Northumbria Research Link

How to conduct sociolinguistic research in online public video

Selina Jeanne Sutton

PhD

2020

How to conduct sociolinguistic research in online public video

Selina Jeanne Sutton

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of Doctor of Philosophy

Department of Computer and
Information Sciences

November 2020

Abstract

There has been an increase in the sharing of video, and thus speech, in social media. Yet research has focused on written language. Considering our communications are continually becoming more computer-mediated, researching of the impact of such interaction contexts upon our speech is overdue. In this thesis I ask, "how can we conduct sociolinguistic research in online public video?". Sociolinguistics is the study of the interplay between social factors and speech. Four key aspects that construct a sociolinguistic research method are identified - i) Formulating Research Questions, ii) Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis - and theorised in relation to online public video research. A case study is used as a vehicle through which the research practices of these four key aspects are explored.

The case study asks, "Is speech influenced by written comments in online public video?". YouTube is rationalised as an interaction context where explicit feedback is received via viewer comments, but who is commenting is ambiguous. Hence, the sociolinguistic theory under examination is Audience Design which assumes intraspeaker variation is an automatic response to one's audience. It is hypothesised that a YouTuber will adjust their speech as they gain information about their audience via the comments. This thesis reports on the quantitative analysis of comments and the speech variable uptalk, as well as an online ethnography that motivates the quantitative analysis of a second speech variable, word-medial trochaic /t/. The relationship between the comments and speech appears to be dependent upon the YouTuber's career stage and their engagement with the comments.

The contributions of this thesis are illustrating the value of considering speech when researching social media, and defining resources to guide sociolinguistically-aligned research in online public video.

List of Contents

List of Tables

List of Figures

Acknowledgements

"Whatever you do, don't congratulate yourself too much or berate yourself either. Your choices are half chance, so are everybody else's"

- Baz Luhrmann, "Everybody's Free (to wear sunscreen)"

Declaration

I declare that no outputs submitted for this degree have been submitted for a research degree of any other institution. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

A portion of the work contained in this thesis has been published as conference poster abstracts:

Sutton S, Foulkes P. (2018) Speech on YouTube – Vloggers and theories of style. *British Association of Academic Phoneticians (BAAP 2018),* University of Kent, UK, 12th – 14th April 2018.

Sutton S J. (2020) Speech on YouTube – Audience Design, Defining Place, and Speech Style over time. *British Association of Academic Phoneticians (BAAP 2020),* University of York, UK, 1st – 3rd April 2020.

Any ethical clearance for the research presented in this commentary has been approved. Approval has been sought and granted by the Department Ethics Committee.

I declare that the word count of this thesis is 88,679 words.

Name: Selina Jeanne Sutton

Date: 30/06/2021

Chapter 1.
Introduction

"[Computer-Mediated Communication] researchers do not need to reinvent the wheel, since they can draw on long standing traditions from other linguistic fields and communication studies"
(Bolander and Locher, 2014, p. 19)

The research presented in this thesis explores how speech can be studied within the context of online public video sharing site, specifically YouTube. In recent years there has been a dramatic increase in the sharing of audio, and thus speech, in social media as well as other computer-mediated communication, such as the consumption and sharing of video publicly and privately. Yet investigations in this area to date have focused on the use of written language; examining speech is a rarity. This thesis presents a methodological contribution in the form of guidelines for research practice that can benefit researchers who aim to analyse speech in online public video. In doing so, the work in this thesis opens up a new topic to explore that is in keeping with technological developments, and the insights gained could influence the practice of content creators and the industry they work within.

This introductory chapter establishes the research context of the thesis. First, the site of study (YouTube) and the specific types of YouTube data that are focused on in this thesis is described (section 1.1 and 1.11 respectively). Then, the interrelated research topics of social media, computer-mediated communication, social computing, and new media are outlined (section 1.2), and the research

107

field of sociolinguistics is defined (section 3.3). Knowledge from these previously separate fields will be brought together throughout this thesis. Section 1.4 briefly reviews the researching of the influence of mass media in sociolinguistics, not so this literature can contribute to this thesis but to differentiate clearly how the work herein minimally overlaps with these research interests. In section 1.5 the thesis research aims are stated, section 1.6 provides an overview of the thesis, chapter by chapter, and section 1.7 states the thesis' contributions.

1.1 Online Public Video

YouTube is the most prolific video sharing website. Not only has the brand name become synonymous with engaging with video online (e.g. "I'll YouTube it"), it now boasts 1 billion users, with content in 80 languages and 1 billion hours of content consumed per day (YouTube, 2019e). In 2011, it was estimated that the total size of YouTube was 448 million videos with an aggregated length of 2,649 years (Ding *et al.*, 2011). Thus, it has great potential to provide useful, interesting speech data.

At its simplest, YouTube is a website where users can upload videos, view and rate others' videos and share them on other platforms by using weblinks or embedding the YouTube video player into webpages. The uploader of a video is able to make it private or publicly viewable, as well as sign up to the YouTube Partner Programme: a scheme enabling the uploader to make money from having advertisements added to their videos and via other revenue streams. The site has a relatively stable structure, with each user having a channel where the videos they upload are collated, and the ability to create collections of their own and others' videos through playlists. Users can subscribe to channels as well, signing up to be notified by email or via other means that a channel that they subscribe to has uploaded a new video. Once simply a free video repository, it is now possible to watch video live streamed as well as broadcast television (YouTube, 2019f), download content to watch later, and pay a fee for an

advertisement-free experience (YouTube, 2019d). It now also has specific services for music (YouTube, 2019c), children (YouTube, 2019b), and a Creator Academy providing guidance to those who want to create a career or at least earn some income from creating content for the site (YouTube, 2019a).

YouTube contains many different types of video content. Ding et al (2011) categorise YouTube content as either i) user generated content: a video the user has recorded with the intention of uploading it to YouTube, or ii) user copied content: video that was recorded for other purposes and originally distributed outside of YouTube. This can include video from television, movies, music videos, live streams on other sites etc. However, this binary distinction is quite restrictive and overlooks the possibility that a video may contain both user generated and user copied content. In contrast, Liikkanen and Salovaara (2015) identified 3 main types and 12 subtypes of video when considering music videos alone, with many of the subtypes acknowledging user editing of copied content and the splicing of user generated and user copied content. Thus, YouTube could be rationalised both as a unique interaction context where users communicate to their audiences through user generated and user curated content, and equally as a video data repository of content that was created for other purposes and originally engaged with elsewhere in some way, providing two distinct areas with research potential.

Because of its prolificacy, YouTube would likely be the first venue that researchers would think of if they were interested in online video. Thus, conducting a thesis on how to perform sociolinguistic research using online video in YouTube will maximise its contribution and impact. It should be noted, however, that the insights gained from the work described herein and the resulting guidance provided could also be applied to other online sites and platforms that contain public video. For example, while the YouTube data engaged with herein is pre-recorded and edited, there are sites that allow for or are a repository of live-streamed video. Twitch (Twitch, 2019), for instance, is a site dedicated to live-streaming video and many other sites incorporate such functionality (e.g. Facebook Live on Facebook (Facebook, 2019b)). Further, public video messaging and micro-videos (cf. Redi *et al.*, 2014) are becoming

increasingly popular. Sites dedicated to micro-videos include TikTok (TikTok, 2019) and the now archived Vine (Vine, 2019), while Twitter (Twitter, 2019), Facebook (Facebook, 2019a) and Instagram (Instagram, 2019) also allow videos to be shared. Thus, while the work herein is focused on one site, its findings reach far more broadly.

1.1.1 YouTube data

In attempting to define the YouTube site and its content it becomes apparent that, as Burgess and Green (2009, p. 88) describe, YouTube is "a massive, heterogeneous, but for the most part accidental and disordered, public archive". Thus, it would be inappropriate, even impossible, to be inclusive of all possible video types in this thesis. Therefore, from the outset this thesis delimits its interests to 'user generated content' (a video the user has recorded with the intention of uploading it to YouTube), to use Ding et al's (2011) categorisations. At its simplest, other video types (e.g. user copied content) are not themselves forms of computer-mediated communication (defined in section 1.2) although viewers may interact with them using YouTube's various mechanisms. Rather, as already mentioned above, it is YouTube's provision of a unique interaction context where video creators communicate to their viewers via video and viewers communicate back that is of interest in this thesis. In stating this, it is important to highlight that it is not the video alone which is the data of interest but also data that evidences viewer interaction with said video, such as various YouTube metrics (number of views, likes, and subscribers), and richer data sources such as the content of comments.

1.2 Computer-Mediated Communication

From a practical perspective, the term 'online public video' communicates clearly the object of study in this thesis is. However, when considering this object from a theoretical or conceptual perspective it becomes far more difficult

to define. This reflects the multiple interrelated fields of research that are relevant to this topic, their differing foci and terminology. As will be described below, these include Social Media, Computer-Mediated Communication, Social Computing, and New Media.

First, 'online public video' can be classed as a form of 'social media'. There are many definitions of social media, and these seem to have evolved over time as the technology that underpins them has changed. Ouirdi et al (2014, p. 119) designed the following inclusive definition of social media as a result of a content analysis of research literature:

> "a set of mobile and web-based platforms built on Web 2.0 technologies, and allowing users at the micro-, meso- and macro-levels to share and geo-tag user-generated content (images, text, audio, video and games), to collaborate, and to build networks and communities, with the possibility of reaching and involving large audiences."

From this definition it is clear that all social media is a form of computer-mediated communication (CMC). However, again, defining CMC is not straight forward. There are multiple definitions (e.g. Herring, 1996), but essentially all refer to "any human communication achieved through, or with the help of, computer technology" (Thurlow, Lengel and Tomic, 2004, p. 15). Further, social media is inherently social and thus can be classed as 'social computing'; "'Social Computing' describes any type of computing application in which software serves as an intermediary or a focus for a social relation" (Schuler, 1994, p. 29). Further still, the media element (images, text, audio, and video) of social media is also fundamental, media becoming 'new media' in the social media context. New Media is, again, difficult to define because the term is used so inclusively. But, essentially, new media is the result of developments in media and computing converging. Lister et al (2009) argue that the coining of the term 'new media' was in part to emphasise an unprecedented degree of change, culturally as well as technologically. In 2001, Manovich (2001, p. 19) argued:

> "we are in the middle of a new media revolution – the shift of all culture to computer-mediated forms of production, distribution and communication".

These multiple overlapping terms evidence how it is not possible to untangle the social, communication, and media elements and their technological enablers that are at play in 'online public video'. Thus, literature, concepts, theories and insights from all of these research fields - i) Social Media, ii) Computer-Mediated Communication, iii) Social Computing, and iv) New Media – will contribute, to different degrees, to this thesis. But for simplicity, the use of the term 'Computer-Mediated Communication' (CMC) will be preferred throughout the thesis unless it is deemed that a distinction between CMC and another of the entities described above is necessary. This will also provide a continual reminder that *communication* is at the centre of this work.

1.3 Sociolinguistics

'Sociolinguistics' is the study of language in relation to social factors (Tagliamonte, 2006; Gordon, 2013), the term 'language' being used broadly to refer to an array of communication resources (speech, verbal language, written language etc). Therefore, when conducting sociolinguistic research, one must define specifically the communicative resource being studied, and this thesis will focus on speech. Colloquially and in other fields, the term 'speech' is used to refer to verbal communication. However, for the work herein it is important to differentiate between studying spoken words and studying the sounds that are used to make up those words (phonetics (Bussmann, 1998c)). If a study was to record someone speaking and then examine the words that are used then the study's focus is *spoken language*. In contrast, if the focus of the work was the accent of the speaker and the sounds used to create that accent then the study's focus would be *speech*.

In sociolinguistics one may study the interplay between phonetic and social factors that influence the production and perception of speech (Foulkes and

Docherty, 2006; Foulkes, Scobbie and Watt, 2010). Each person's speech is unique, and speech varies across ('inter-speaker variation') and within ('intra-speaker variation') speakers. The same person may speak differently at different times, and factors contributing to this variation include the speaker's knowledge of and relationship to the listener, the topic discussed (including its emotional content), and the physical environment in which the conversation is taking place. Speech features are also 'indexical'; they convey information about the speaker. A wealth of knowledge has been generated about the speech features that convey regional, social and cultural background including speaker age, class, gender, ethnicity and membership to communities of practice. Identity is complex, and speech is used as a resource to portray and emphasise different facets of one's identity at different times. A person's speech can also *change over time*. Social factors play a vital role in long-term change, such as migration, moving cities, or building relationships with new people. It is this speech change and variation, the reasons why and the processes by which it occurs, that are the focus of sociolinguistics.

The term 'Sociolinguistics' also encompasses many different branches of investigation including the Sociology of Language, Interactional Sociolinguistic, and Linguistic Anthropology, all of which predominantly employ qualitative methods. However, there is one branch which takes a quantitative approach. Variationist sociolinguistics investigates the correlations between linguistic features and social factors using statistical models (Foulkes and Docherty, 2006), with the data more often being speech. From a practical perspective, the nature of speech variables allows for sufficient data for quantitative analysis to be collected far more efficiently in comparison to other linguistic variables. When studying a speech variable (e.g. a single speech sound) the likelihood of a sufficient number of tokens (examples of the speech variable) arising during data collection (e.g. an interview) is high. In comparison, other linguistic variables (e.g. a specific word or grammatical structure) are less likely to occur and so collecting a sufficient number of tokens to allow for statistical analysis can be difficult (Tagliamonte, 2006).

However, as will become evident in sections 2.2 and 2.3, social media now provides the opportunity for online written language data to be collected and analysed far more efficiently than off-line language data. Further, the unique sets of interactional qualities that online platforms and media bring together provides novel contexts in which communication can be studied. The overall aim of this thesis is to identify why studies have focused on language and overlooked speech to date and how this can be addressed.

1.4 Media in Sociolinguistics

From the title it may at first seem that prior sociolinguistic research on the influence of media would be of fundamental importance to this thesis. Actually, the research herein runs tangential to these interests, rather than overlapping with them. This is because sociolinguistics has so far considered the role of mass media in language change in *everyday speech* whereas this thesis questions how we may conduct sociolinguistic research on language change *within online public video*. To crystallise how the research herein is complementary to prior sociolinguistic work, a short literature review is provided.

For sociolinguists the fundamental questions in regard to mass media to date have been: 'is it involved in changes in speech?' and if so, 'how?'. To clarify, this is different to asking how mass media may act as a source through which speakers can access stylistic resources, such as vocabulary, to incorporate into their own repertoire for interactions, or how it could offer or assist in the construction of new social meanings of linguistic features. In comparison, it has generally been agreed that speech is different. As Trudgill (2014, p. 216) emphasises, it is generally believed that regular face-to-face contact is a necessary pre-requisite for changes in speech otherwise "everyone in the British Isles would now have an American accent, or at least there would be progress in that direction." However, there is a small but steady stream of empirical work going back to as early as the 80s that has alluded to the potential for the media to play some sort of role, probably minor, in speech change (see Sayers, 2014).

Trudgill (1988, p. 44) himself suggested TV was "softening-up" viewers, readying them to take on a change in speech.

To date, the only investigation that has tested experimentally the role of broadcast media in change to speech is referred to as the Glasgow Media Project (Stuart-Smith, 2002 - 2005). Here, Jane Stuart-Smith and colleagues found that emotional investment in the content of the broadcast media was required for its linguistic content to have an influence on viewers, rather than mere prolonged exposure. Even then, it was a minor predictor in the statistical model that included many other factors. Further, the influence of the broadcast media was catalytic; it accelerated linguistic changes that had already begun rather than triggering new ones. Thus, Trudgill's (1988) intuition appears to have been along the right lines.

The other notable contribution to the discussion of broadcast media's role in language change has been Sayers's Media Innovation Model (Sayers, 2014b). Here, Sayers describes a linguistic feature being taken from a source community and incorporated into a media text (e.g. the script of a television show), this process being the 'mediation' of the linguistic feature. Then, this media text is broadcast to the community that is in the process of adopting this speech feature. This model is built from the practices of the Glasgow Media Project (Stuart-Smith, 2002-2005) which, unlike the other literature that Sayers reviewed, examined i) speech data of participants from the adopting community, ii) the content of media texts, *and* iii) the media engagement practices of the participants with said media texts, along with other data (e.g. attitudinal data).

Sayers's paper was heavily debated in a series of articles in edition 18(2) of the Journal of Sociolinguistics (2014). One of the most notable criticisms levied at the Media Innovation Model is that its motivation and the research project on which it is primarily based are somewhat incongruent. Sayers (Sayers, 2014b) argues that the motivation for the model is the observation that the same changes had occurred in geographically dispersed locations, a phenomenon referred to as globalisation and these speech features being described as 'global linguistic speech features' (Sayers, 2014b). However, in response, Stuart-Smith (2014)

emphasises that the speech features studied in the Glasgow Media Project, and their linguistic constraints, are local; they do not simply mirror the mediated speech feature. She argues instead that speech features are reinforced or enhanced by engagement with mediated versions, resulting in media acting as a catalyst, as was found. Further, Stuart-Smith (2014) interpreted Sayers's model as deterministic, although Sayers pre-empted this criticism by emphasising that detail had been sacrificed for simplicity. Overall, little headway has to been made in discovering what kind of influence mass media may have upon change in speech.

In contrast, the research herein considers how the configuration of online public video interfaces creates a novel interaction context. Whether and how speech change can occur within this interaction context is the focus of this work, rather than whether the media that is online public video is an instigator that triggers or catalyses speech change in other, non-mediated contexts. At its simplest, this thesis is about speech change *on* YouTube rather than *from* YouTube.

1.5 Research Aims

This thesis explores how speech can be researched within the context of YouTube. The aims of the thesis are:

i)   to gain insights into how real-world sociolinguistic research methods can be manipulated to be transferred to online research contexts and data;

ii)   to develop an understanding of the advantages, challenges, and limitations in performing sociolinguistic research in online publicly shared video.

1.6 Overview of the Thesis

To generate the knowledge required to address the research aims I will conduct a sociolinguistic study of speech in online public video (referred to throughout as *the case study*), recording and reflecting upon the thought and practical processes that were required to design and perform these research activities and how successful these research activities were. Thus, this thesis is effectively a research study within a research study.

This has affected how I have written this thesis in a number of ways. First, chapter 4 onward I have written in chronological order. Second, from chapter 4 onward I have written this thesis in the first person. While both of these practices are not typical of an academic thesis, this style of writing allows me to explicitly communicate how, when and why my practices developed, because it is through this refinement from experience that this thesis' primary contribution (the guidelines) can come to fruition.

