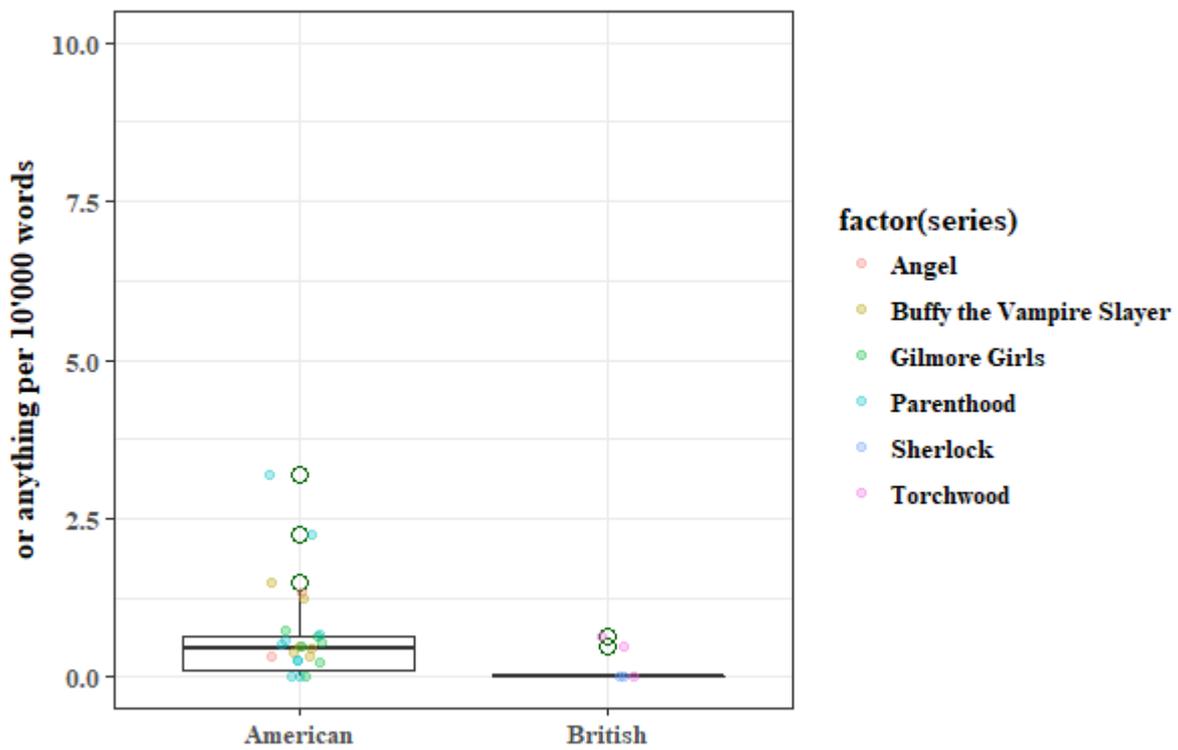
General extenders