Third, a clear distinction will be made between the **thesis questions** and the **case study questions** throughout. Because this thesis is effectively a research study within a research study there is the danger that the two become conflated, rather than the case study providing the opportunity to perform the practices and gain the experience necessary to develop the guidelines, which are the main contribution of the thesis. Thus, the following strategies are used. At the end of chapter 3 the two overall thesis aims (defined above) are refined into four **thesis questions**. To keep track of when each thesis question is being addressed, in chapters 4 through 7: i) the introduction will state which thesis questions the learnings in that chapter will contribute towards answering, and ii) there will be a "Reflecting on Thesis Research Questions" section at the end of each chapter. In chapter 2 the **case study question** will be defined. After which, more detailed versions and refinements will follow as the data is interrogated and variables are identified (chapter 4 through 7). Finally, to emphasise whether the thesis questions or the case study question is being discussed, these terms will be in bold text to encourage reader attention.

In regard to the thesis structure, chapter 2 begins with a summary of the key theoretical concepts used in sociolinguistics. Then, a state-of-the-art literature review of CMC is performed. This provides an overview of the research generated across these fields in regard to speech and language in real world and online contexts. In this chapter, the sociolinguistic theory under examination in the case study is also defined (Bell's (1984, 2001) Audience Design) and YouTube is rationalised as an interaction context within which explicit feedback can be received (via viewer comments), but who is commenting can be ambiguous (an example of 'Context Collapse' (Wesch, 2009; Marwick and boyd, 2010). This dynamic and the wealth of data available allows for speaker behaviour in relation to audience to be considered at a finer level of detail (particularly with respect to time) than previously.

Chapter 3 provides a review of the current methods used in sociolinguistics and CMC research that considers language and speech, and a number of barriers to researching speech in online public video are identified as a result. These primarily relate to difficulties in transitioning research methods for real world studies into these online contexts and media. The key methodological issues for four elements that make up a research method are identified: i) Formulating Research Questions, ii) Research Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis. And four **thesis questions** are defined in response.

Chapter 4 introduces the subject of the case study; Zoella, one of the most well-known YouTubers who progressed from being a 'microcelebrity' (defined by Jerslev, 2016) to 'A List vlogger' (defined by Bishop, 2018). In addition to outlining the case study, chapter 4 reports on the data collection and analysis of the independent variable (content of YouTube comments).

Chapters 5 through 7 report on the collection, analysis, and results of several data sets. Each chapter ends with a critical reflection on the decisions made and practices performed (the "Reflecting on Thesis Questions" sections). Chapter 5 focuses on the first dependent variable of the exploratory case study: uptalk. Chapter 6 reports the observations made from the online ethnography, observations that provide a clear direction for the kind of speech variable to be

investigated next. Chapter 7 focuses on the second dependent variable of the exploratory case study: word medial trochaic /t/. The rationales behind investigating these speech features will be revealed in the relevant chapters.

Finally, chapter 8 is a critical review of the insights gained from the exploratory case study, a discussion of the key findings and how they came about, as well as a description of how they were reformulated and collated into a set of working guidelines for future practice. How the use and impact of the guidelines will be monitored will also be outlined, and suggestions made for future work.

1.7 Thesis Contributions

The primary contribution of this thesis is a methodological one and takes the form of guidelines for research practice for researchers who aim to take a sociolinguistic approach to the study of speech in online public video. These guidelines may also be applied to other online video sharing platforms and thus have broader impact.

The second contribution of the work in this thesis is insight into the relationship between viewer feedback and YouTuber speech. This new knowledge could influence the practice of content creators and subsequently the industry that they work within. Further, illustrating the value of considering speech when researching social media opens up a new topic for researchers and industry practitioners to explore that is in keeping with technological development.

Chapter 2.

Research Review

The internet "*affords a panoramic and quantitatively unprecedented vantage point from which to study linguistic events*"
(Pfrehm, 2018, p. 122)

The first half of this chapter provides a review of the literature across a number of disciplines and topics that relate to the research aims of the thesis. Section 2.1 describes foundational sociolinguistic concepts that will be essential for understanding the review of computer-mediated communication (CMC) literature in the following sections (2.2 and 2.3) and the work in this thesis. Section 2.2 provides a summary history of different approaches to researching language online to contextualise the evolution of interests and practices over time. This provides both context and a framework within which the computer-mediated communication (CMC) literature can be reviewed with the aims of the thesis in mind (section 2.3). Based on this literature review, I argue that as the design of CMC changes so do the behaviours that they record and the online contexts in which they occur. Therefore, as is summarised in section 2.4, researchers that identify a CMC as an interesting source of data must design their research methods and practices to be responsive to that particular CMC.

The second half of this chapter defines the case study through which research methods and practices that intend to respond to the interaction context of YouTube and researching the phenomena of speech can be explored. In other words, this case study provides the opportunity to perform the practices and gain the experience necessary to develop methodological guidelines, the main contribution of the thesis. First, I will outline how YouTube is a new kind of interaction context by bringing together key social computing theories (section 2.5). Next, theories of speech style are reviewed (section 2.6), as well as the little research that has considered speech style in YouTube to date (section 2.6.4). In section 2.7, I argue that applying the sociolinguistic theory of Audience Design (Bell 2001) to the interaction context of YouTube and the data generated therein offers an ideal opportunity to investigate YouTuber-commenter interaction via speech. The focus of the case study and its research questions will be defined in section 2.8 and the chapter summarised in 2.9.

2.1 Speech Offline: Sociolinguistic concepts

In order to engage with CMC speech one must first grasp three foundational sociolinguistic concepts: i) the vernacular, ii) the three waves, and iii) the indexical field.

2.1.1 The Vernacular

Central to sociolinguistic work is 'the vernacular': "the style which is most regular in its structure and in its relation to the evolution of the language" (Labov, 1972, p. 112). In other words, the belief was, and still is to some, that this way of speaking is the truest reflection of the speech system at that moment in time and in that location, and thus is the best type of data with which to observe speech variation and change. Labov (2013, p. 3) defines the vernacular further as "the form of language first learned, most perfectly acquired, which we use automatically and unthinkingly in conversation with family and intimate

friends.".". The implication is that the vernacular is a local and/or nonstandard way of speaking and that other speech styles (e.g. more formal) are defined in reference to it.

However, accessing the vernacular is not simple. There is a conundrum known as the 'Observer's Paradox': a researcher must be present to witness or record behaviour but their very presence may cause the participant to alter that behaviour (Labov, 1972). This is particularly the case for the vernacular because this style of speech occurs when the speaker pays minimal attention to how they are talking. While it is agreed that the Observer's Paradox can never be fully resolved, researchers have continually sought ways to mitigate its impact on the speech data collected. Labov's (1966) sociolinguistic interview for his study of New York City included questions specifically designed to minimise the speaker's attention to their speech, such as the 'Near death experience' question (see section 3.1.1.1). More contemporary attempts have included allowing participants to talk to one another without the interviewer present (e.g. Docherty and Foulkes, 1999; Stuart-Smith, Timmins and Tweedie, 2007; Snell, 2010) and asking participants to record their own data (e.g. Robert J. Podesva, 2007; Podesva, 2011; Sharma, 2011; Boyd *et al.*, 2015; Hall-Lew and Boyd, 2017; Leemann, Kolly and Britain, 2018; Kim *et al.*, 2019).

However, the existence, the prestige and, simultaneously, the research value of the vernacular has been questioned. Natalie Schilling (2013b, p. 128) argues against the existence of the vernacular, for "there is no such thing as non-observed language data, and hence no such thing as one single 'most important' type of language for linguistic theory". Thus, rather than attempting to overcome the Observer's Paradox via the different methods described above, she suggests adopting an anthropological and/or ethnographic approach. These approaches take context into account so the situational factors and their impact on language can be considered, rather than abstracted away.

For many cultures, in contrast to the vernacular is 'the standard': the way of speaking that the 'establishment' (the education system, those that are socioeconomically in power) imposes. According to Coupland (2003), both the

concepts of the vernacular and the standard are dependent upon the perspective that there are authentic and inauthentic ways of speaking, with only the vernacular and the standard being authentic. However, Coupland (2003) also drew attention to an ongoing theoretical realignment. He observes that there is a moving away from conceptualising authenticities as either "standard authenticity" or "vernacular authenticity" to something more three-dimensional. This reflects how social categories (and thus the binding of social categories to linguistic varieties) have become less stable (e.g. class, gender). Thus, he posits that for speakers "the quest for authenticity [is] *more rather than less* necessary" (ibid 2003, p. 427) as authentic ways of speaking expand beyond the vernacular-standard binary.

Coupland (2003) points to several arguments that are made under this theoretical realignment that need investigation. The most relevant to this thesis is:

> "Face-to-face networks are being complemented and complicated by fast, remotely mediated networks: electronically mediated social interaction is providing new means of achieving intimacy, rapport and sociality." (ibid 2003, p. 426)

clearly indicating the value of researching online language from a sociolinguistic standpoint.

Referring to Coupland's (2003) arguments, Penelope Eckert (2003) points to how sociolinguistics has conflated authenticity and automaticity: that only speech styles that are ingrained / instinctual / produced unconsciously are authentic (e.g. the vernacular), and ways of speaking that appear to be more intentional / conscious / strategic are not. Unpacking why this ideology has been so pervasive, she observes that for many the view is that "what is interesting in language is what is beyond the conscious control of speaker agency" (ibid, p.394). But, in arguing that intentional / conscious / strategic ways of speaking should not be excluded from sociolinguistic research, she highlights that "[s]ociolinguistics should be located not at the edge of social variability, but squarely in the center" (ibid, p. 396). This centre is between the two extremes of fully automatic speech and fully intentional speech, such as phenomena that can

be explained by Allan Bell's (1984, 2001) theory of Audience Design (see section 2.6.2), the focus of the case study within this thesis (as will be explained in chapter 2).

2.1.2 The Three Waves

The second of the three key concepts from sociolinguistics that are essential for understanding the work in this thesis is 'the three waves'. How the social meaning of speech variation has been treated in sociolinguistics has evolved through three waves of analytical practice, as discussed by Penelope Eckert in (2012) and (2016). Each wave can be defined by two elements, i) where social meaning of speech features comes from, and ii) the methods used to demonstrate how they contrast. The first wave views social meaning of speech features as coming from macrosocial categories and uses surveys. The second wave views social meaning of speech features as coming from local categories and uses ethnography. The third wave views the social meaning of speech features to be multiple and as coming from speakers using them in interaction. A mixture of methods is used in third wave studies. For example, a researcher could assess the patterning of variables across individuals and groups quantitatively, and then examine how these variables are used within interaction by employing discourse analysis or conversational analysis.

The first wave began in the 1960's with Labov's and others' seminal works (e.g. Labov, 1966; Wolfram, 1969; Trudgill, 1974; Macaulay, 1977). Research in the first wave views variation in speech as resulting from macrosocial categories, predominantly class and then other demographics (e.g. sex, age) in association with class. Speech variation was conceived in reference to the standard and the vernacular, such that speech was described in terms of similarity to or difference from the standard, and thus the vernacular, along a one-dimensional continuum. Centralising class, this variation was rationalised as speakers' ability to self-correct towards the standard and the degree of exposure or access to the standard. Thus, speakers were in effect deemed to be passive, having minimal agency in how they spoke, and the way they spoke being a direct repercussion of

their position in the social hierarchy. The primary method was a survey, ensuring the collection of data from participants who fulfilled the criteria for a set of demographics (e.g. older, middle-class, female; younger middle-class female; older, middle-class, male; etc) (Eckert, 2012, 2016).

In the 1980's the second wave departed from the first by considering the meaning of variation in speech to be defined locally in relation to local identity, and then this being associated with broader macrosocial categories. Thus, there was a shift towards thinking that speech features indicated the characteristics of local identities, rather than pre-existing categories. There was also a realisation that when a survey method is taken, as is typical in first wave studies, these local identities become subsumed into macrosocial categories. The standard and vernacular were still at the centre of researching speech variation, but now the speaker was attributed social agency in their use, with the vernacular having the potential to have positive connotations locally. In order to understand local identities and their characteristics, there was a shift towards using ethnography as a method in the second wave (Eckert, 2012, 2016). For example, Milroy and Milroy (1978, 1985, 1992) conducted an ethnography of three working class neighbourhoods in Belfast along with sociolinguistic data collection. Based on findings from first wave studies, it was predicted that men would use more vernacular speech features than women, and women would use more standard speech features than men. In regard to one speech feature studied, this prediction held true for the Ballymacarrett neighbourhood but not Hammer, and in Clonard the pattern was the opposite: women used the vernacular speech feature more than men, and men used the standard speech feature more than women. This related to employment. There was a lack of jobs in Clonard so many men had to find work elsewhere, spending less time in the local area and loosening their connection to it. In contrast, the women in Clonard both worked and socialised there. Thus, use of the vernacular speech feature reflected the women's local identity of being strongly tied to Clonard. Thus, speech variation reflected local identities, not the macrosocial category of sex, and this was only revealed by spending time in the communities, rather than surveying them.

The third wave began to emerge in the 2000s, with its establishment confirmed through Eckert's writings (e.g. 2012, 2016). The third wave sees more retheorizing of how and why certain ways of speaking are associated with certain groups of people. Unlike the first and second waves, which saw variation as a consequence of social categories and hierarchies, the third wave attributes further agency to speakers. The view is that speech features gather meaning through speakers repeatedly using them when taking a certain stance during an interaction (called 'stance accretion' according to Dubois (2002) and Rauniomaa (2003), both referenced in (Eckert, 2012)). An illustration (author's own) that uses gesture instead of speech is:

> A teenage girl is being told off by her parents. At one point, she flicks her hair from her collarbone over her shoulder. A few days later, a teacher at school attempts to discipline her. She performs this same action, flicking her hair over her shoulder. Her peers may witness this interaction, with the hand gesture being a salient feature, and then use it themselves in their own interactions when taking a similar stance. Thus, by repeatedly using the same gesture in different interactions where a similar power dynamic is present and the user is taking a similar stance towards their interlocutor, the gesture becomes associated with defiance to authority. Her peers taking this up in their own interactions allows the gesture to spread, and if they also use it in interactions where they are being defiant to someone in authority or those being authoritative over them, then the meaning associated with the interaction will also spread.

In the third wave, variation is hypothesised to come from 'stylistic practice' (the interpretation and the production of styles (Eckert, 2008)), through which variation comes to reflect social identity over time. The hair-flick-over-shoulder gesture may be one feature in a repertoire that becomes associated with authority-defiant teenage girls. This associating is 'enregisterment': when one (e.g. a gesture) or more acts become differentiable from others and socially recognised as belonging to a population (Agha, 2003). Thus, taking our gestural illustration above, if flick-hair-over-shoulder distinguishes authority-defiant-teenage-girls from others then the gesture could either be: a) used by others to

stigmatise this group and the additional meaning of stroppiness invoked; or b) used by others to invoke the qualities that this group is admired for, which could be sassiness. These two scenarios hint at how the meaning of speech features is mutable and multiple, not fixed as in the first and second waves (how a speech feature can have multiple meanings will be explored further in the next section 2.1.3). Also, scenario (b) is an example of "bricolage": taking a speech feature and recombining it with others to re-inscribe meaning. Others using the authority-defiant-teenage-girls gesture of flick-hair-over-shoulder in their own interactions are taking a feature from the authority-defiant-teenage-girls register, demonstrating that "[r]egisters are both an important source of stylistic resources and a potential end product of bricolage" (Eckert, 2012, p. 96).

2.1.3 The Indexical Field

The third and final key concept from sociolinguistics that is essential to this thesis is 'the indexical field'. Contemporary sociolinguistic research has realised that the meaning of linguistic variables is underspecified, multiple and mutable. In other words, that we should not view the relationships between speech features and social categories as fixed. A speech feature may clearly index an accent for a listener at the time of hearing it, but it might not do so for all listeners or for all time. Specifically, "[v]ariables do not have static meanings, but rather general meanings that become more specific in the context of styles" (Eckert 2012, p. 453). Penelope Eckert, in her seminal piece (2008), coined the term "the indexical field" to describe the mapping out of the flexibility and multiplicity of a variant's meaning, which we explore in more detail now.

As first wave and second wave studies evidence, speech features can reflect social identity. A key contribution of third wave work is not only the theory of the processes by which this association occurs (as outlined above), but also that associations between speech feature and social identity are *indirect*. In other words, if it were not for a mediator the speech feature and social identity wouldn't be connected. It is hypothesised that the mediators between the speech feature and the social identity are *qualities and stances*.

Returning to Eckert's work of white adolescent speech in a high school in a Detroit suburb, the suburban living but urban-oriented and school-alienated kids (known as Burnouts) adopted some of the speech features found in urban Detroit. She argues:

> "The urban kids that [the Burnouts] identified with were white kids who knew how to cope in the dangerous urban environment – kids they saw as autonomous, tough, and street-smart. Presumably in adopting urban forms, suburban kids were *affiliating with those qualities* [autonomy, toughness, and street-smart], not claiming to be urban." (Eckert, 2008b, p. 459, my emphasis)

The mediator between the social identity of 'urban white Detroit kid' and certain speech features were the qualities of autonomy, toughness, and being street-smart. Therefore, the Burnouts (suburban living but urban-oriented kids) adopted these speech features because they wanted to indicate to others that they were autonomous, tough, and street-smart, not that they held the social identity of 'urban white Detroit kid'. Similarly, revisiting the gesture example given above, the hair-flick becomes associated with the stance of defiance towards authority. Those that admire this quality of the authority-defiant teenage girls may affiliate the gesture with the positive quality of sassiness, and thus adopt or recruit the gesture into their own repertoire of resources for their own interactions. Further, an important point to emphasise is that the meaning of speech features is activated when in use in an interaction. That is, it is only through who is using the speech feature (social context) and what stance they are taking in an interaction that one of the speech feature's meanings, from the multitude that are available, is invoked. For example, if the hair-flick gesture was used by teenage boys (and so might become more of a shoulder brush) it is unlikely that the gesture's meaning of 'sassiness' would be activated and perhaps more likely the quality of 'toughness' might become affiliated with the gesture and thus activated in future interactions. Further, if the gesture was to be used by a parent or teacher, it is unlikely that the qualities of sassiness or toughness would be activated either. Rather, because of the relationship dynamic between and the relative social stances of authority-defiant teenage girls and those in authority, the gesture could become affiliated with 'stroppiness' as a result of its use, by

teachers or parents. Thus, this meaning could be activated when used in future interactions. Further, one can imagine that the use of this gesture by parents or teachers would have to be in a sarcastic or mocking tone. These descriptions also illustrate that the indexical field of a speech (or gesture) feature is fluid; "each new activation [use of the feature by someone when taking a stance in an interaction] has the potential to change the field by building on ideological connections" (Eckert, 2008, p. 454).