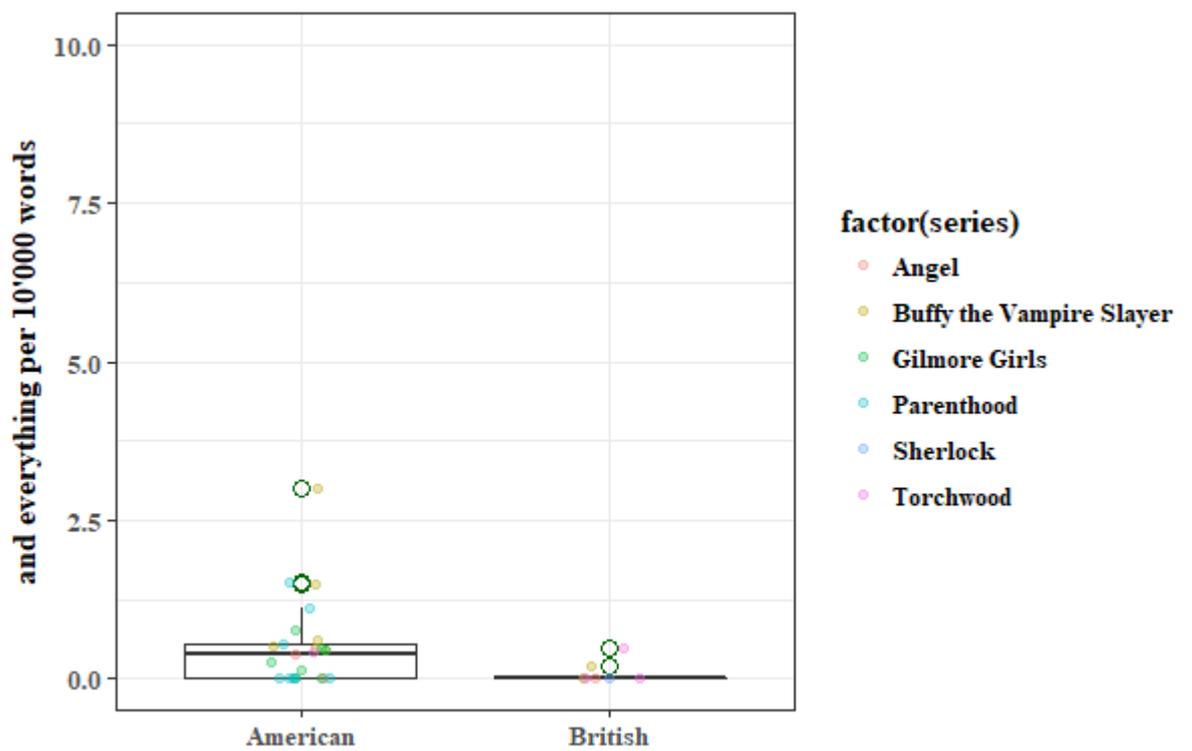
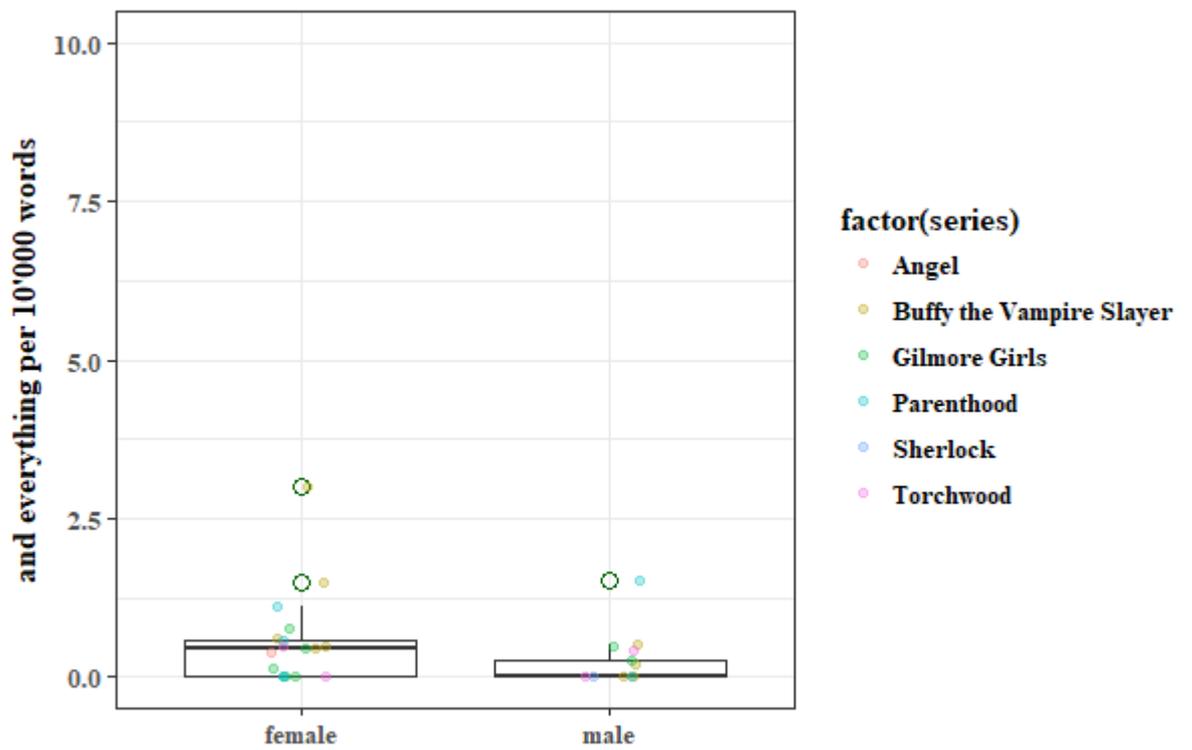




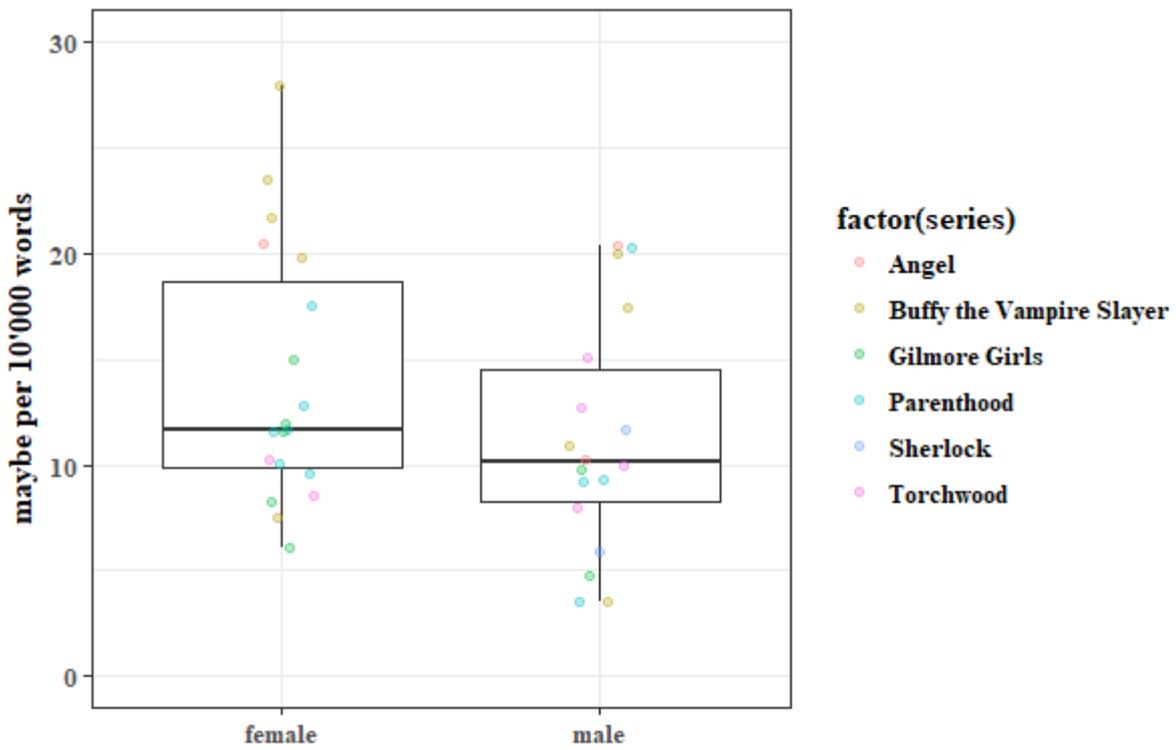
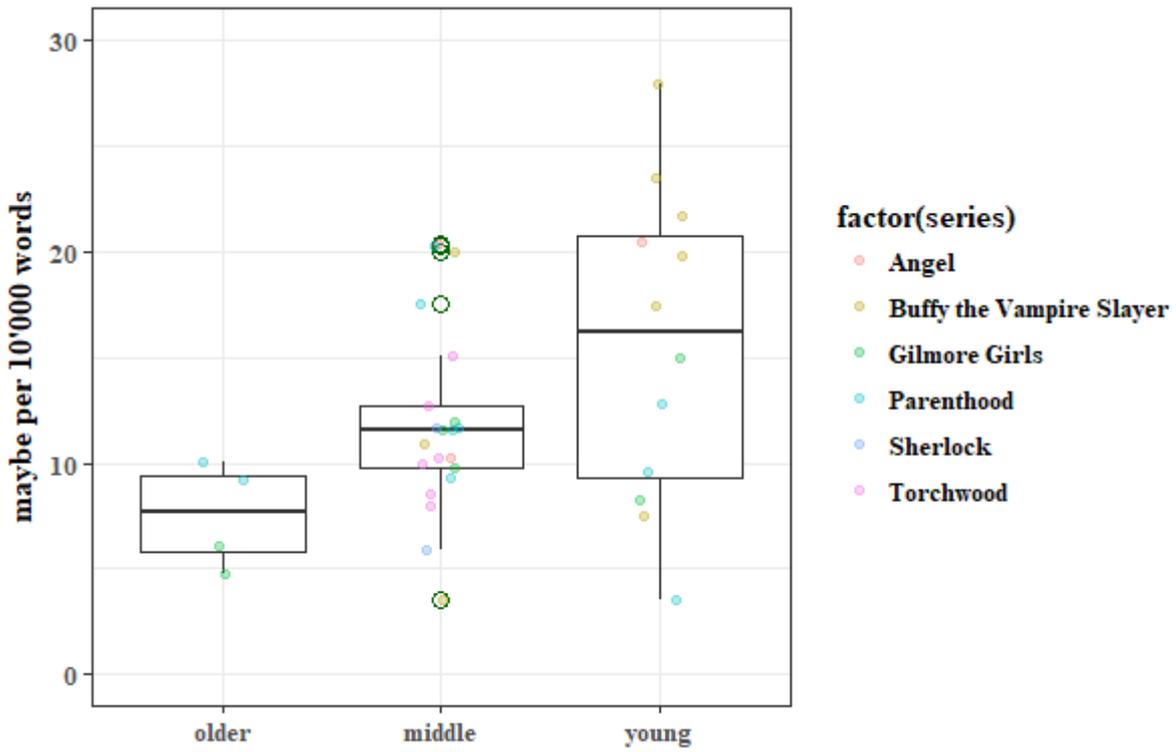