A final, important concept in regard to indexical fields is Silverstein's notion of indexical order (2003). Simply, different speech features have different degrees of social saliency and this social saliency contributes to defining the speech feature's indexical field of meaning. Here, the term 'variant' will be used to refer to a speech feature (a full description of a variant will be given later on, in section 7.1.1). For example, a first-order index variant is one that is associated with membership to a social group but has not attracted any attention and so is not open to be discussed by non-linguists. This lack of social saliency means it is a less rich resource for making meaning in an interaction. Hence, the indexical fields of first-order index variants are restricted and small, relative to second-order speech features. Second-order variants are socially salient speech features and the social meaning affiliated to them can be activated in interaction, and thus it and its indexicality is available for continual reinterpretation and redefining. It is possible for first-order variants to become second-order variants because social saliency is always in flux and how a social group is evaluated changes over time. Note, second-order index variants can also become first-order index variants again through the reverse of this process, with variant's social meaning falling out of use. While their numbering implies linearity, this would actually contradict the fluidity of the indexical field. Rather, rationalising variants as either having a 1st indexical value, or a $n + 1^{st}$ indexical value would be more accurate.

We will revisit the gestural example above to illustrate these concepts. The writing of the scenario in section 2.1.2 makes it appear that when the young teenage girl who was being told off by her parents flicked her hair that this was the first time that she had used that gesture. Also, the apparent spread of the

gesture amongst her friends and then wider peer group further suggested that they had never used this gesture before either. The impression given in this scenario was that this gesture was something new. However, this is in fact highly unlikely. It is more likely that this social group was already regularly using this hair-flick in their interactions, but it wasn't noticeable and, in tandem, carried little meaning. This would be an example of a first-order index speech feature.

In contrast, if the hair-flick-over-shoulder gesture was recognised as belonging to authority-defiant teenage girls and differentiated from other gestures that this group may use, then it is a second-order index variant. Further, this gesture's socially saliency means that it can be used to mean 'defiance-against-authority' in interactions. Plus, its adoption by teenage boys to communicate toughness, and parents or teachers using it in a derogatory way to communicate 'stroppiness', illustrates that the hair-flick gesture and what it indexes is open to continual reinterpretation and redefining. Thus, the scenario in section 2.1.2 is an illustration of how a first-order index variant may become a second-order index variant.

While one reading of Eckert (Eckert, 2008) can focus on the indexical field as conceptual (ideological), it may also be rationalised as primarily physical. In most of the examples she gives, a 'here' versus 'there' distinction is paramount. Zhang's (2008) work in Beijing compared the speech of managers in foreign-owned businesses with that of managers in state-owned businesses. As described above, Eckert's (1989) ethnography in a Detroit high school found that speech differences reflected the opposition between city and suburb orientated social groups. Similarly, Labov's (1963) participants' speech indicated affiliation with either the traditional island economy or the contemporary island economy that was dependent upon the mainland. Speech is symbolic; it doesn't inherently carry meaning thus it cannot be understood in isolation. We can only understand the indexical meaning of speech when used in one context by comparing it to its use in another context. Here, the term 'context' is used to refer to the physical, but social and temporal contexts overlay also, of course. If we return to thinking of the indexical field as primarily physical, Eckert (Eckert, 2008) explains:

> "meaning is based in ideologies about what the locality is about – what kinds of people live there and what activities, beliefs, and practices make it what it is. Local identity is never an association with a generic locale but with a particular construction of that locale as distinct from some other. Local identity claims are about what it means to be from 'here' as opposed to some identified 'there.'"

Therefore, the locality of speakers and listeners is fundamental to the defining of their indexical fields. Locality provides access to the symbolic material (speech) and a knowledge of its prior and current social meaning. Further, as was explained above, it is not merely that the speech type exists within someone's realm of experience, but that the listener is able to conduct a similar social evaluation of it as is conducted by those around them so that they tap into the same notions of indexical order. In the same breath, this allows one to envision another's indexical field.

2.2 Researching Language Online: A brief history

In regard to researching language in CMC, there are two technology-related traditions that should be considered. The first will be referred to as 'Sociolinguistic CMC' that, it can be argued, was first defined by Androutsopoulos (2006a). The second is 'Computational Sociolinguistics', its name reflecting its origin in Computational Linguistics (Nguyen *et al.*, 2016). A brief history of each of these is provided in order to contextualise the literature review that follows in section 2.3.

2.2.1 Sociolinguistic CMC

When overviewing linguistic studies in CMC, two waves of analytical practice can be identified. Androutsopoulos (2006a) describes the first wave as focused on technology producing new language varieties. The 1990s saw linguistics

researchers defining, for example, the language of emails, the language of online chatgroups, the language of texting, etc., as distinctive varieties. Thus, the goal of a taxonomy began to emerge, with classification being determined by the technology's features (most pivotal being whether the communication being facilitated was synchronous or asynchronous) and the apparently technologically determined language features used (e.g. abbreviations such as 'lol' and 'brb', and emoticons). For examples of this work, see Susan Herring's (1996) edited volume.

It is not hard to view this approach as reminiscent of linguistics prior to sociolinguistics emerging (as described in section 1.3) in that any variation within each CMC's language variety would be viewed as errorful deviation from the norm or 'noise' in the data. Of course, any heterogeneity is more likely to actually reflect the diversity in the people using CMC technologies, the purposes for using them, and the social contexts in which their use occurs. And just like in sociolinguistics, there was a paradigm shift towards not only acknowledging but also integrating variation in the analysis.

The second wave of language studies in CMC was a shift towards considering how users harness the affordances of multimedia for different interactional purposes (Androutsopoulos, 2006a). That is, where once macro factors (the technology) were at the centre of an enquiry, now the user and their community was. This refocusing was synonymous with the taking up of discourse analysis as the main investigative method (the framework for which being defined by Susan Herring (2004)). Thus, from the mid 2000s, discourse analysis dominated sociolinguistic investigations of CMC (see Androutsopoulos, 2006b; Thurlow & Mroczek, 2011). Subsequently, the methods used diversified to include other qualitative practices such as conversation analysis, semiotics, and ethnography (Thurlow and Mroczek, 2011; Herring, Stein and Virtanen, 2013; Tannen and Trester, 2013; Georgakopoulou and Spilioti, 2016).

However, the 'waves' of researching sociolinguistics in CMC can be restructured based on the technology investigated. This, evidently, has been a somewhat symbiotic relationship. As the technology that can be studied has

evolved, the sorts of research questions that are asked have surfed this wave of change. In the first wave, in the late 1990s and early 2000s, the main contexts studied were blogs and one-to-one communication, such as email and text messaging. The second wave moved on to researching social media, with the amount of work increasing greatly in the latter part of the second wave after interest and use of the social network site Facebook exploded (e.g. Dovchin, 2015; Seargeant, Tagg, & Ngampramuan, 2012; Sharma, 2012). Finally, as the research field of computational sociolinguistics (described next in section 2.3.2) has developed, the focus has moved to Twitter (Huang *et al.*, 2016; Grieve, Nini and Guo, 2018; Strelluf, 2019).

2.2.2 Computational Sociolinguistics

Computational Sociolinguistics evolved separately from Sociolinguistic CMC. Computational Sociolinguistics is

> "the emerging research field that integrates aspects of sociolinguistics with computer science in studying the relation between language and society from a computational perspective" (Nguyen *et al.*, 2016, p. 540).

There are two separate motivations that have led to computational sociolinguistic work. The first is to accommodate the social dimension of language, and not just the informational dimension, in natural language processing in order to refine models and improve their performance. Here, the focus is on maximising the model's predictive accuracy – its ability to predict the social demographics of the person who wrote the text. The second is the use of computational linguistic techniques for the processing of very large datasets to answer sociolinguistic questions and, to quote Huang et al. (2016, p. 254), "examine the dynamics of linguistic characteristics and their spread at finer spatial–temporal resolutions". Here, the focus is on maximising the interpretability to the model – its ability to explain what combination of social factors lead to what kind of language (Nguyen *et al.*, 2016). Much of the literature in section 2.3.1 can be classified as work performed under this second motivation.

From examining the literature, it appears that the second motivation branched out of the first, and a key factor in this development was the increasing use of social media as data. At first, computational linguistics used corpora of formal, non-CMC texts, and then moved on to blogs. However, these data sources are not interaction-rich – there is minimal to no turn-taking or exchanging of utterances. When computational linguistics moved on to using social media data as the corpora for modelling it signified a move towards incorporating a more social perspective of language into this work, rather than minimising it as many would argue the use of these previous data sources did.

Thus far, the social factors of gender, age, and location have dominated computational sociolinguistics, reflecting the "first-wave" approach of sociolinguistics. Thus, the same critique of first-wave-like computational sociolinguistics can be made of first-wave sociolinguistics: it is questionable "the extent to which the socio-demographic factors inferred from metadata can be used to explain socially meaningful patterns of […] variation" (Ilbury, 2019, p. 4).

2.3 Researching Language Online: Literature Review

Now that a brief overview of the history of researching language in CMC has provided context, literature that is more directly related to the thesis topic can be reviewed. In order to navigate such varying fields of enquiry, the literature review below is broadly structured as follows: 2.3.1 Quantitative Written CMC; 2.3.2 Qualitative Written CMC; 2.3.3 Video and Audio. It is through this simple categorising that themes in regard to research interests and methods can be clearly delineated, and thus gaps, and initial rationales for why these gaps have remained, identified. Studies that use multiple materials or sources of data and thus multiple research methods, also exist. Here, these will be categorised based on the topic of interest that they make the greatest contribution to. Finally, it is

worth noting that not all language focused qualitative studies of written CMC are reviewed, but only the papers whose primary contribution is to sociolinguistics.

2.3.1 Quantitative Written CMC

As the literature review below will reveal, in quantitative research on language variation and change in CMC, three main topics have developed: i) Language use in CMC contexts – mostly focusing on how language is used in different online medias (section 2.3.1.1); ii) Social media as a proxy for offline communication – the rationale being that online language will reflect certain features of offline language to a degree (section 2.3.1.2); and iii) Comparing CMC and offline language – the aim being to identify if language use differs across mediums and, if so, how and why (section 2.3.1.3).

2.3.1.1 Language Variation in CMC Contexts

When language in CMC first began to be studied the focus was on the resources different people used in different online contexts. It was observed, for example, that emoticons are used more often by females on online message boards than by males (Witmer and Katzman, 1997), and by teenage males on their blogs more than teenage females (Huffaker and Calvert, 2005). Siebenhaar's (2006) mixed methods study examined dialect usage in Swiss-German internet relay chat channels. The language variety used differed according to the topic of the internet relay chat channel; those that were region focused (e.g. "#bern", "#zuerich") were dominated by the corresponding dialect. Similarly, Androutsopulos and Ziegler (2004) observed the use of dialect speech features for "nicht" (*not*) continuously increased in German region internet relay chats (e.g. "#mannheim", "#bremen") from North to South. Some contemporary research continued this trend. For example, Sali Tagliamonte (2016) studied how young people's use of i) acronyms, short forms and initialisms (e.g. 'lol', 'tho' for 'though', and 'ppl' for 'people', respectively), ii) intensifiers (e.g. 'very', 'really' and 'so'), and iii) future temporal references (e.g. 'will', 'ill' for 'I'll',

and 'gona') varied across email, instant messaging and phone based texting. The patterns of use were found to be consistent across the three registers.

The findings from such work soon began to demonstrate that language use is tailored to audience. Pavalanathan and Eisenstein (2015) found that the smaller the target audience, the more nonstandard lexical features were used, both in nonstandard American English variables and nonstandard regional variables. It is theorised that the nonstandard variables (being more geographically specific) reflect the narrowing of the audience, whereas more standard terms are used to be inclusive to a more geographically diverse audience. Similarly, Shoemark and colleagues (2017) argue that in tweets from Scottish users less local terms were used when talking to a broader audience. Scottish terms were used more by users of pro-independence hashtags, but overall Scottish terms were used less in Scottish referendum hash-tagged tweets than in general tweets. Teresa Gil-Lopez et al. (2018) considered if there was a relationship between network size and language style variability in status updates on Facebook. Language styled variability was determined by examining the function words (e.g. pronouns and prepositions), words that indicate cognitive processes (e.g. "think", "know"), and indicators of informal language (e.g. swear words, fillers like "um") used in status updates. They found that language style variability was negatively associated with the size and heterogeneity of the network across a 12 month period. This indicates that "people manage their online self-presentation in ways that are consistent with lowest common denominator" (ibid, p. 127). Paolillo (2001) looked more closely at the strength of the relationship between the message sender and receiver in internet relay chats. He conducted a network analysis to test Milroy and Milroy's (1985) network theory of language change within an online context. The network theory of language change posits that i) dense networks are resistant to change and most linguistic change is initiated by weak links, ii) vernacular forms correlate with network density, and iii) standard forms are found in networks with weaker ties, indicating how local speech varieties endure regardless of being stigmatised. Unexpectedly, the distribution of the vernacular speech features 'r' and 'u' (as in "r u ok?") was as was predicted for the standard variables ("are you ok?"). That is, they were used when the network tie between the message giver and receiver was strong.

Paolillo posits the theory that their use is now sufficiently widespread that they can be viewed as 'standard' in internet relay chats rather than being innovative 'vernacular' features. Finally, Miriam Hansen et al. (2015) conducted an experiment to see if people aligned the linguistic content of their email to the email that they were responding to. They found that participants aligned their communication style in their response email to the communication style (Western vs Asian) used in an email from a peer asking for help. Also, the ethnicity cue (German vs Chinese name) influenced the wording of their response, their perception of the sender's personality and their willingness to help.

In three of the studies above (Pavalanathan and Eisenstein, 2015; Shoemark *et al.*, 2017; Gil-Lopez *et al.*, 2018) the fact that social media platforms are *networks,* where one can follow others and be followed, is fundamental to the audience-related behaviour. While these authors make no specific claims in this regard, if and how the design of the technology influences one's use of language, and thus potentially language variation and change, has just begun to emerge as a topic of interest when researching language in CMC. For example, Bohmann (2016) found that the restriction of the number of characters in tweets on Twitter (densification) encouraged *because*-complementation (e.g. "Early morning gym because fat", example 1b p.149). Thus, the design of the platform acted as a catalyst or driver in a linguistic change. To quote danah boyd (2008, p. 93), "Computer code does not determine practice, but as a form of architecture (Lessig 2006) it does shape the way in which people can interact." Thus, when conducting sociolinguistic research in CMC whether certain technological features encourage or discourage certain linguistic practices should be considered.

2.3.1.2 Social Media as a proxy for Offline Communication

From the latter half of the 2010s until now, a dominant type of study is examining offline language variation and change via social media, predominantly Twitter. It can be argued that Delia Mocanu et al.'s (2013) "Twitter of Babel", which clearly laid out the potential of Twitter data through

analysis at different geographical levels (country, region, and city) and over time (seasonal variation) was a moment of legitimisation for this kind of work. As is outlined in many of the papers in this topic area, to conduct similar work via the traditional means of a linguistic fieldworker surveying representative speakers would be far more resource intensive. In regard to variation, Huang et al. (2016) examined regional linguistic variation in the form of lexical alternations in the US; Gonçalves and Sánchez (2014) defined different varieties of Spanish across the globe; and Strelluf (2019) looked at regional variation in the use of positive "anymore" in the USA. However, only recently was it confirmed that the regional variation found on Twitter broadly aligns with offline variation (survey data) (Grieve *et al.*, 2019). Expanding the use of such data further, Coats (2016) questioned whether specific varieties of English online differ from English on Twitter in general. He collected English tweets that were geo-tagged within Finland and compared the frequency of specific grammatical features in these tweets to English, non-geotagged tweets from an established corpus, finding differences in the use of a range of features that clearly distinguishes English from Finland on Twitter, and global English on Twitter.

A key theme within this dominant study type is examining the representation of pronunciation through orthography as a proxy for speech. Specifically, this refers to the theory that users adjust spellings to reflect their accent and, to a degree, encourage the reader to pronounce said word in a certain way. Both Jones (2015, 2016) and Eisenstein (2015) mapped their Twitter data onto the United States Census. Both argue that the use of nonstandard spellings that reflect African American Vernacular English[1] speech correlates with areas in the US that have a high proportion of African Americans. Jones (2015) looked at features such as th-fronting (e.g. "something" becomes "sumfin" and "brother" becomes "bruva"), and ey-raising ('yeen' for 'you ain't' and "reenin" for "raining") whereas Eisenstein (2015) looked at g-deletion ('ing' to 'in' in verbs, e.g. 'walking' to 'walkin') and t/d deletion (e.g. 'just' becomes 'jus' and 'passed' becomes 'pass'). Callier (2016) extended this type of study further by examining how these nonstandard spellings covaried. Callier reported an

---

[1] An accent and dialect associated with working class black Americans.

analysis of tweets that used the standard/nonstandard spellings 'this/dis', 'that/dat', and 'they/dey' to convey DH-stopping (when the fricative /ð/ is produced as the plosive [d]), and their covariance with other orthographic representations of pronunciation. Most interesting is how, through a multiple correspondence analysis, the three DH-stopping keywords ("dis", "dat", "dey") showed different and highly distinctive co-occurrence patterns with the other variables such as R-lessness [ɹ] (deletion after a vowel so "where" becomes "wea"), and the presence/absence of internet initialisms (e.g. 'OMG', 'lol'). Callier proposes that this may indicate these three forms occur in different communicative situations.

Insights into the processes via which linguistic change occurs has also been gained. For example, both Grieve et al. (2018) and Eisenstein et al. (2014) tracked lexical innovation. Both identified a number of words (e.g. 'baeless' meaning 'to be single' and 'tookah' meaning 'marijuana'; the acronym 'ctfu' which stands for 'cracking the fuck up' (laughing); respectively) whose use dramatically increased during their dataset's timeframe (1 year for Grieve et al. (2018), 3 years for Eisenstein et al. (2014)) and then tracked their use over time and (geographic) space.

A number of themes are evident from this body of work. First, these studies exemplify elements of computational sociolinguistics: large datasets (with the studies referenced here having datasets of tweets between 12,273 (Jones, 2015) and 5×10^7 (Gonçalves and Sanchez, 2014), prepped using computational methods (mostly part-of-speech tagging), and often using complex statistical analyses (e.g. principal component analysis, multiple correspondence analysis) are three distinct features. Second, the importance of geo-tagging in such work is clear, as it is through this metadata that patterns in the data can be interpreted.

It should be noted that, unlike the prior wave of Sociolinguistic CMC research, these studies minimally consider demographics other than location. Unlike automatic geo-tagging, user information is inputted or declared by the user themselves, thus there is less certainty over the accuracy of this metadata. There are other ways of inferring such demographics, such as the user's avatar or photographs that they share of themselves (Jones, 2015). But this would require

engaging with every user's timeline that is in the dataset, an approach that would counter all the advantages of big data and computational methods. This raises problems for traditional variationist methods which assume that reliable information about participant gender, age, social class, race, […], etc., is available to the researcher' (Herring, 2001, p. 621). In the studies above where it was found that Twitter data correlated with census data about race, this issue is skirted by Jones (2015) who rationalises that the subject of study is African American Vernacular English use, not race. Finally, just because a tweet is geo-tagged in a certain location does not mean that the tweeter uses a linguistic variety that is representative of that location (e.g. university students studying in cities that are not where they grew up).

However, the tide has begun to turn on the assumption that by analysing online data at a macro level we can observe offline language variation and change. Bamman et al. (2014) took a more nuanced approach to gender and, through clustering topics of interest and linguistic style, found that individuals who deviated from population-level gender patterns had online social networks that included significantly fewer same-gender connections. Thus, online network homophily correlated with the use of same-gender language marker. Further, Ilbury (2019) found that stylistic spellings associated with African American Vernacular English were frequently used in tweets by 10 White gay men from the south of the UK. These included orthographic representations of speech features such as: 'hurr' for 'hair'; g-deletion in verbs so 'sipping' becomes 'sippin'; DH-stopping ('that' to 'dat', 'they' to 'dey', 'this' to 'dis', 'them' to 'dem'), and some which are ambiguous in regard to their intended pronunciation such as 'gurl' for 'girl', 'werk' and 'werq' for 'work, and 'fuq ' for 'fuck'. He argues that these non-standard spellings are ideologically associated with stereotypes of "sassy" and "fierce" Black women, qualities that are appreciated within certain subcultures of the gay community. Thus, the use of such features cannot be explained by applying macro-social categories (demographics such as age, location, gender), but can be when the personas that these young men might want to index is considered.

Finally, as is also indicated by Ilbury (2019), researchers may have begun to be less interested in broad patterns and more interested in the individual user, moving from a very small amount of data (possibly even just one token plus metadata) from thousands or even millions of users to being focused on a lot of data from a few individuals. Notably, Clarke and Grieve (2019) focused on linguistic change and how this contributed to different styles in 21,739 tweets from one particular person: Donald Trump. Thus, focusing on just one person does not necessarily lead to less data or relinquish the need for computational methods or complex statistical analysis.

2.3.1.3 Comparing CMC and Offline Language

In comparison to the first theme of examining offline language variation and change via CMC, a second significant theme of research is comparing CMC and non-CMC language. This work either effectively verifies the first theme of work (that CMC can be used as a lens onto offline communication) or identifies the reasons why CMC and non-CMC language can differ.

The results of several studies support the premise that CMC language can be used as a relatively accurate lens on non-CMC language. As one of the first studies to compare offline language to CMC language, Tagliamonte and Dennis (2008) analysed four grammatical features in instant messages and speech data from the same participants over two years. The instant message variation and change over time reflected that of offline variation and change and thus was deemed to be a part of a much broader trend of language variation and change. LaFave (2016) examined English adjective gradation. Two data types (i. Instant messages, ii. Spoken) were collected from several corpora and examined in regard to the influence that linguistic and social factors had, particularly on synthetic (e.g. "old", "older", "oldest") versus analytic (e.g. "beautiful", "more beautiful", "most beautiful") adjective gradation. But, through various statistical analyses, LaFave reveals that there is no statistically significant difference between instant messenger and speech data. And Wieling et al. (2016) studied hesitations ("um" and "uh") in American English and Dutch in spoken and Twitter corpora. They argue that the results patterned similarly, including by the

social factors of age and gender, taking into account the register difference of speech vs writing.

However, multiple studies have also found a difference between CMC and non-CMC language. Becky Childs (2016) analysed instant messaging and spoken data collected from the same participants from an Appalachian African American community. She found the instant messaging and spoken data to be distinct. Postvocalic /r/ was used more, and expletives and slang terms were used less in the spoken data, with the opposite pattern (less postvocalic /r/, more expletives and slang) used in instant messaging. The lack of overlap in using these features was due to identity construction and performance being different across these contexts, with instant messaging being the main avenue for engaging with the broader African American community and speaking being for engaging with the local Appalachian African American community. Thus, for example, words such as 'holler' were spelt without the 'r' that came after the vowel in the unstressed syllable ('holla'). This spelling aligned more closely to the r-less accent of the broader African American community (who were contacted via instant messenger) than the r-full accent of the local Appalachian African American community. However, this alternate spelling did not occur in all possible instances suggesting it was used strategically on particular words in particular exchanges.

In a similar vein, Nadine Chariatte (2015) studied speakers of Spanish in Malaga. Broad transcriptions of corpora speech were compared with written Facebook status updates and comments, and answers to a specially designed survey supported the interpretation of the data. More nonstandard features were used in the Facebook data, reflecting a specific style, she argues. Further, social factors appear to behave differently in these two datasets; middle-aged men use the most nonstandard features in speech, while young women use the most nonstandard features on Facebook. However, what orthographic constructions reflect what speech features is not clarified. Further, as previously mentioned, Grieve et al. (2019) found British English dialect variation on Twitter broadly aligned with offline survey data.

2.3.2 Qualitative Written CMC

The dominant theme when examining CMC language using qualitative techniques has been multilingualism and how different languages are used. Androutsopoulos (2015, p. 185) proposed the term 'networked multilingualism' to categorise much of the participant behaviour that was observed in this work:

> "multilingual practices that are shaped by two interrelated processes: being networked, i.e. digitally connected to other individuals and groups, and being in the network, i.e. embedded in the global mediascape of the web".

Androutsopoulos emphasises that 'networked multilingualism' is a broad, encompassing term. As was evident in his four-week long ethnography of the Facebook pages of secondary school students with a Greek-background living in Hamburg, "networked multilingual practices are individualised, genre-shaped, and based on wide and stratified repertoires" (ibid, p.185).

Sharma (2012) and Seargeant et al. (2012) both found code-switching (switching between languages) and code-mixing (two languages together in the same utterance) that differed from what is expected in face-to-face interaction. In Sharma's (2012) two year ethnography of three Nepalese undergraduate students on Facebook, he found that participants mixed two languages to construct bilingual identities. While most of the actual social relationships of the participants were local, they constructed cosmopolitan affiliations and identities (real and aspirational) through the mixing of English and Nepalese in innovative ways compared to face-to-face interactions in Nepal. Seargeant et al. (2012) also found their participants performing complex code-switching on Facebook, despite Thai being the default choice in face-to-face interactions. They specifically emphasise addressivity: who the participant intended to address. Participants developed and maintained multiple, separate lines of conversation under the same status post/update. Here, it was not simply that choosing English over Thai or vice versa indicated the community being addressed; English was used as a resource for orienting to the local (UK) sometimes, but to the global at

other times. Similarly, Dovchin's (2015) 40 university students on Facebook in Mongolia recombined linguistic, cultural and semiotic resources from English, Russian, Japanese, Korean and Turkish regardless of the cultural pressures of "Linguistic Dystopia" – the belief that the use of foreign languages within local Mongolian contexts is either endangering the Mongolian language or perceived as being inauthentic. In fact, rather than defying or clashing with the ideology of linguistic dystopia, these young people use translingual practices to claim their own authenticities, that are both multiple and coexisting, metalinguistically. How they do this depends on their individual beliefs and identities, and thus vary greatly.

Multilingualism within the context of initialising interaction in social media is particularly interesting. Seargeant et al. (2012, p. 519) found that "Code-mixing tends to occur, although not exclusively, where the initial post is directed at a specific individual." Rather than code-mixing, Androutsopoulos (2014) found three other strategies that his participants (five teenagers, all living in Germany, two born to Greek parents, two born to Taiwanese parents) used to maximise their audience for initial contributions. The first was to choose the "common-denominator language", in other words the lingua franca of their network. The second was replicating the content in two languages or more, and the third was to partition the audience by posting in a language so that only those who are competent in that language can engage. They argue that the tension between intimacy and publicness heightens metalinguistic awareness of the language options available for delimiting the audience. Finally, Christiansen's (2015) work looked at how code-mixing and switching was used in identity construction. The research was an ethnography on Facebook of five participants who either lived in Chicago, US and spent holidays in Mexico or vice versa, all with family in Mexico and socialised in *ranchero* culture. *Rancheros* "are a subpopulation of Mexican peasants considered with some ambivalence by the larger Mexican society" (Christiansen, 2015, p. 689). Christiansen found that through the strategic use of Standard English, Mexican Spanish and *Ranchero* Spanish (a stigmatised variety) the participants constructed a transnational identity. Most importantly, participants detached themselves from negative *Rancheros* stereotypes by using *Ranchero* Spanish in a mocking or ironic way.

2.3.3 Video and Audio in CMC

To date, a sociolinguistic perspective in researching video and audio CMC is rare. Key themes of work to date have been conversation analysis of multi speaker audio chats (e.g. Brandt & Jenks, 2013; Hung, 2017) and video conferencing (e.g. Santos Muñoz, 2016) and speech and language in online video with qualitative analysis (e.g. Tolson, 2010; Pihlaja, 2011; Georgakopoulou, 2015; Mendoza-Denton, 2016). Automated analysis of speech in online video has predominantly been used to classify speakers (e.g. Biel, Tsiminaki, Dines, & Gatica-Perez, 2013), speaker behaviour (e.g. Park, Shim, Chatterjee, Sagae, & Morency, 2014) or video content (e.g. Biel & Gatica-Perez, 2011), the intention being to assist in the automatic description, categorisation and organisation of online videos. A rare example of automated analysis being used with the intention to consider the speech itself is Coat's analysis of speech rate in regional varieties in the USA (2019). In contrast, non-automated analyses have begun to emerge in student work, notably Sarah Lee (2017) and Kelsey McDonald (2018). Both these studies examined how speech varied across formal / planned and informal / spontaneous videos (this work will be discussed in more detail in section 2.6.4).

The majority of sociolinguistic studies that have engaged with online videos have focused on analysing the comments that are left by viewers and examining the metalinguistic commentary they contain. Thus, the speech produced in the video data is only analysed to the degree that it provides reference through which the content of the comments can be understood. Further, most of these video data have not been produced for the purposes of YouTube but were originally broadcast as a part of a television program (Ivković, 2013; Aslan and Vásquez, 2018), a television advertisement (Jones & Schieffelin, 2009), or in films (Cutler, 2016). There are two exceptions. First, Betsy Rymes and Andrea Leone-Pizzighella (2018), who analysed "Accent Challenge" videos: a type of video where one would film themselves reading out a series of words or responding to prompts to produce specific vocabulary. Second, Rachael Tatman (Tatman, 2017) directly compared the use of New York English speech features

by a sports announcer (Mike Francesa), and by a fan (Mike Zaun) while mimicking Francesa in a parody video and in spellings on Twitter. Four features were examined, two of note being i) R-lessness - [ɹ] deletion after a vowel (so "we're" becomes "weah", and "air" becomes "aih"), and ii) DH-stopping (when the fricative /ð/ is produced as the plosive [d], so "that" becomes "dat"). Zaun used [ɹ] deletion and DH-stopping at a higher rate than Francesa both in the parody video and on Twitter suggesting it is a key part of his performance.

One piece of research, although no linguistic analysis was conducted, would be particularly relevant to sociolinguistics studies of YouTube: Sophie Bishop's (2019) observations of "vlogging parlance". Bishop (2019) argues that YouTube's use of auto-generated caption data in their search algorithm encourages several YouTuber behaviours that are used in the hope of creating caption data that results in their video receiving greater visibility. An illustrative example will assist here. A YouTuber videos a haul of winter coats. Upon uploading the video the closed captions are automatically generated. While some of the times that the YouTuber says 'coats' is correctly transcribed, at others the caption generated is 'cots' or 'corks'. Now we transition to a user who wants to watch a video of a haul of winter coats. They use the search function, entering "winter coat haul". From the potentially hundreds of thousands of videos about winter coats existing on YouTube, how does the system decide which to display on the first results page? The YouTube system categorises videos into topics based on keywords found in various data such as video titles but also the content of the closed captions. Crudely, one can assume that the more times a keyword appears the more likely it will be one of the first results displayed when a user searches using that keyword. Therefore, YouTubers believe that the more times the keyword is used in their video, and thus appears in the closed captions, the more visible their video will be. This leads to a behaviour Bishop (2019) describes as 'keyword stuffing', repeating the video's keywords many times so it appears in the captions many times. But, more relevant to sociolinguistics, is how a YouTuber may alter their pronunciation for the accuracy of the auto-generated closed captions dictates how many times the keyword will appear. For our YouTuber who videoed the winter coat haul, how visible will her video become considering the YouTube system thinks her video is about 'cots' and

'corks' as well as coats. This can lead to YouTubers "carefully and crisply pronouncing keywords" (ibid, p.27), a behaviour that sociolinguistic researchers will need to consider carefully.

Finally, it should be acknowledged that it could be argued that work around voice user interfaces, such as Apple's Siri (Apple, no date), could be classified as researching audio online because of its utilisation of cloud technologies. However, in this thesis the term 'online' is used with connotations of publicness (known or unknown others can act witness to the communication during the original event or afterward via a recording) and so literature on voice user interfaces will not be reviewed. Further, the same rationale is made in regard to synchronous voice CMC technologies, such as video conferencing, and this literature is not reviewed either. For a summary of synchronous voice-based CMC see (Jenks and Firth, 2008).

2.4 Summary

To summarise so far, the first half of this chapter has outlined the key sociolinguistic concepts of the vernacular and the indexical field, and has described the three waves of analytical practice. Two traditions of research that have evolved in parallel but very separately, Sociolinguistic CMC and Computational Sociolinguistics, were then described. In regard to the research of speech in CMC, the findings from the literature review are aptly summarised by Androutsopoulos (2006a, p. 425):

> "[r]elatively few studies of language use in CMC are based on quantitative methodologies [...], and even fewer make an explicit connection to variationism [...]. [This is perhaps partly] due to the absence of the main type of linguistic variable in the correlative paradigm, that is, phonetics/phonology.".

However, the latest papers demonstrate significant progress has been made in using CMC data in the variationist paradigm, primarily as computational

methods have been applied with sociolinguistics in mind and as relevant data has become relatively easily available, mostly in the form of Tweets.

This literature review leads to the question of why; why is there a paucity in studying speech online? Another notable finding is that when categorising the research papers either by interests and practices, the type of technology (e.g. blogs, Facebook, Twitter) or over time, a very similar pattern is revealed. This indicates that these three aspects are interconnected or co-dependent; as new forms of CMC are designed and/or become popular researchers may identify them as sources of interesting information to explore. Because CMC technologies shape the way we interact, the topic of the research is in response to the behaviours that it is possible for users to perform, and so the methods and practices used to research these must also be designed in response. Therefore, to identify an appropriate case study through which methods and practices for researching speech can be explored, the interaction context that is YouTube must be defined.

2.5 YouTube: An interaction context

Below, several theories from social computing are brought together in an attempt to understand YouTube as an interaction context, particularly in regard to the YouTuber receiving comments, how the audience is configured, how the feedback received in the comments differs from face-to-face feedback, and the potential ramifications of these factors.

2.5.1 Context Collapse

When a speaker films themselves, they are looking at a camera. Talking to a camera provides little to no information to the speaker, thus they have little idea who will watch the video, what the viewer will think and feel about it, and where or when the video will be watched. This is unlike face-to-face interactions where the speaker has an abundance of contextual information to absorb that can

influence "how [they] will act, what [they] will say, and how [they] might try to construct and present ourselves" (Wesch, 2009, p. 22). This phenomenon is known as context collapse: "an infinite number of contexts collapsing upon one another into that single moment of recording" (Wesch, 2009, p. 23).

Context collapse prompts a behaviour that is very important when considering speech style: the speaker imagines their audience (boyd, 2007). However, this is not an unusual behaviour nor one that is specific to online interactions. In fact, it can be argued that all communicative acts in all contexts, digital or not, involve an imagined audience of some kind as the speaker cannot be certain of who is witnessing the interaction. For example, 'privacy' requires that those near-by are not eavesdropping, and even when privacy can be guaranteed the speaker has a mental construction of the audience's likely responses, emotions and thoughts that may be based on impression and not fact.

From researching Twitter use, Marwick and boyd (2010) propose that there are elements of the writer's audience (Ede and Lunsford, 1984) in the imagined audiences of such context collapsed spaces. According to Ede and Lunsford (1984), the writer's audience can be conceptualised in two ways: i) the audience addressed (the actual reader), and ii) the audience invoked (the reader that the writer imagines). Just like a writer, a YouTuber has an invoked and an addressed audience. How the YouTuber defines the imagined audience is fundamental because "the imagined audience defines the social context" (boyd 2014, p32). Thus, an ambiguous audience results in an ambiguous context.

The key contribution of Marwick and boyd (2010) is outlining the concept of the networked audience; the audience that a user must navigate when using social networking technologies. They describe the networked audience as a combination of the writer's audience (both addressed and invoked audiences as described above) and elements of the broadcast audience (Livingstone, 2005). The broadcast audience has been traditionally viewed as a mass of passive, unidentifiable consumers, as demographic groups are flattened to become indistinct (Livingstone, 2005). Twitter, and other social networking sites, combine a person's individual connections to flatten their discrete audiences into

one, just as broadcast media does with audience demographics. That is, all those that can view a person's content are not just connected with that person but also with each other. The final key configuration of the networked audience (Marwick and boyd 2010) is the opportunity to give feedback. Unlike a writer's audience and a broadcast audience, "the networked audience has a clear way to communicate with the speaker" (Marwick and boyd 2010, 129).