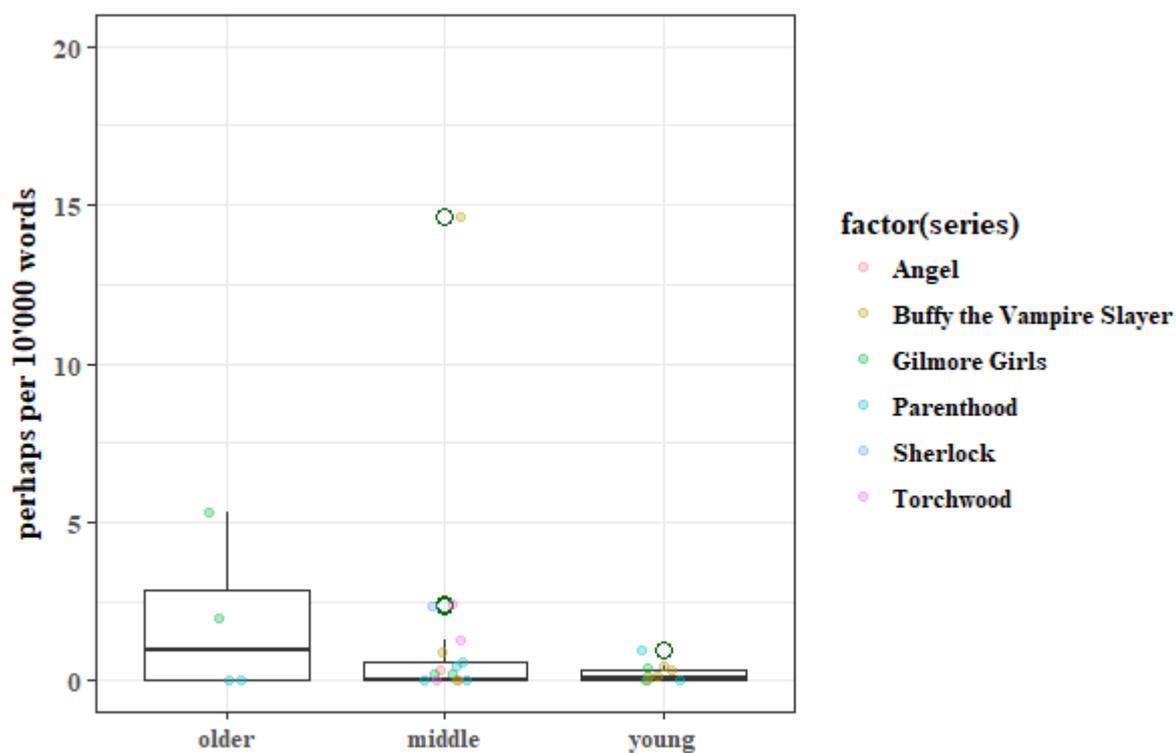