The term 'networked publics' is used to refer to the amalgamation of the concepts of context collapse (Wesch, 2009) and the networked audience (Marwick and boyd, 2010), as occurs on many social media sites. That is, networked publics are both "(1) the space constructed through networked technologies and (2) the imagined community that emerges as a result of the intersection of people, technology, and practice" (boyd, 2014, p. 8). First defined by boyd (2007), the argument is that audiences differ from publics in the connotation that 'audience' implies passivity whereas 'publics' implies active, critical engagement (Livingstone, 2005) as can occur through the variety of feedback functions in social media sites. In the next section, the properties of the feedback in YouTube will be considered specifically.

2.5.2 Feedback configuration

As was outlined in section 3.1.4, affordances (the properties or characteristics of an environment) shape interaction contexts and thus can encourage certain types of practices (boyd, 2010). Therefore, the design of technology that mediates communication creates interaction contexts that differ from traditional, physical ones and so may encourage different communication practices. Herein, I consider the affordances of the YouTube interface in regard to shaping the interaction context within which the viewers interact with the YouTuber via the comments.

Considering the relationship between YouTuber and viewer, as previously mentioned, there is a one-to-many and many-to-one interaction. This is different from broadcast media where the interaction is generally only one-to-many from

the speaker to audience, and the majority of face-to-face interactions that typically occur (e.g. one-to-one, and multiple one-to-several exchanges, from speaker to listener and vice versa). The feedback that a YouTuber receives in the comments embodies different qualities from the feedback received in face-to-face interactions. First, it is overt, and comments are often direct and explicit (e.g. "You're so funny"). Second, feedback is provided in a written medium. In contrast, feedback in face-to-face interaction is subtle, not overt or explicit, and is communicated through feedback mechanisms such as gesture, body posture, and facial expressions. Text is generally devoid of these qualities, although emoji and emoticons can provide some paralinguistic information. Third, as has been explained in the section above, the YouTuber does not know who the feedback is from or has very little information about the commenter. Thus, there is a tension between the overtness of the written feedback and the uncertainty of who it is from.

2.6 Speech Style

As was explained in section 1.1.3, a person's speech is not always consistent. The same person may speak differently at different times (intraspeaker variation). Factors that align with this variation are the topic discussed including its emotional content, and the physical environment in which the conversation is taking place. But, most importantly, intraspeaker variation can relate to the listener (or, in other words, the audience) and the speaker's knowledge of and relationship with them. The overlaying of these different factors creates a way of speaking specifically for a certain topic, in a certain place, with a certain audience: in other words, a speech style. Thus, interrogating a speech style can answer the question "*Why* did *this speaker* say it *this way* on *this occasion*?" (Bell, 2001, p. 139). However, speech style means different things across the three waves of sociolinguistics and it can play differing roles in sociolinguistic studies. Furthermore, the analysis of style can be quantitative or qualitative.

There are three major sociolinguistic theories for explaining intraspeaker variation: i) Attention to Speech (Labov, 1966) (section 2.6.1), ii) Audience Design (Bell, 1984, 2001) (section 2.6.2) and iii) Speaker Design (Schilling, 2013a) (section 2.6.3).

2.6.1 Attention to Speech

In the first wave, style was a marginal concern in variationist sociolinguistics (Coupland, 2007, p. ix). Style was not seen as an investigative factor but as a tool to identify and access speech that is "natural", "casual" and thus representative of the participants' "vernacular", the original focus of variationist sociolinguistics. Hence, style was rationalised as the degree to which the speaker paid attention to their speech. Labov's sociolinguistic interview (see section 3.1.1. also) attempts to tap into this through five tasks, each increasing in the formality of the communication context, with i) 'casual speech' being the most informal and iv) 'word lists' being the most formal, and as the formality increases so does the speaker's attention to their speech. Labov (1966), and others, found usage of standard speech features increases as the formality of style increases. For example, Trudgill (1974) looked at the use of "ing" and "in" (e.g. 'walking' versus 'walkin') in Norwich. The use of the standard variable (ing) increased as the formality of the situation did, with 'casual speech' having the least use of "ing" and the 'word list' having the most.

This conceptualisation of style is rather restricted and simplistic for several reasons, as fully untangled by Coupland (2007). This is as a result of the theory being rooted in the first wave of variationism (Eckert 2012), thus only considering the most broad social categories, predominantly gender and class in studies of style as attention to speech. First, only one continuum of speech variation - standard to nonstandard, with prestige and stigma as its poles – is contemplated, ignoring other continua along which speech may vary. Second, this continuum is only considered in a controlled, and not ecologically valid interaction context. While the sociolinguistic interview provides a structured interaction context within which data that is comparable across participants can

be collected, how representative this is of real interaction contexts that occur in everyday life is debatable. Finally, style is operationalised as one dimensional (attention to speech) ignoring the other motivations that may prompt speakers to adjust their speech, such as those that are outlined below.

2.6.2 Audience Design

The second major theory for explaining intraspeaker variation views the addressee as the primary influence of speech change. Audience Design (Bell 1984, 2001; Bell and Johnson, 1997) posits that speakers produce their speech primarily for and in response to their audience. Thus, this theory is categorised as being of the second wave of sociolinguistic study (Eckert, 2012) because the speaker, their addressee, and the relationship between the two of them is rationalised as the most influential factor. In a seminal study, Bell (1984) demonstrated that a radio newsreader changed his speech depending on the radio channel he was being broadcast on, because of their differing audiences, even though the broadcasts were of the same text, from the same studio, on the same day. Thus, in studies testing the theory of Audience Design the speech features of the same speaker are quantified and compared across interactions with multiple, different audiences. The relative differences in the rates of using certain the speech features are compared, and the results interpreted with the speaker/audience relationship as a backdrop.

The Audience Design model builds on speech accommodation theory, first proposed by Giles, Taylor and Bourhis (1973). This theory posits that speakers converge their speech to that of their interlocutors, and sometimes diverge from it, in order to signal their relationship. However, Audience Design differs from Speech Accommodation Theory in that style shifting does not have to be along one continuum of being more or less like that of the addressee's speech but can vary along several axes. Furthermore, the Audience Design model acknowledges that members of a speaker's audience other than the addressee may also influence speech. These other audience members are referred to as i) auditors, ii)

over-hearers, and iii) referees[2]. An auditor is a known and permitted member of the interaction but is not directly addressed and an over-hearer is someone who does not participate in the interaction but is known to be within hearing distance of it (Bell, 1984). For these two audience members, a speaker adjusting their speech would be a *responsive style shift*. But framing speech style as a solely responsive behaviour cannot account for all stylistic variation. This leaves creative style shifts that appear to be initiated by the speaker unexplained. Hence the inclusion of referees in the list of audience members.

Referees are not present but are so important to the speaker that they influence their speech; an effect coined "referee design". In referee design, the speaker's focus moves from the audience to a person or persons that are not present and "the linguistic features associated with [this] reference group can be used to express identification with that group" (Bell 2001:147). This is referred to as an *initiative style shift*. Referee design can be ingroup or outgroup, a speaker emphasising their own identity and way of speaking to distance themselves from their addressee or aligning themselves with an identity and way of speaking that is not actually their own which has prestige for their addressee, respectively. An example of ingroup referee design would be an Irish person talking to an English person in an English city but emphasising aspects of their Irish accent. Here, they are initiating a style shift to reference their Irish identity, although their addressee is English and no other Irish speakers are present. An example of outgroup referee design would be a middle-aged person being interviewed by another middle-aged person for a job working with teenagers. The interviewee may initiate a style shift towards what they believe to be the way that teenagers in the local area speak in order to demonstrate to the interviewer that they would be appropriate for the job.

Although this initiative component was present from its inception, referee design was at first considered a small part of the Audience Design model. Nearly 20 years after its publication, Bell (2001) reworked the Audience Design model,

---

[2] Bell (1984) also defines "eavesdroppers" in the Audience Design model. This is a listener that the speaker is not aware of. Such listeners are excluded from this discussion because "[e]avesdroppers, being unknown, by definition cannot affect a speaker's style." (ibid, p. 160).

primarily based on Bell and Johnson (1997), a highly controlled series of experiments with gender and ethnicity as the independent variables. He conceded that audience design and referee design operate in parallel, and thus initiative style-shifts are just as important as responsive style-shifts

Finally, a notable point for this thesis is Bell's (1984, p. 191) statement that "[a] good case can be made for regarding all mass media language as referee designed." This is because the speaker does not know who their addressee is, and so a responsive style shift is impossible. In other words, "[b]ecause the mass communicator is cut off from the audience, there is no effective, equal-terms feedback" and thus "the media audience is, for the communicator, unspecific". Therefore, any style shifts in media must be initiative.

2.6.3 Speaker Design

The third major theory for explaining intraspeaker variation was defined in response to "a perceived inadequacy in existing theories of style-shifting" (Geere, Everett and MacLeod, 2015, p. 12). The "inadequacy" identified in both Labov's Attention to Speech model and Bell's Audience Design model is that they view speakers as primarily reactive to an external change, thus passive in creating and using speech style. Behaviours that could not be related to a change of audience were not left unexplained, however. As is detailed above, Bell included referee design in the original Audience Design model (1984) and redressed the balance between Audience Design and referee design in its reconceptualization (2001) so that initiative style-shifts were just as important as responsive style-shifts. However, Natalie Schilling (2013a) argues the supposed initiative shifts of referee design are actually still reactive because whether the speaker makes reference to an ingroup or an outgroup is dependent upon who the addressee is. Schilling also argues that such shifts are described as incidences where speech styles that are not normatively associated with the speaker or the speaking context are utilised. Therefore, emphasis is still placed on pre-established linguistic-social associations providing meaning.

In response to these issues, Schilling (2013a) posits the theory of 'Speaker Design': a theory that focuses on the agentive uses of sociolinguistic variation. That is, speakers are active in fashioning themselves and the context that they are within by creatively using speech features, depending on their communicative motives. The formulation of Speaker Design is exemplified by an analysis of the "self-conscious" speech of a participant, Rex, in a project about Ocracoke English (Schilling-Estes, 1998). Simultaneously described as "the performance register", Rex's exaggerated production of particular phrases, that highlight the most salient features of the dialect, was attributed to fulfilling the role of 'participant' in the sociolinguistic study that the data was taken from. Rex could have equally produced speech that is more standard, as would be expected when one's speech is under study (this would be explained by Labov's (1966) Attention to Speech model), or more nonstandard considering the audience of other islander friends and the research assistant who Rex had begun to build a friendship with (and would be explained by Bell's (1984, 2001) Audience Design model). Rex, however, "has a choice as to how he appears" and so "opts to assume the role of the quintessential quaint islander" (Schilling-Estes, 1998, p. 75). Schilling-Estes argues (1998, p. 53) "the incorporation of performance speech into the variationist-based study of style-shifting offers support for the growing belief that style-shifting may be primarily proactive rather than reactive."

Unlike style shifting studies from an Attention to Speech (Labov, 1966) or Audience Design (Bell, 2001) perspective, those that utilise Speaker Design to explain their results vary to a greater degree in terms of study design and method. The Speaker Design approach asserts that "the social meaning of linguistic variation is located in the qualitative patterning of stylistic variation in interaction rather than the quantitative patterning of linguistic-social group variation" (Schilling, 2013a, p. 339). This focus on the individual utilising resources, and in analysing speech at the level of interaction evidences its association with the third wave of variationism (Eckert, 2012). Examples of Speaker Design studies that have taken a primarily qualitative approach include Coupland's (2001) observations of a radio DJ and guest invoking Welsh speech

features for persona management, and Podesva's (2007) examination of a participant's use of falsetto to modify the identity portrayed.

Cutillas-Espinosa, Hernández-Campoy, and Schilling-Estes (2010), however, examined multiple speakers and used quantitative data to evidence one speaker constructing a speech style that was unexpected. They identified "hypervernacularisation" in the speech of a female Murcian politician who has a working-class background. Hypervernacularisation is the correct use of nonstandard speech features in linguistic terms but incorrect or inappropriate use according to socio-demographic and/or stylistic parameters. A simple example would be using a strong/broad regional accent when giving a formal speech. In comparison to other speakers (both male and female, Murcian and non-Murcian, politician and non-politician, and from lower, middle and upper classes) the former President of the autonomous region of Murcia showed relatively low usage of many standard features when speaking in work related, less-formal contexts. This violates expectations based on occupation, social class and gender, thus Cutillas-Espinosa and colleagues rationalise this unexpected behaviour by attributing it to wanting to project a "socialist identity" (2010, p. 47), and "downward social mobility and a working-class image" in pursuit of her political goals (2010, p. 49).

2.6.4 Speech style on YouTube

To date there have been two sociolinguistic investigations of speech style on YouTube: Lee (2017) and McDonald (2018). These studies attempted to consider style under the three approaches outlined earlier: i) Attention to Speech (Labov, 1966), ii) Audience Design (Bell, 1984, 2001), and iii) Speaker Design (Schilling, 2013a). Both took different types of videos from along a continuum of the degree of planning and prior preparation. Lee (2017) described the scale as "scripted" to "unscripted" whereas McDonald (2018) used the terms "planned" and "spontaneous". Lee (2017) considered style-shifting across 4 types of video from the same YouTuber whereas McDonald (2018) contrasted 2 types of video – a planned makeup tutorial, and a spontaneous vlog - from 4

YouTubers from different regions of the UK. Both found the speaker's style shifted in relation to the type of video, the correlation being that the use of the prestigious speech feature (from a national, class perspective) increased as the degree of prior preparation for the video increased. This result can be explained by Labov's (1966) theory of Attention to Speech. It is questionable, however, whether the investigation design used by Lee (2017) and McDonald (2018) would have revealed speech style phenomena that can be explained by Audience Design (Bell, 2001) or Speaker Design (Schilling, 2013a) because neither study considered who the YouTuber's imagined their audience to be.

Aside from these two quantitative studies, all previous literature of style on YouTube has comprised reports of qualitative, multi-modal analysis. The most relevant to this study is Maximiliane Frobenius' work (2014). Here, Bell's (1984) Audience Design (along with the theories of Goffman's (1981) participation framework and Clark and Carlson's (1982) Audience Design) is considered in the analysis of 30 vlogs in conjunction with their comments. She reports on a variety of involvement strategies that speakers use to adapt to this interaction context where there is no immediate feedback and those being addressed are imaginary. However, speech is a minor consideration amongst many others (linguistic content, conversational history, physical arrangement, and gaze/gesture). Thus, Clark and Carlson's (1982) Audience Design dominated throughout, evidenced by adapting this model to structure the reporting of the results. Frobenius's finding that "there is a form of audience involvement present in vlogs resembling that of face-to-face conversation" (2014, p. 70), despite the interaction being asynchronous and unidirectional, provides further evidence to support the argument for the investigation reported in this thesis.

Most importantly in relation to this case study, however, is that Lee (2017), McDonald (2018), and Frobenius (2014) show no engagement with the literature and research findings from social computing and other related fields as I do here. In particular, while these studies do consider the idea of an imagined audience none integrate Wesch's (2009) context collapse, Marwick and boyd's (2010) networked audience, or boyd's (2014) networked publics and technological

affordances. Integrating these theories would have helped conceptualise the YouTube space, and the qualities of the interactions that occur within the YouTube space. Thus, a strength of the work in this thesis is its interdisciplinary nature.

2.7 Selecting a Speech Style Theory

So far in this chapter, section 2.5 has defined YouTube as an interaction context through applying the theory of context collapse (boyd, 2007; Wesch, 2009) and considering the feedback mechanisms within the platform. Section 2.6 has described the three prominent theories of style in sociolinguistics (Attention to Speech (Labov, 1966), Audience Design (Bell, 2001), and Speaker-Design (Schilling, 2013a) and reviewed the initial research that has considered these theories of style within a YouTube context.

One of the most important points to remember about the three sociolinguistic theories of style is that each does not totally discount the other. Besides each being associated with a different wave of sociolinguistic enquiry and thus different investigative practices, each theory has emerged from highlighting and then addressing the limitations of its predecessor. For each theory proposed there have been, and will be, speech behaviours that are exceptions. Thus, a new theory was brought forward to account for those exceptions. While the respective authors may argue their theory of style can most accurately account for speech behaviour, a more neutral view is that each of these theories covers style in a different way, and thus one theory cannot cover style in all ways. Communication is a complex interaction between external and internal factors at different levels, e.g. at the individual, community, or social demographic, and one may argue that each theory is able to explain speech style behaviour when the force of these internal and external factors are configured in different ways.

In choosing a theory of style to use for this case study, each theory's limitations needed to be considered. I argue that the most notable limitation is that prioritising or emphasising the influence of internal factors leaves speaker

behaviour unpredictable in many interaction contexts. External factors can be identified, described, and a hypothesis about resulting speaker behaviour can be made prior to its observation. The Attention to Speech and Audience Design theories both focus on external factors; i) the relative, perceived formality of a series of interactions contexts, and ii) the relationship between the speaker and audience whichever kind they may be, respectively. However, speaker motivation is predominantly an internal factor and so is difficult to identify and describe prior to analysing interaction. Thus, it is very difficult to define and provide rationale for a research hypothesis under the theory of Speaker Design. Further, one could even argue that for evidence that supports Speaker Design to emerge, speech behaviour in an interaction context has to be predicted based on the Attention to Speech and Audience Design theories first. In other words, there is an expectation that we should understand an interaction context as is conceived by Attention to Speech and Audience Design before we can talk about Speaker Design. As described in section 2.6.4, both sociolinguistic studies of YouTube to date (Lee, 2017; McDonald, 2018) evidence that, at least initially, behaviours that can be explained by Labov's (1966) theory of style are present in this online interaction context. This provides a fundamental foundation on which to build this case study.

The suitability of these theories within the complex interaction context that is YouTube also needs to be considered. For this case study, I decided to focus on Bell's (2001) Audience Design. First, what makes the context of YouTube so interesting and exciting is also what differentiates it so strongly from interaction contexts that have been investigated so far. What YouTube offers investigations in speech and style in relation to audience is an abundance of potentially detailed, explicit feedback in the form of comments. To date, the factor of "audience" has been operationalised as a categorical data in sociolinguistic studies, for example Bell and Johnson's (1997) experimental study design of changing the conversation partner. Therefore, the differences that comment feedback embodies compared to face-to-face feedback (as described in section 2.5.2) allows for the feedback to be quantifiable and the factor of "audience" to be operationalised at a much finer grained level.