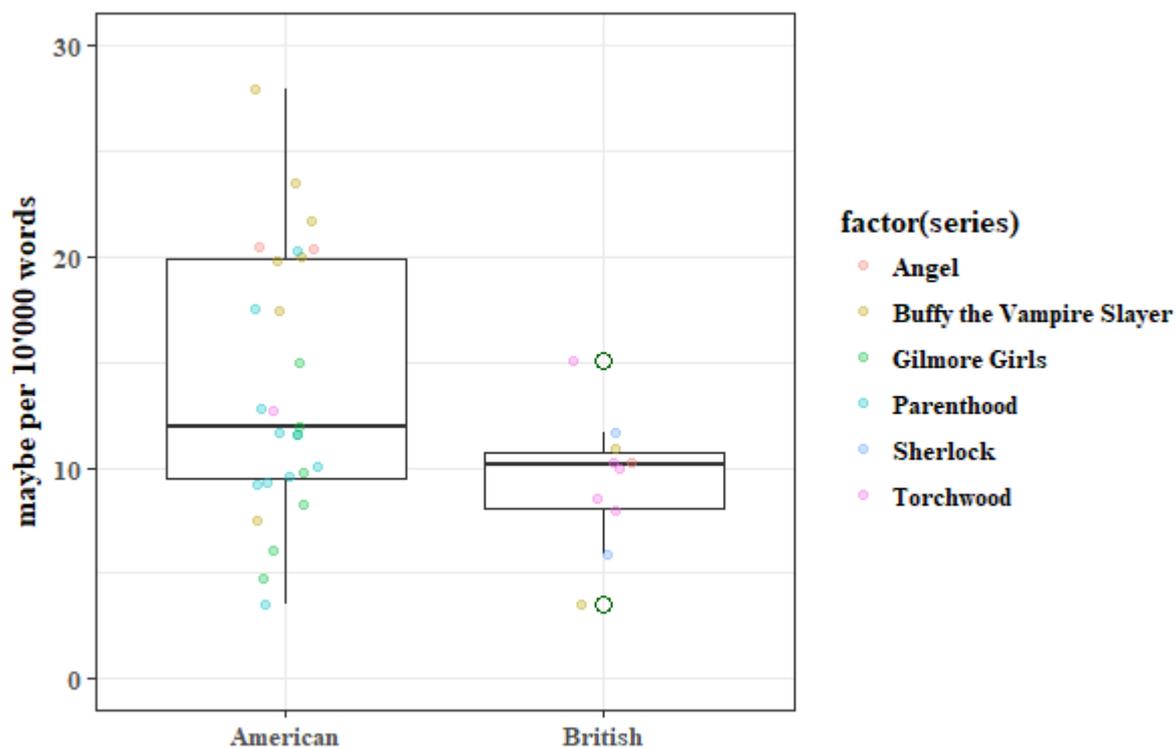


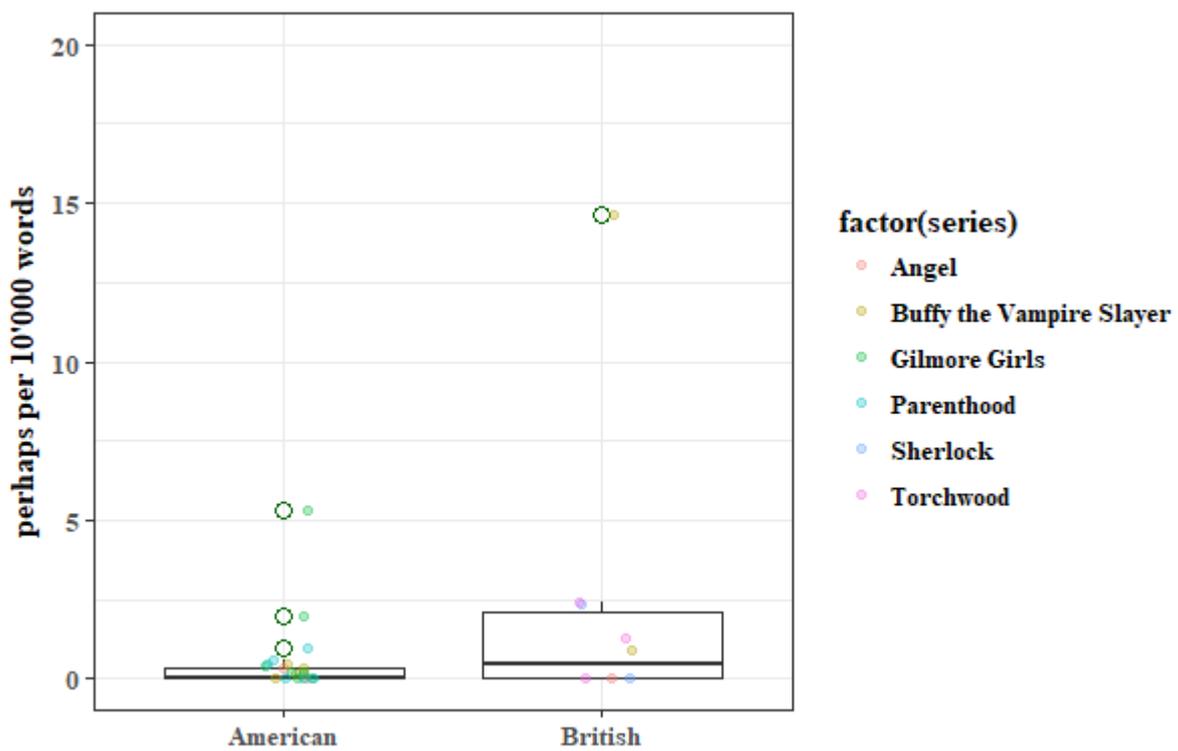
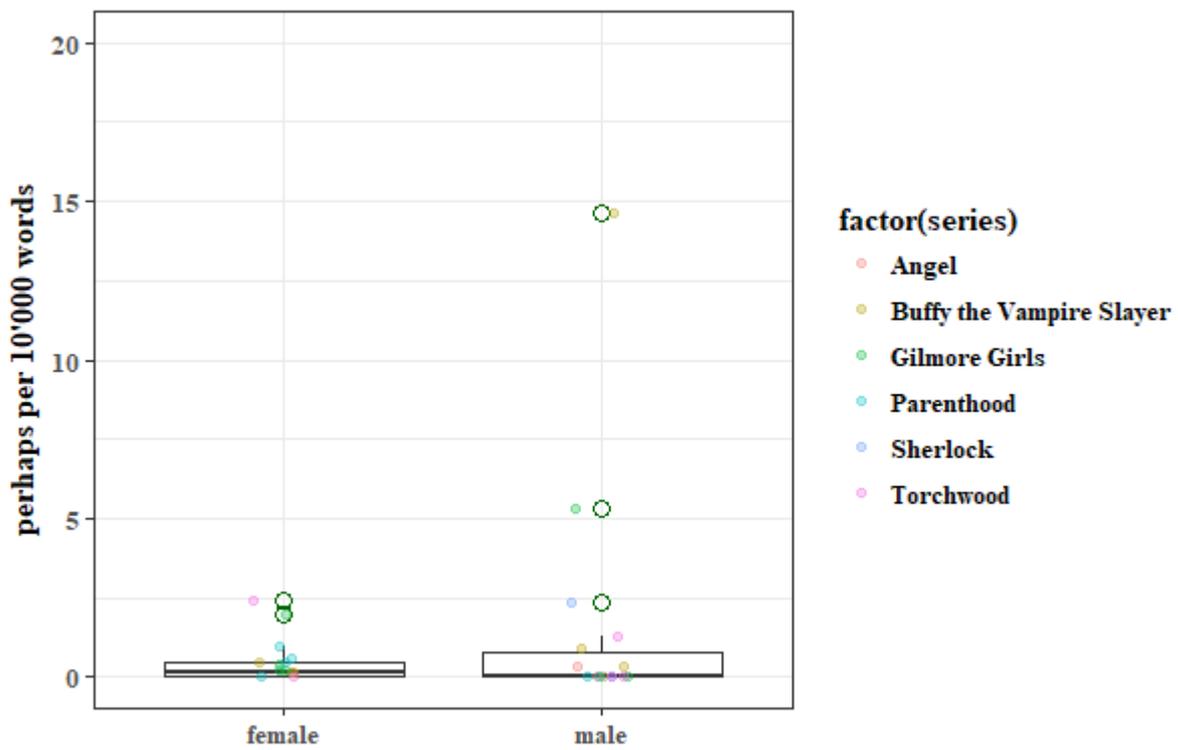


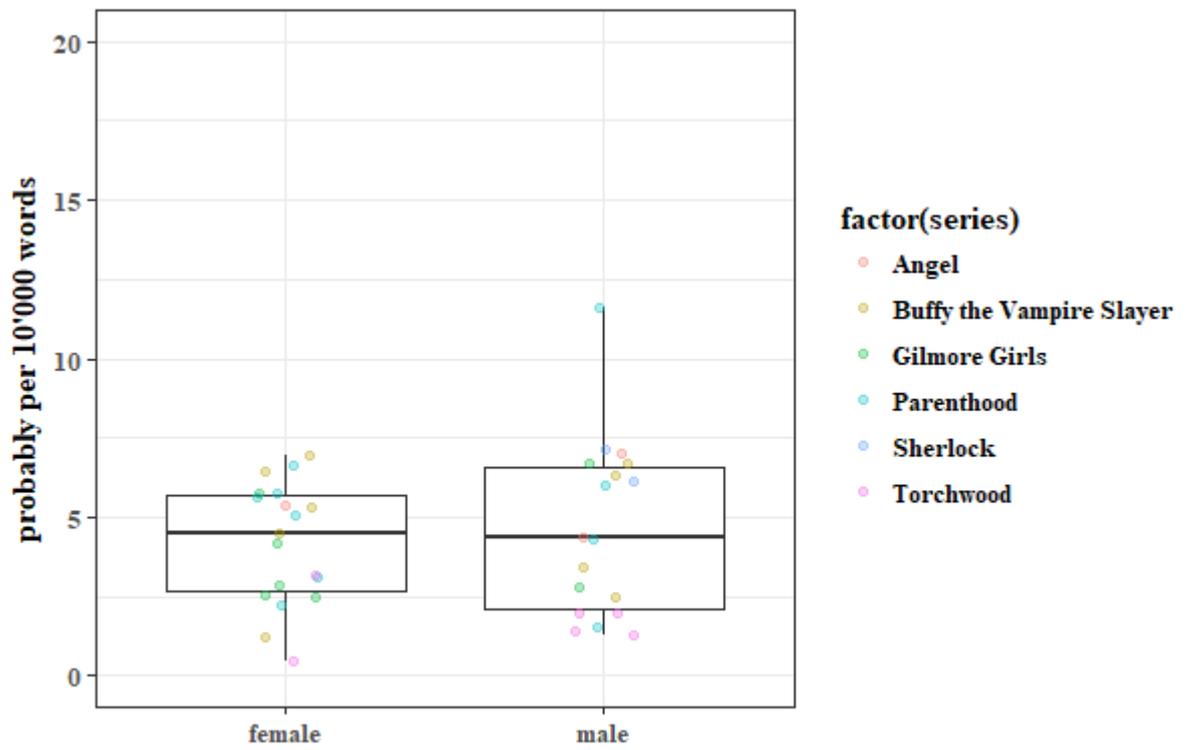