Second, due to context collapse (Wesch, 2009), the YouTuber has to imagine their audience. But because viewers can communicate with the YouTuber via the commenting function, there is the opportunity for the YouTuber to gain feedback from their actual viewers and modify their imagined audience in light of this information. Therefore, unlike previous studies, the effect of far smaller, subtler changes in audience upon speech style can be examined. Of course, the commenters may not be representative of the whole viewership, nor do the comments inform knowledge of the viewership in regard to demographics, but in their opinions. Therefore, the imaginary viewership is actually a refraction, not a reflection, of the commenters.

Finally, previous work of written computer-mediated communication (reviewed in 2.3.1.1) indicates such a study will be fruitful. Prior work suggests that people change linguistic behaviour in relation to audience whether that is on the same media (Pavalanathan and Eisenstein, 2015; Shoemark *et al.*, 2017; Gil-Lopez *et al.*, 2018) or different media (Chariatte, 2015; Childs, 2016). This provides a strong argument for expanding this topic of research into speech behaviour.

2.8 Case Study Research Question and Hypotheses

This study explores speech style in relation to viewership from a sociolinguistic perspective within YouTube. In doing so, this study aims to contribute to our understanding of the impact of YouTuber-commenter interaction. Overall, this case study asks:

- Does the direct written feedback received through the commenting function influence a YouTuber's speech?

As has been outlined above, the sociolinguistic theory of style that will be explored within this study is Audience Design (Bell, 2001). Audience Design assumes that intraspeaker variation is a largely automatic response to known and/or assumed knowledge of one's audience. Under this premise, it would be predicted that a YouTuber would adjust their speech in relation to their imagined viewership as they gain more information about them through the comments.

The null hypothesis is that a YouTuber is not influenced by the comments, thus there should be no evidence of a causal relationship between the content of the comments and the use of the speech feature. Further, the patterning of the data may be indistinct, indicating that a YouTuber adjusts their speech in response to a combination of factors.

Finally, it should also be acknowledged that a YouTuber's behaviour may differ over time. First, because the content of the comments may change, or second, because their conceptualisation of their relationship with their imagined audience. Thus, these scenarios will also be considered in the study.

2.8.1 A note on terminology

Both Bell (Bell, 1984, 2001) and Marwick and boyd (2010) use the term 'audience' in the names of their theories. This could become problematic for several reasons. First, referring to the YouTuber's 'audience' could unintentionally communicate bias towards Audience Design (1984, 2001) being present in the context of YouTube. Second, while Marwick and boyd (2010) do not explicitly discuss the active or passive nature of the networked audience, boyd (2014) goes on to define networked publics, explicitly stating that the term publics indicates activity whereas audience indicates passivity. Further, using the term 'audience' may cause confusion as to whether Bell's (1984, 2001) Audience Design, Marwick and boyd's (2010) networked audience, or some other conceptualisation of 'audience' is being discussed. Therefore, the neutral term 'viewership' will be used to refer to the collective of people who watch a YouTuber's video, with 'viewer' being used as the singular. The passivity of this

term will be utilised by assisting in differentiating viewers (those who just watch a YouTube video) from those who watch but also comment ('commenters'). In regard to 'viewership' two more terms should be defined: 'actual viewership' and 'imagined viewership'. Actual viewership shall be used to describe the group of real but unknown people who watch a YouTuber's videos. Imagined viewership shall be used to refer to the group of people that the YouTuber envisions is watching their videos. It is important to state and contrast these two terms because the actual viewership and the imagined viewership may overlap to different degrees or not be similar at all.

2.9 Chapter Summary

To conclude, in the second half of this chapter the exploratory case study through which the thesis questions will be answered has been outlined. YouTube as an interaction context was described and an appropriate theory of speech in relation to audience that could be explored within the context of YouTube was selected. The case study's research question and hypotheses were defined. As was stated in the first half of the chapter, the design of research methods and practices needs to be responsive to the particular CMC of interest and it is highly likely that this will require modifications of more well-established research methods and practices currently applied in offline research. The next chapter will focus on what aspects of current typical sociolinguistic research practices are less likely to transfer well from offline speech or online written variationist studies to speech in online video and thus motivate the rest of the thesis investigations.

Chapter 3.

Methods Review

"The question is much more interesting, potentially, than whether old methods can be adapted to fit new technologies. New technologies might, rather, provide an opportunity for interrogating and understanding our methodological commitments."

(Hines 2005:9)

The literature review of chapter 2 has provided a summary of online investigations and evidenced online speech as an underexplored area. Our focus can therefore now shift onto *how* variationist investigations of online speech could be conducted. Despite the growing interest in sociolinguistic investigations of CMC, there is little literature in regard to methodology and methods. What few exceptions there are (Herring, 2007; Androutsopoulos and Beißwenger, 2008; Androutsopoulos, 2013; Bolander and Locher, 2014; Lim and Sudweeks, 2014) come from a discourse analysis standpoint and thus their relevance to the variationist analysis of speech is limited. Although Bolander and Locher's (2014) is the broadest discussion and does touch upon some issues that are relevant to this thesis (e.g. ethics) that will be discussed later on.

An overview of the methods constructed within sociolinguistic investigations is given below in order to identify the aspects that do not easily transfer from

studies of offline speech to studies of speech online. Equally, where relevant, methods used when researching offline language and language-online are also considered as well as notable CMC related literature. Four elements that are key to constructing a method are identified and their theoretical and practical complexities considered in depth: i) Formulating Research Questions, ii) Research Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis (sections 3.1, 3.2, 3.3, and 3.4 respectively). Within and across the discussion of these elements, decision making and research practices are presented as sequential, although in reality they may be cyclical or spiral (as noted in Feagin, 2013). From these discussions, the **thesis questions** are defined (sections 3.5). As outlined in chapter 2, while much of the work in online speech has been qualitative, and much of the quantitative work in online contexts and CMC relate to language, the literature in both these areas was considered to provide context and to gain useful insights and further inform the **thesis questions**. Prior to the chapter summary (section 3.7), the topic of auditory and acoustic analysis is also addressed (section 3.6).

3.1 Formulating Research Questions

The foundation of a research project is the research question (Blaikie, 2007). Yet accounts of the process by which researchers formulate such questions are rarely published. What research question can be answered is determined by the data's qualities. Some fundamental data qualities in linguistics are the language, dialect or accent being spoken. For sociolinguistics, there are also social factors such as the age, gender and social class of the speakers; technical factors such as the recording equipment used, file types, and sampling rate; situational factors such as the type of speech activity that was recorded, who the speakers are and how they know each other; as well as more practical elements such as how many speakers there are and how much data there is per speaker. In other words, what the data is like determines what a researcher can ask of it, or conversely "[t]he choice of research question determines the kind and amount of data you need" (Hazen, 2014, p. 9).

The data's qualities are determined by the data collection methods used. To quote Milroy and Gordon (2003, p. 49), and Macaulay (2009, p. 30), respectively, "[w]hat constitutes "good data" depends on the research objectives, as do the methods for collecting such data" and "what investigators choose to sample and how depends upon the question they want to answer". Thus, the process of designing a research project is a negotiation between these three elements: i) research question, ii) data collection methods, and iii) data qualities, until a satisfactory compromise is reached.

One convenience of sociolinguistics is that any recorded speech potentially holds interesting material, and so many different types of data have been used in research. Adopting Cieri and Yaeger-Dror's (2018) terminology, data types could be categorised as:

i)   "Tailored" - "collections that the linguistic researcher has designed and executed for a specific research agenda" (p.54),

ii)   "Found" – "alternative data, designed and developed for some other purpose but discovered by the linguist as relevant to her research agenda in some way" (p.54) which may include intra- or cross-disciplinary use, or

iii)   "Raw" – "created for an entirely different purpose" (p.54) and it is "material that has not been collected for a specific language-related research need" (p. 60),

with each potentially possessing very different data qualities as a result.

These three data types reflect a continuum of control. Here, I use the term 'control' to refer to the ability to ensure that the data's qualities align with the research question, so that a researcher has 'good data' for their study. The most control would be held by the researcher in question (tailored data). Next would be found data, such as a corpus, where most of the researcher's control is in selecting a portion of the data available (e.g. all young speakers, only the read passages). But also, while the secondary researcher (the one accessing the corpus) relinquishes much control, the primary researcher (the one who collected

the original data) tailored the data collection to their research question and so had a greater degree of control. Thus, a secondary researcher can have a greater degree of confidence that the data will possess certain qualities that make it suitable for research. Finally, a researcher would have least control over the data qualities of raw data and least confidence initially that the data will be suitable for research because it was not collected for this purpose.

This section will review tailored, found and raw data in sociolinguistics as well as online written data and YouTube data. By understanding how these data types are collected and the data qualities they possess, we can also know the types of research questions that can be asked, and gain insight into the decision-making process researchers may take to formulate these research questions.

3.1.1 Sociolinguistics: Tailored Data

Formulating a research question when a researcher has the opportunity to collect tailored data is predominantly guided by prior literature and theory. Some practical implications may also have an influence, such as already being a member of, or having access to, a relevant speech community. Plus, there is always the risk that a researcher has to adjust their original research question in response to unforeseen events during data collection (e.g. being unable to record enough speakers). The predominant method for collecting tailored data in sociolinguistics is the sociolinguistic interview, if Becker's (2013, p. 91) definition that "any face-to-face interaction that is recorded for use as sociolinguistic data" is taken. Experiments have also been used, and ethnography is also often employed to collect complementary data to help understand the linguistic data. All of these data collection methods will be reviewed below, and more recent innovations in sociolinguistic data collection will also be considered.

3.1.1.1 The Sociolinguistic Interview

The founding data collection method within variationist sociolinguistics was audio recording a sociolinguistic interview. Developed by Labov for his 1966 study of New York City (2006), it included:

i) Personal narratives (to encourage casual speech, e.g. the Near death experience question "Have you ever been in a situation where you thought there was a serious danger of your being killed? That you thought to yourself, "This is it?" (Labov, 2006, p. 415)),

ii) Responding to interviewer questions (to encourage careful speech, e.g. "What country were you born in?", "Are you married?", "Have you any children?" (Labov, 2006, p. 409)),

iii) Reading passage (to collect read speech),

iv) Word lists (to gain an overview of the speaker's pronunciations),

v) Minimal pair list (a specially designed list of paired words where only the speech feature/s of interest differs in each pairing).

In the words of Labov, "each part of the interview had at least two purposes: first, to provide the context for a given style of speech," (as noted above) "and second, to obtain the specific information proper to the questions themselves" (2006, p. 91). Thus, some of the questions in tasks 1 and 2 covered topics such as Linguistic Attitudes (Section VII) by asking "What do you think of your own speech?" and "What do you think of [insert area of study, e.g. New York City] speech?" (Labov, 2006, p. 420), the answers to which provide invaluable social insight and context that can be integrated into the data analysis.

However, the strict method as described in (Labov 2006) is rarely followed in contemporary studies. It is complicated and long with many very different tasks. In addition to asking open/narrative questions (Personal narratives) and closed/demographic questions (Responding to interviewer questions) and the reading of passages and word lists, the interviewer is also required to play recordings of other people's speech for the interviewee to respond to and rate on a scale (Section VI: Subjective Reaction Test – Labov, 2006, p. 419) and design

descriptions of objects or activities to elicit the corresponding local words
(Section II: Lexicon - Labov, 2006, p. 411). To follow Labov's prescribed
method in full, the interview would include 9 sections with most of these
containing 6 or 7 questions and/or tasks of different kinds. Although this allows
for thorough data collection of a wide range of relevant information, modern
sociolinguistic studies are usually more targeted with the interview tasks being
designed to elicit many tokens of the target linguistic variables. Examples of
other data elicitation methods include the map task, as used in e.g. (Grabe,
2004). Each participant has a map and one participant is asked to describe to the
other a route from point A to point B. The linguistic content of the map (e.g. the
names of streets, towns, rivers or landmarks, types of shops and amenities) is
designed to ensure the linguistic variable of interest is produced, and the two
participants' maps are different so many repetitions of the same words are made
as the participants double check what each one has said and describe the
differences in their maps to each other. The set up for the spot-the-difference
picture task (as described in Van Engen *et al.*, 2010; Baker and Hazan, 2011) is
similar in that the content of the image is designed to encourage the participants
to say certain words. Therefore, the term 'sociolinguistic interview' now seems
to "stand for any face-to-face interaction that is recorded for use as
sociolinguistic data" (Becker, 2013, p. 91).

3.1.1.2 Experiments

It can be argued that experiments have been used in variationist sociolinguistics
since Labov pioneered the sociolinguistic interview, if one were to describe
reading passages and word lists as experiments (as Clopper (2013) does) with
the language stimuli being the independent variable and the speech produced
being the dependent variable. Just like in the sociolinguistic interview, the main
strength of experiments is the degree of control that the researcher has over the
content of the speech produced. Notably, this control can be realised over the
linguistic content of the speech. Thus, the researcher can be confident that a
sufficient number of tokens of the linguistic variable of interest will occur. This
is particularly advantageous when the linguistic variable is rare. Further, other
linguistic factors can also be controlled for. For example, when studying a

speech sound, it is known that the speech sounds immediately before and after it will influence its production (Johnson, 2012). Controlling for linguistic factors increases confidence that any results are due to the social factors under study and that they are not a by-product of some linguistic factor.

Of course, this section focuses on experiments targeting speech production as this most closely resembles the way in which YouTube data is examined in this thesis. For a brief review of these see (Drager, 2014) and for a more thorough understanding and practical guidance see (Drager, 2018).

3.1.1.3 Ethnography

Ethnography within sociolinguistics can be described as prolonged participant-observation in order to gain insight into how the members of a community behave (in our case, how they speak) and why they behave in that way (Levon, 2013a). Unlike first wave studies, where broad social categories are imposed upon participants by the researcher, studies that include ethnographic practice can be classified as second wave studies and aim to understand variation in relation to local meaning and practice (Eckert, 2012). That is, the speech community being studied defines its own social groups and their associated activities (Schilling, 2013b).

The first variationist sociolinguistic study (Martha's Vineyard, (Labov, 1963)) included an ethnographic perspective somewhat in that the patterning of the data could only be understood by considering participants' opinions of traditional island life. According to Natalie Schilling (2013b), this aspect of Labov's ground-breaking study was overlooked at the time, with most subsequent work from other researchers focusing on objective social categories. Ethnography re-emerged, however, in the 1980s with Penelope Eckert's (1989) study of teenagers at a Detroit high school being the most notable.

Ethnography as a standalone method would not fit into the variationist paradigm as it does not focus on collecting quantitative data. But since the 1990's there has been an ever increasing use of ethnography to complement sociolinguistic

data collection methods, in the same vein as Labov's (1963) research in Martha's Vineyard. A more detailed literature review of use of ethnography in sociolinguistics will be given in section 6.3.1.

3.1.1.4 Innovation in Data Collection

There have recently been some innovations in data collection practices in sociolinguistics, such as specially designed data collection apps for smart phones (e.g. Leemann, Kolly and Britain, 2018) and self-recorded data (e.g. Podesva, 2007, 2011; Sharma, 2011; Boyd *et al.*, 2015; Hall-Lew and Boyd, 2017; Leemann, Kolly and Britain, 2018).

With an app, such as "English Dialects" (Leemann, Kolly and Britain, 2018), a potential participant downloads it from the relevant app store and then makes recordings of read speech. Users can also listen to others' recordings, which are displayed on a map to reflect where the participant is from. A present limitation of this design is that only read speech, and not conversational speech, is collected. Also, most of the current examples are not webapps and so are not editable without updates, meaning there is little need for the user to interact with the map once they have recorded themselves reading all the passages it contains. However, we are in the first generation of apps of this kind and there is great potential for this method to continue developing.

As the focus is on user-generated YouTube video, somewhat more relevant to this thesis is the use of self-recorded data. Hall-Lew and Boyd (2017:89) define 'self-recordings' as "recordings made without the researchers acting in researcher roles." Podesva's (2007:483) motivation for requesting self-recorded data from participants was in the hope that researcher absence "would yield more naturalistic, less self-conscious recordings". Levon (2013b) suggested that the recording situation (the physical location, communication partners, topic of discussion etc) is the strongest predictor for speech behaviours and thus self-recordings do not lead to significant differences in the content of data or study results. Recently this claim has been tested, and self-recorded data has been explicitly compared to more traditional methods. Notably, Boyd et al (2015)

found more advanced productions of five sounds that were currently undergoing change when examined in self-recorded data than in data obtained from classic sociolinguistic interview methods. This indicates that self-recordings result in a markedly different speech style, and data that is closer to the vernacular (see section 2.1.1 for a definition).

3.1.2 Sociolinguistics: Found Data

To recap, found data is "alternative data, designed and developed for some other purpose but discovered by the linguist as relevant to her research agenda in some way" (Cieri and Yaeger-Dror, 2018, p. 54). There is a growing trend for sharing research data. In the UK, this is now often a requirement from national funding bodies and they facilitate this activity through the UK Data Service (see UK Data Service, no date, for further information). Further, researchers publicise the availability of their corpus in a variety of ways including webpages, social media, academic email lists, at conferences and through publications (e.g. Gold, Ross and Earnshaw, 2018). Thus, secondary researchers can become aware of corpora relatively easily if some time is spent searching these sources. This is promising for secondary researchers who have a research topic in mind and wish to find an appropriate corpus of data to use. Alternatively, secondary researchers may happen upon a corpus and be struck with inspiration for a research project. In other words, secondary researchers may approach corpus data with a research question, or one may develop as they become familiar with a corpus. Hazen's (2014, p. 9) observations of how researchers interrogate a corpus also reflects this distinction: "[l]arge corpora can be searched in either an exploratory way to develop research questions or in a research-directed manner after crafting a research question". Of course, in both cases the corpora's data qualities impose restrictions on what the secondary researcher can ask, and thus "[t]he immense and growing supply of found data shifts the researcher's challenge from creating recordings to selecting speakers and sessions of interest" (Cieri and Yaeger-Dror, 2018, p. 67).