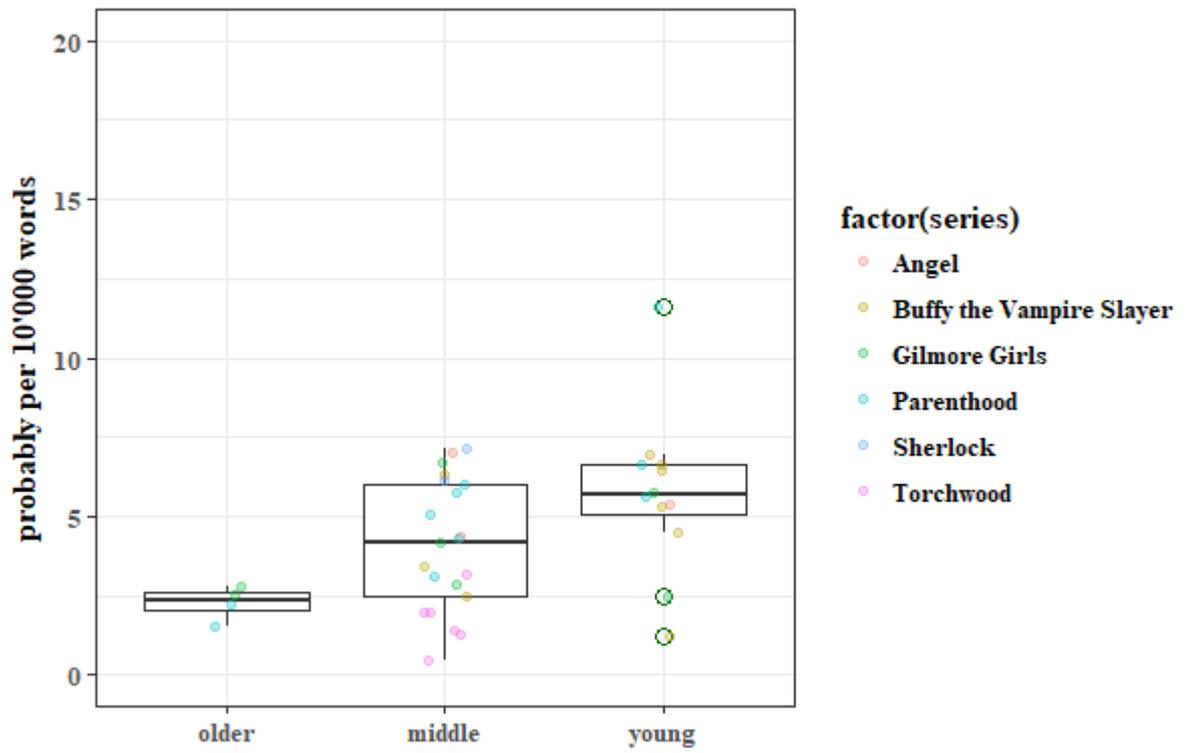


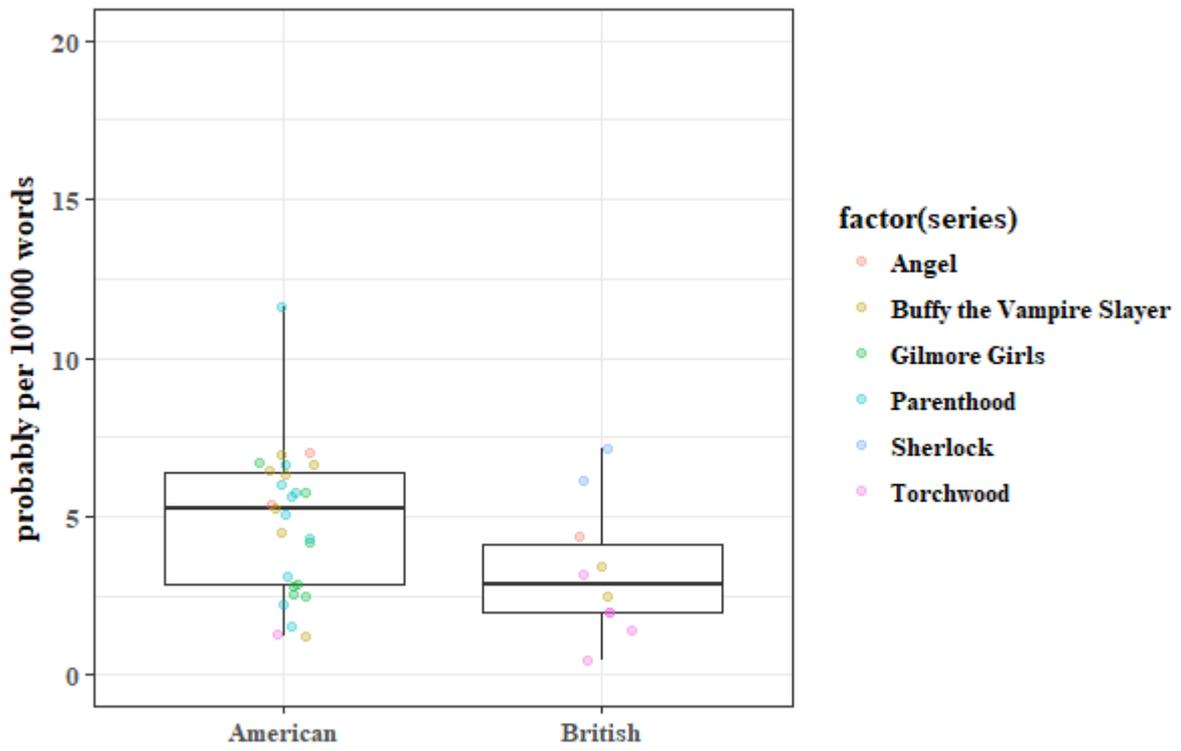
*Modal adverbs*











*Intensifiers*