Found data may have been collected for research purposes that are sociolinguistic (intra-disciplinary corpora) or not (cross-disciplinary corpora). Interestingly, the few cross-disciplinary corpora highlighted by Cieri and Yaeger-Dror (2018) have been compiled for developing speech recognition technology, either by collecting tailored data (e.g. the Greybeard corpus (Brandschain *et al.*, 2010)) or taking advantage of raw data, mostly radio and TV broadcasts (e.g. Boston University Radio Speech Corpus (Ostendorf, Price and Shattuck-Hufnagel, 1996)). Use of cross-disciplinary corpora is rare in sociolinguistics and presents similar if not the same issues to consider as intra-disciplinary corpora, and so the discussion in this thesis is restricted to the latter.

3.1.2.1 Intra-disciplinary Research Corpora

With some sociolinguistic studies of speech, the intention from the outset is to create a corpus from the data collected so that secondary researchers can interrogate the data with their own questions. While the research questions that the secondary researcher (the one accessing the corpus) will ask of the corpora will be different from those asked by the primary researcher (the one who collected the original data), secondary researchers can have confidence when using corpora because the data is likely to embody certain qualities. It is highly likely that the recordings will be good quality, that the data has been collected in an ethically sound way (e.g. participants have gone through a formal consenting procedure), that all or some of the activities recorded are very similar to those described above in section 3.1.1.1, and the participants recruited may even be balanced in regard to some social factors (e.g. an equal number of female/male, older/younger participants). For example, the IViE corpus (Grabe, Post and Nolan, 2001) was collected to investigate how intonation (see section 5.2) varies across the UK (9 locations), by sex, and across 5 speaking styles (e.g. read sentences, conversation with a peer). This provides 36 hours of data where other speech features are, of course, used frequently and thus can be examined using many more research questions other than that of the primary researcher.

Other speech corpora have been formed by combining a number of already existing resources, and corpora can also be supplemented by collecting

additional material. The ONZE corpus (see Gordon *et al.*, 2004) is an example of this. It brings together three collections: 1) The Mobile Unit – personal and group oral histories of 300 speakers recorded between 1946 and 48 (speakers born 1851-1910); 2) The Intermediate Archive - data from 4 sources, one being recordings of some of the descendants of the speakers in the Mobile Unit archive, and the other three could be likened to personal oral histories, some made for the purpose of research, some for radio shows; 3) The Canterbury corpus - modern day (from 1994) recordings made by linguistics students at the University of Canterbury to be a part of an archive of New Zealand English. The speakers were born between 1930 and 1984 and recruited to create a socially balanced sample (equal numbers of male/female, young/old, higher/lower social classes). Further, a notable innovation in regard to speech corpora is SPeech Across Dialects of English (SPADE) (Stuart-Smith, Sonderegger and Mielke, no date). This project's aim is to develop open-access software that links multiple speech corpora and the automatic searching and analysis of data. But more importantly, it will be possible to search across multiple datasets with the speech feature as the search criteria and retrieve data measurements without needing to access the raw audio recordings, circumventing many ethical issues.

3.1.3 Sociolinguistics: Raw Data

Cieri and Yaeger-Dror (2018) define raw data as "created for an entirely different purpose" (p. 54) and the line between cross-disciplinary found data and raw data is "[t]he term raw data is reserved here for material that has not been collected for a specific language-related research need" (p. 60). While still relatively rare, such data is increasingly being used in sociolinguistic research. Broadcast media dominates, with Queen (2013) estimating about 80% of the data used is unscripted media such as talk shows. However, the most notable sociolinguistics work does include scripted data such as documentary voice overs (Cham, 2016) and speeches (Harrington, Palethorpe and Watson, 2000; Kirkham and Moore, 2016; Hall-Lew, Friskney and Scobbie, 2017), semi-scripted sermons (Stanley and Renwick, 2016; Rodríguez, 2019) and reality TV

(Levon and Holmes-Elliott, 2013) as well as unscripted reality TV (Sonderegger, Bane and Graff, 2017).

The actual sources of these data vary and how they were discovered and then accessed goes unreported. Hall-Lew, Friskney and Scobbie (2017) subscribed to an archive of parliamentary recordings, and Harrington et al (2000) accessed an archive held at the BBC. It is assumed that a similar repository was archived by Sonderegger, Bane and Graff (2017) and Levon and Holmes-Elliott (2013) who "gratefully acknowledge permission from Channel4/Endemol to access footage" (p. 598) and describe taking scenes from "high definition downloaded files" (p. 114) respectively. In comparison, Hall-Lew, Coppock and Starr (2010) and Kirkham and Moore (2016) illustrate the use of YouTube as a repository of video recorded as a 3rd party witness, although Kirkham and Moore (2016) do not state their data source explicitly but confirmed via personal communication.

In regard to formulating research questions, data qualities will impose constraints on what research question can be asked just as they would when considering found data. However, because raw data was not collected for research purposes, it will take a researcher more time to realise what qualities the data possesses and what the constraints they impose are, in comparison to the more thoroughly documented metadata of found data. Some aspects of different types of raw data may make assessing their data qualities less challenging. For example, Harrington et al (2000) knew from the outset that the archive they were accessing would be one recording per year of approximately the same length of a scripted speech from the same speaker. Equally, both Sonderegger et al (2017) and Levon and Holmes-Elliott (2013) would have known the main speakers in their relevant TV programme and how many were men and how many were women. However, documenting how much each speaker spoke in order to design a study where an equivalent amount of data from each speaker is included would have took some time to figure out. This example illustrates that Cieri and Yaeger-Dror's (2018, p. 67) observation that the researcher's challenge shifts "from creating recordings to selecting speakers and sessions of interest" is equally as relevant to raw data as to found data.

Finally, it should be noted that some metadata that could be essential may be missing from raw data. This may include demographic information about the speakers, or their attitudes and perspectives on relevant topics. Most of the raw data sources described above are centred on public figures, and so it is likely that there are many other sources of information for a researcher to access to enrich their study. For example, in regard to using the archive of the oral arguments given in the Supreme Court of the United States, Cieri and Yaeger-Dror (2018, p. 60) argue that

> "although no fieldworker was on hand to collect demographic, situational and attitudinal metadata during the oral arguments, journalism, biographies, memoirs and scholarly treatises tell us more about the justices, their backgrounds, ideologies and interactions than we have for the average speaker in tailored sociolinguistic corpora."

3.1.4 Online Written Data Collection Methods

Just like with interviews and experiments, sociolinguistic researchers have taken existing data collection methods for online data and harnessed them for their own interests and purposes. Online data collection methods are often unique to digital media, and their practical details are bespoke to the platform or site where the data is held. Here, Twitter data will be used as an illustrative example as it is the dominant source of online written data for sociolinguistics (as outlined in section 2.3).

According to boyd (2010), bits (the building blocks of digital structures, like atoms are the building blocks of physical structures) embody four interrelated affordances. These are:

1. Persistence (the ability to record and archive),
2. Replicability (the ability to make exact duplicates of these records),
3. Scalability (the potential visibility of content)
4. Searchability (the ability to access records through searching)

Three of these are fundamental to being able to find online data of interest: Persistence allows for a repository of data to exist; Replicability allows for an exact copy of the data to be made and downloaded by the researcher; and Searchability provides the mechanism through which one can find data. However, while all digital data is made up of bits, it would be more appropriate to describe online data as having the *potential* for these affordances to be realised. The main reason for this is accessing such data requires permission. This is likely to be at two levels: i) the company that owns the platform on which the data was posted (and thus the servers where the data is saved), and ii) the individual users who have produced that data. The issues and rhetoric around data access and ethics is covered in the next section, with the discussion in the rest of this section being based on the scenario that both the users and the company allow for their online data to be accessed.

The collection of much online data is performed by accessing a platform or site's servers via an application programming interface (API). An API could be thought of as a librarian (a guardian of data), and a server as a library (a large, systematic storage of data). Servers are rarely open to anyone to access, and so permission must be granted by applying for an account with which to talk to the API, just like one must open a library account. The process of applying for an account and the requirements that must be met for the account to be awarded varies significantly across sites. Further, unlike most libraries, even with a valid account one is not allowed to go browsing the data held on a server. Data access is performed through an API and it is typical for large sites to have several APIs through which different databases can be accessed. Thus, once one has an account, a script must be written in software and then executed to tell the relevant API not only what data is requested but also where it can be collected from in the server. Server structures vary by site and so each requires a unique script.

One important element to emphasise in regard to APIs is that the data they contain and one's access to it is not unlimited. Just like a library card, each site has limits on how much data can be retrieved, how often or regularly, and what kind of data. Thus far, with most social media sites these limits regularly change

as the platform or site grows in popularity, or in response to events and user demands. With some sites, access to data is restricted depending on what kind of access has been purchased. For example, with the Twitter Search API whether one can retrieve tweets posted in the last 7 days, 30 days, or before 30 days ago depends on the subscription paid for (Twitter, 2020b). Further, not all the data the site holds will be accessible. This is for legal reasons (for example, what data could be accessed significantly changed as a result of GDPR) but also because data and API access are at the site's discretion. Finally, what data can be searched for depends on what is in the API and how the server is structured. For example, as will be further discussed in section 3.1.5, one cannot search within the captions of YouTube videos via an API. To search a YouTube video's captions one must first download them and, because it is possible to have multiple captions per video, one must know the caption's unique identification code in order to do that. Thus, this caption id code must first be collected by downloading the metadata that is related to the video of interest.

Finally, it is important to note that most sites design their access to data from a business perspective, the main tasks being companies identifying potential customers via demographics for targeted ad campaigns, analysing who engages with their social media to continually refine these demographics, and for this and other content to be integrated with their website, for example embedded YouTube videos. Few sites explicitly accommodate academic research in the design of their data access. However, Twitter is a notable exception as it provides specific services and support around collecting data for research purposes (see Twitter, no date, for further details).

Using Cieri and Yaeger-Dror's (2018) terminology, online written data such as Twitter could be categorised as raw data because its creation (user's tweeting) was not for research purposes. However, the searching of the Twitter API and downloading the data can be likened to accessing a very large corpus, which would be classified as found data. In contrast to the finite data of a corpus and most raw data sources, Twitter continually grows meaning some of the data's qualities (e.g. number of tweets from men/women, from different countries, about different topics, including certain GIFs, emoji or hashtags) will continually

be in flux. What research question can be answered is determined by the data's qualities, and so formulating research questions to ask of potentially unlimited data that is continually growing is complex. But because Twitter data is searchable the search criteria can impose constraints on the data qualities and therefore the potential research questions, just like when a secondary researcher searches a corpus. The way the data's API or server is structured dictates what can be searched for and how, and so whether the data is structured in and searchable by sociolinguistically meaningful criteria fundamentally shapes the research question. Such criteria could include location in the form of geotags, the topic represented by hashtags, or a cross referencing of both (e.g. Shoemark *et al.*, 2017). One can search for particular linguistic speech features of interest, for example 'this/dis', 'that/dat', and 'they/dey' in Callier (2016), or even identify them by collecting a mass of data within a specific timeframe from across a specific geography and then filter the dataset to reveal interesting linguistic phenomena (Grieve, Nini and Guo, 2018). To summarise, online written data such as Twitter data may be infinite, but its organisation allows researchers to search using sociolinguistically meaningful data qualities as criteria and this naturally imposes constraints on the research questions that can be answered.

Finally, it should be noted that there are some online data collection methods that do not necessarily require the downloading of data. An example would be online ethnography[3] where other types of data and thus data collection techniques (e.g. notes in a field work diary, screen shots) are used. A more detailed description of online ethnography and how it differs from "offline" ethnography will be given in section 6.2. Thus, online data may either be analysed "live" as it is accessed, viewing and interrogating the data simultaneously and recording observations in some way (e.g. note taking while streaming a video), or a copy of the data could be retained (e.g. downloading a video) for analysis later on or within a software.

---

[3] Other terms used in the literature include "virtual ethnography", "cyberethnography", "discourse-centred online ethnography", "Internet ethnography", "ethnography on the internet", "ethnography of virtual spaces", "ethnographic research on the internet", "internet-related ethnography", and "netnography" (Varis, 2016, p. 55) p. 55 Piia Varis)

3.1.5 YouTube Data Collection

YouTube could be viewed as raw data because it was "created for an entirely different purpose" (Cieri and Yaeger-Dror, 2018, p. 54) from that of research. Many of the same issues and potential limitations in regard to formulating research questions that arise when other kinds of raw data are used in sociolinguistics (see section 3.1.3) and when researching online written data (section 3.1.4) are predicted to also be relevant to researching YouTube data. However, YouTube also presents some unique challenges that are aptly reflected in this quote from Burgess and Green (2009, p. 88): "YouTube is [...] a massive, heterogeneous, but for the most part accidental and disordered, public archive".

Here, these four descriptors, i) massive, ii) heterogeneous, iii) accidental and iv) disordered, will be used to navigate the predicted challenges that a sociolinguist (and probably other researchers) may experience in trying to formulate research questions to ask of YouTube data.

The first challenge is how massive YouTube is. While having an abundance of data can be a positive aspect, the size of the exponentially growing platform is truly overwhelming. Nearly a decade ago, it was estimated that the total size of YouTube was 448 million videos with an aggregated length of 2,649 years (Ding *et al.*, 2011). One advantage of other sources of raw data and of found data is they naturally constrain what research questions can be asked through the data's qualities. These limitations could be the number of speakers, amount of data per speaker, the topics being discussed or the location of the recording, for example. On Twitter potential research questions can naturally be constrained by searching by location, tweet text or other content (e.g. GIFs, emoji or hashtags). But it is unclear how a researcher would begin to assess these aspects with a source like YouTube. Even if a research project was narrowed to one content creator, many have been posting videos regularly for many years, amassing hundreds of hours of footage to be considered. Thus, a researcher may find it difficult to decide what data may be interesting to study.

The second challenge is the heterogeneity of YouTube's content. Content ranges from videos of unexpected humorous mishaps akin to home movies, to vlogs of daily life, 'how to' instructional videos, highly edited bricolages and remediation of previous media content, to professional content produced by large corporations. Thus, a researcher has no control over the qualities of the data available to them, just like when using other found and raw data sources. But unlike other types of found and raw data, the inconsistency of content across YouTube makes each video's data qualities highly unpredictable and this is unhelpful in navigating YouTube to select a portion to form a research corpus.

The third challenge is that YouTube is continually evolving. To quote Burgess and Green (2018, p. vi):

> "YouTube has transformed significantly in the past ten years. It has of course continued to grow at dizzying rates, but it has also changed in terms if its business model, its interface and features, its cultural role, and the extent to which it regulates content and behaviours."

Some of these changes have been to the platform itself (e.g. features and functions) and some have been in YouTube user practices (both video creators and consumers). Thus, YouTube is co-created by YouTube Inc and by YouTube users. Both parties may disagree with the statement that their practice has been "accidental", but their cumulative effect has been and is certainly unforeseeable. For example, in regard to user practices, video fads or trends regularly occur, but what creative idea catches on and how long for is unpredictable. Equally, users cannot predict which videos YouTube will promote on their home page or trending page, and thus be encouraged by the platform. Regarding YouTube Inc's practices, in 2020 some significant changes were made in the USA regarding channels that are targeted towards children (see Alexander, 2020) as a result of violating child privacy laws (see Kelly, 2019). How users will respond to these changes is unknown. It may be that YouTube will be abandoned for other platforms, or it may be that creative work arounds are devised to circumvent certain features and tools no longer being available (e.g. push notifications, comments).

YouTube's continual and drastic evolution to date adds another layer of diversity to an already heterogenous data source. Again, it makes predicting a video's data qualities and selecting data to form a research corpus more difficult. This is of particular note if the study is being conducted in relatively real-time as unexpected changes could cut the research project short, require an amendment to the research question, or cause the project to be abandoned.

The fourth challenge to formulating research questions for YouTube data is finding the data. Just like all digital data, YouTube data embodies the four affordances laid out by boyd (2010). However, the searchability of YouTube data is not aligned with sociolinguistic interests and could be described as "disordered" from a sociolinguistic, and other, researcher's perspective.

Technically there are two ways of searching YouTube data: 1) the webpage search box, and 2) the search resource on YouTube's Data API. Most users would be familiar with the search box that sits in the top middle of the screen consistently on all YouTube pages. A user can type any search term they wish into this box and YouTube will retrieve the most relevant content. The data that YouTube uses to decide upon search results includes video titles, video descriptions, channel names, user names, hashtags, and even "the video itself" apparently (YouTube Creator Academy, 2020), which may suggest that video captions are also searched. In deciding what to retrieve the YouTube algorithm also apparently considers "which videos have driven the most watch-time and engagement for a search phrase" (ibid). Further, YouTube tries to personalise the content retrieved to the user. Therefore, what has been searched for and engaged with on YouTube on that user account prior to the current search is also took into consideration.

The second method of searching is using the search source on the YouTube Data API (see YouTube, no date a for full API documentation). What data can be searched for and how depends on what is in the API and how it is structured. The search source includes a parameter that allows any search term to be used, similar to how one would use the search box, and the same resources are searched as would be if the search box was used. One key difference when using

the API, however, is the degree of control in regard to what search optimised resource is being searched. For example, the "location" and "locationRadius" parameters can be used to specify a video's location (if this is included in the metadata) using latitude and longitude coordinates, so one can be certain that videos uploaded within the boundaries of a specific city will be retrieved, compared to entering "Newcastle" into the search box which would retrieve any content related to "Newcastle" (e.g. interviews of Newcastle United football players, or clips from the TV programme Geordie Shore). Second, searches can be more specific along many parameters. For instance, specific dates can be used to limit an API search compared to merely filtering the results from a search box search by "last hour", "today", "this week", "this month" or "this year".

However, regardless of which facility is used, searching YouTube data based on sociolinguistic interests or desired data qualities is very difficult. Most notable is that, unlike Twitter, a researcher cannot search for specific language features of interest on YouTube. Even though the vast majority of videos now include automatic captions it is not possible to search these directly via the search box or the API, so the researcher cannot gain insight into what language or speech a video contains. Theoretically, it would be possible to select a series of videos, download their captions using the API and then export them into another format to search for specific linguistic features, but this is not as streamlined compared to the Twitter API. Of course, one cannot search for speech features but would have to search for specific words that the speech feature of interest may be in and hope the automatic captioning has not mistranscribed the word.

As Caron et al (2017, p. 53) report, in their experience "these tools are not specifically designed for researchers, making our searches a bit like finding a needle in a haystack". Caron et al (2017) provide a detailed description of how they developed a search strategy to find YouTube videos related to social change-oriented young people. While some inspiration can be took from this work the searching was to identify videos with a specific type of content, so unless a research question was built specifically around the way a certain topic affects speech (which is a valid area of research, e.g. Love and Walker, 2013)

other bespoke guidance for finding video data suitable for sociolinguistic research would be useful.

Finally, there is the practicality of how to access YouTube data to analyse it. As was stated above in section 3.1.4, online data can be analysed in real-time as it is consumed, or a copy downloaded to be analysed at a later date. This is also true for YouTube data. As mentioned, like other platforms YouTube has an API from which a copy of certain data can be requested such as the comments, captions, and number of likes for a video. However, video and audio data cannot be downloaded via YouTube's API or it's web interface. Here, the practicalities of collecting video and audio data overlap with ethical considerations and so will be discussed in the next section (3.2).

3.1.6 Summary

Sociolinguistics uses a range of methods to collect linguistic data and other methods, such as ethnography, to collect complementary data to help understand the linguistic data. The key factor in regard to collection methods for linguistic data is the degree of control the researcher has over the data's qualities such as the linguistic content, its quantity and the factors that are predicted to affect its production. The data's qualities constrain what research questions can be asked. In tailored data the data collection method and data qualities will align with the research question the researcher has formulated beforehand. But when using found and raw data a researcher would have to formulate a research question that works within the constraints that the data qualities of the source impose.
A second key issue is discovering sources of found or raw data. Unfortunately, most research papers do not document how they discovered and negotiated access to their found or raw data. Further, these papers do not document how discovering the data and formulating the research question relate. In other words, it isn't clear how often researchers happen upon a data source and are struck with inspiration compared to having a topic or even drafted research question in mind and then search for potentially useful data.

As has already been described, online, public video such as YouTube data would be classified as raw data. However, a notable difference between YouTube data and found and other sources of raw data is that most corpora are finite whereas YouTube is infinite. A second notable difference is most found and raw data sources are relatively internally homogenous, whereas YouTube data is highly heterogenous. A limited amount of data and the data qualities being relatively consistent across the source naturally sets boundaries and constrains what research questions can be asked. And the ability to utilise other online sources that are highly heterogenous and continually expanding (e.g. Twitter) is attributed to being able to search for linguistic features of interest, rather than for pieces of data (in this case tweets) which the researcher hopes will be useful. In comparison, YouTube's apparent infinity and diversity, and its search functions not being aligned with sociolinguistic interests, renders finding sociolinguistically appropriate data and formulating a question to ask that data to be like finding a needle in a field full of haystacks (to extend Caron et al's (2017, p. 53) observation).

3.2 Research Ethics

As is expected with all research, the ethical implications of the work need to be considered. Universities require a formal record of these considerations and a plan for how they will be addressed when performing the research. This plan is critiqued in a review process and only once approved does the researcher have permission to begin the research (sometimes referred to as an Institutional Review Board (IRB)). Overviews of the most pertinent ethical issues within sociolinguistic research (3.2.1), research using online data (3.2.2), and YouTube data (3.2.3) are given below. It is evident that while many ethical issues overlap with offline data and online written data, YouTube data has its own ethical nuances, many of which would be amplified if the focus of the research was speech.

3.2.1 Sociolinguistics and Research Ethics

All sociolinguistic data collection methods require direct interaction with participants who are aware they are being studied (e.g. sociolinguistic interview, experiment). Thus, as in all interventional human subject research, informed consent is mandatory. Along with 1) informed consent, the main ethical guidelines for working with participants are 2) guaranteed anonymity; 3) voluntary participation; and 4) access to researcher and research findings (Tagliamonte, 2006, p. 33).

This second guideline, guaranteed anonymity, has its own ethical complexities when the research topic is speech. First, it should be bore in mind that (as was laid out in section 1.3) each person has a unique voice and so supposedly anonymised data would be identifiable to a familiar listener. How likely this is to happen or what the ramifications of this may be is very difficult to predict, but it should be acknowledged that speech can never be fully anonymised. Thus, how the data will be used as part of disseminating the research should also be made clear to participants. For example, they should be aware that clips from audio recordings or quotes from their transcripts may be used in presentations as well as a part of companion webpages of publications as is increasingly becoming the norm. Further, participants should also be fully informed if the data is intended to be included in a corpus, and how it would be accessed by other researchers. Second, participants may want their contribution to be explicitly attributed to them as acknowledgement of their expertise or to leave a legacy. While Sara Trechter (2013) makes this point in relation to collecting data of a minority language, any participant may object to their data being anonymised.

It goes without saying that a researcher needs approval from their Institutional Review Board (IRB) (Schilling, 2013b) before conducting their research, this process being a formal demonstration of commitment to principles of ethical responsibility (Besnier, 2013). However, while the guidelines of governing bodies (e.g. (Linguistic Society of America, 2009)) can be referred to for guidance, there is variation across the requirements IRBs impose. For example, Sali Tagliamonte (2006) relays the differing requirements from two institution

IRBs in regard to how much detail the participant should be given about the focus of the research project. One institution was satisfied that describing the focus of a project as an interest in language as a part of the history and culture of the community was sufficient, but another institution required the researchers to state to the participants the specific linguistic features that would be analysed.

Further, there can be ethical issues that are difficult to articulate on an IRB form or may need to be responded to reflexively. For example, researchers should take particular care if the community they're engaging with could be defined as vulnerable as well as carefully consider how the speech community, vulnerable or not, is represented in the dissemination of the research (see (Besnier, 2013; Mann, 2013), respectively, for discussions). Sara Trechter (2013, p. 43) summarises "the ethics board may be both too lenient […] and too strict" and argues sociolinguists should be involved with their local IRB to represent and communicate the sociolinguist perspectives on ethics and research.

Finally, as sociolinguistic research has started to venture into CMC some initial discussions and recommendations in regard to ethics have been made. These will be included in the multidisciplinary discussion of online data and research ethics below.

3.2.2 Online Data and Research Ethics

Beyond the ethical considerations that you would expect from conducting research offline, such as how to work with vulnerable populations and represent participants fairly, collecting and conducting research on public online data introduces many complexities. The most prominent complexity is negotiating the issues of public data and informed consent, and a tension between anonymity and credit. These are set within a context where site terms of service, the law, and user preferences may jar or directly conflict.

A key question is whether informed consent is required from the data producer before one can collect and analyse it and publish the results. At first, the

distinction between private data and public data was the primary factor in this decision making. Initially, for example as Susan Herring (1996) stated in the introduction of her edited volume, 'public' was defined from a technical perspective and thus synonymous with open-access. Thus, public data was viewed as not being bound by the same ethical restrictions in regard to informed consent as traditional research with human subjects. Indeed, some research disciplines and Institutional Review Boards (IRBs) still take the view that informed consent is not required when collecting public data because the data is not collected through intervention or interaction with those producing it (Bruckman, 2014; Vitak, Shilton and Ashktorab, 2016; Vitak *et al.*, 2017). This is evidenced, for example, by very few papers that use Twitter data mentioning an IRB or ethical review process (Zimmer and Proferes, 2014).

However, the "Taste, Ties and Time" (Lewis *et al.*, 2008) controversy (the public release of a Facebook dataset that contained the profiles of an entire cohort of college students from a US university that quickly became deanonymized) prompted many to review this practice, with several papers that included a discussion of the ethical issues (e.g. (Zimmer, 2010; boyd and Crawford, 2012)) being published in the aftermath. The scandal prompted researchers to reflect on consent and the use of public data, and its anonymity.

A shift has begun towards conceiving public / private as far more nuanced in online contexts and thus the decision-making around the collection and study of such data requiring more critical consideration. For example, Bolander and Locher (2014) cite Landert and Jucker's (2011) model of mass media communication in their discussion of methodological issues in sociolinguistic investigations of online data. This model considers the actual content of the data, such as topic, as well as the communicative context when defining 'private' and 'public'. Helen Nissenbaum's (2004) theory of 'contextual integrity' is also often cited in discussions of online data and privacy. Her rationale is that we expect different levels and types of privacy in different contexts and this conception of privacy is tied to information gathering and dissemination norms. Fiesler and Proferes's (2018) survey of Twitter users found that few knew that researchers could use their content. Also, their responses to this differed greatly

and were highly contextualised, depending on who was conducting the research what it was about, what kind and how much data was being used and how it would be disseminated. Prior to this, Williams et al (2017) also found Twitter users' responses to the idea of their data being used was also highly contextualised. Their participants were least concerned for their data to be used to conduct research at a university but would be far more concerned if a government or commercial body was using the data. Therefore, researchers should bear in mind that, to quote danah boyd and Kate Crawford (2012, p. 672) "Just because content is publicly accessible does not mean that it was meant to be consumed by just anyone".

Of course, gaining informed consent from data producers to include their data in your research would circumvent these issues. In regard to sociolinguistics in particular, Alexandra D'Arcy and Taylor Marie Young (2012, p. 540) argue that "the notion that informed consent can be curtailed in virtual spaces undermines the ethical principles that underpin sociolinguistic research, which crucially acknowledges the subjectivity of social settings". They provide practical advice for conducting ethically sound sociolinguistic research on Facebook including recruitment, the researcher's role, and how to end a project, as well as consent. However, in some topic areas there is a strong trend for researching online data to be synonymous with researching 'big data'. While the definition of big data is highly debated, and the size of the dataset is not a defining characteristic, generally these datasets are too large for researchers to be able to analyse them using manual or traditional methods. Equally, it can be argued that they are too large for researchers to be expected to gain consent from all of the data's content creators (boyd and Crawford, 2012). Thus, researchers who access Twitter data, for example, do not seek consent from each Twitter user, but also do not inform users that their data has been collected for research purposes either (Fiesler and Proferes, 2018).

The other key issue is being able to ensure anonymity. First, as the "Taste, Ties and Time" (Lewis *et al.*, 2008) controversy brought to the fore, anonymising data retrieved online is very difficult. Considering the affordances of digital data (see section 3.1.6) reveals the multitude of ways in which the original of

supposedly anonymised data can be found. First, the persistence (the ability to record and archive) of digital data means that it is possible for it to exist indefinitely. Thus, in contrast to analogue documents or even digital data held on private university computers, it is not possible for a researcher to guarantee anonymity by destroying the original data post-analysis. Second, the replicability (the ability to make exact duplicates of these records) of digital data means that it has the potential to 'travel' without the creator's permission and with little indication that this has happened. For example, a Twitter account may tweet data taken from Reddit, or a Facebook account may upload clips from videos originally posted TikTok. Thus, even if it was possible for the participant to destroy or privatise the original public data it cannot be guaranteed that public copies exist elsewhere on the internet. Third, the searchability of digital data means that very little resource may be needed in order to find data. With most social media sites including a search box, and public data also being accessible via a range of search engines, quotes apparently anonymised in publications can instantly be found. To prevent this, researchers can refer to Amy Bruckman (2002) for guidance. Her guidelines describe different measures that can be taken to ensure different levels of data "disguise". For example, a "light disguise" would include changing names and locations but including verbatim quotes, where as "heavy disguise" would include changing names and locations as well as rewording quotes and possibly even adding false information, e.g. if researching a Facebook group dedicated to sailing, changing the topic to another sport as long as this would not affect the findings. Further, in manipulating the data for disseminating the research, the researcher must also address the possibility that providing a detailed report of the data collection process may make it possible for others to repeat it, providing them with the original, deanonymized data.

A final issue is the tension between anonymity and credit. Bruckman (2002) suggests that thinking of all those who post content on the internet as 'amateur artists' allows for greater reflexivity in considering whether to anonymise or attribute data. However, social media sites may give specific guidance on this in their terms and conditions. This may be either that any data published elsewhere should be verbatim and attributed to its creator or that all data must be

anonymised, suggesting a researcher has no choice in the matter. But whether following a site's terms of service is a legal requirement remains debated (Brake *et al.*, 2020), with the latest work highlighting that terms of service are ambiguous, inconsistent, and lack context (Fiesler, Beard and Keegan, 2020). Hence, scholars have begun to argue that ethical decision-making should not be limited to the site's terms of service but prioritise the specific circumstances of the research (Fiesler, Beard and Keegan, 2020). Further, when asking social media users, 90% of Williams et al's (2017) respondents and 58.2% of Fiesler and Proferes's (2018) respondents wished to be anonymised if their tweet was used in research dissemination, illustrating how following terms of service may conflict with participants' preferences.

The ethical issues surrounding the use of online public data in research is continually evolving as the technology does. While the review above clearly indicates how complicated a researcher may find ethical decision making, there are some guides and advice available, such as (Townsend and Wallace, 2016; Williams, Burnap and Sloan, 2017). The primary guidance of ethics and conducting research on online data is from the Association of Internet Researchers (2020). Advocating a case-by-case approach, this document guides researchers through the different stages and phases of their research project, encouraging reflection on the ethical implications and risks related to research practices. This practice, described as an "Ethic as method, Method as Ethic" stance by Annette Markham (2012), emphasises that

> "our choice of methods vis-à-vis given research questions and design evoke specific ethical issues – but these in turn (should) shape our methodological choices" (Brake *et al.*, 2020, p. 4).

3.2.3 YouTube Data and Ethics

As is evidenced above, the discussion of ethics and social media research has so far focused on considering the use of written data. This is probably as a result of most research data being written content but also reflects how different social media platforms respond to the idea of the data housed on their platform being

used by researchers, for example Twitter's dedicated service to researchers mentioned in 3.1.6. What discussion there has been (e.g. Highfield and Leaver, 2016; Caron *et al.*, 2017; Patterson, 2018) evidences a similar situation to text data in that there is a lack of consensus across researchers in regard to informed consent.

Just like written data, a key question is whether informed consent is required from the data producer before one can collect and analyse it and publish the results. In the previous literature, there is evidence of researchers asking permission (e.g. Harley and Fitzpatrick, 2009; Frobenius, 2014) but also including data where they didn't get a response (e.g. Frobenius, 2011). In other cases, because ethics is not mentioned, it appears that researchers treated the data as not requiring informed consent (eg. Adami, 2009; Porter and Hellsten, 2014; Choi and Behm-Morawitz, 2017). Here, the data would have been categorised as 'public data' and so informed consent for its use deemed not necessary. However, just as the public/private distinction has begun to be considered with more nuance in other social media sites, Patricia Lange's (2007) work evidences that it should be in YouTube too. Through a one-year ethnography she identified how the video content that YouTube users create, and the manipulation of sharing video and other data produces varying degrees of 'publicness'. For example, participants were 'publicly private', sharing their personal identity but restricting who can access their videos, as well as 'privately public', sharing one's video data but keeping personally identifying information private. The content of the videos also feeds into defining its position along the private to public continuum. For example, Lange (2007) found that while her participants made some videos publicly available their content was designed to only be accessible to or understood by a specific intended audience. In other words, the video content was based around in-jokes. Therefore, to repeat boyd and Crawford (2012, p. 672), "Just because content is publicly accessible does not mean that it was meant to be consumed by just anyone".

Another key issue in regard to YouTube data is its 'collection'. All the works referenced above have taken qualitative approaches where their analyses can be performed by streaming the video. ""Streaming" means a contemporaneous

digital transmission of the material by YouTube via the Internet to a user operated Internet enabled device in such a manner that the data is intended for real-time viewing" (YouTube Great Britain, 2010, p. 5.L). Thus, the raw video data is not collected, copied or retained by the researcher, and so there are minimal ethical concerns in this regard. There are software tools such as NCapture and NVivo (QSR International, no date b), to assist in analysing streamed video. NVivo is a video (and other data) annotation software, and NCapture is a web browser extension that links to YouTube and then integrates with NVivo to allow a video to appear in its interface. With its ability to provide timestamped transcripts and annotation, and then automatically quantify the coding, NVivo would be accommodating to auditory analysis (for a description of auditory analysis see section 3.6). However, NCapture merely links NVivo to a YouTube video, thus if the video were to be removed from YouTube NVivo would no longer be able to access the video, leaving the analysis without the material it is referencing (QSR International, no date a).

In comparison, research that requires a copy of the video data to be available outside of the YouTube interface is more complicated ethically. YouTube's terms of service state that video should only be streamed and that it is "not intended to be downloaded (either permanently or temporarily), copied, stored, or redistributed" (YouTube Great Britain, 2010, p. 5.L). Thus, downloading data would be violating these terms. Regardless, there are a multitude of 3rd party websites from where it is possible to download audio and video data posted on YouTube (the ethical and legal issues in this regard are discussed below in section 3.2.3). While rare, there is evidence in the literature of researchers downloading data to analyse. Almost all this work comes from Biel and Gatica-Perez and their colleagues (Biel and Gatica-Perez, 2010, 2011, 2013; Biel *et al.*, 2013; Gatica-Perez *et al.*, 2018) but all these studies seem to be analysing a portion of the same dataset (first mentioned in Biel and Gatica-Perez, 2010) that was collected in November 2009. Unfortunately, based on publicly accessible records (e.g. the YouTube Engineering and Developers Blog (YouTube, no date)), it is unclear whether such downloading was permitted in the Terms of Service at the time. But overall, previous work does suggest that the downloading of video from YouTube is not typical research practice. Further,

researchers may consider downloading data even if the analyses does not require a copy of the data to be retained. Not downloading the video data leaves researchers at the mercy of YouTubers who may privatise or delete videos whenever they choose as Caron and colleagues (2017) found out.

In contrast, there appears to be less of a tension between anonymity and credit, with almost all researchers identifying the YouTuber producing the data regardless of informed consent (Adami, 2009; Harley and Fitzpatrick, 2009a; Porter and Hellsten, 2014; Choi and Behm-Morawitz, 2017). Further, creating content for YouTube has become an established profession, an industry with identifiable sectors (e.g. beauty, gaming), and specific governance (e.g. advertising and marketing (The Advertising Standards Agency and The Committee of Advertising Practice, 2020)). While there is a continuum of success along dimensions such as income and social media based measures (e.g. social media statistics such as number of followers/subscribers or number of likes), some carry the markers of a traditional celebrity (e.g. brand deals, management by talent agents, coverage in tabloids) (Abidin, 2015; Bishop, 2018). Sophie Bishop (2018) refers to these content creators as 'A list Vloggers'.

3.2.4 Summary

To summarise, this section has given an overview of the most prominent ethical issues within Sociolinguistic research, and research using online data and YouTube data. This indicates that, while many ethical issues overlap with offline and online written data, YouTube data has its own ethical nuances, many of which would be amplified if the focus of the research was speech. Further, there is a lack of guidance to assist researchers in navigating these issues. Thus, it is evident that ethics and using YouTube data when researching speech is a topic that needs to be explored.

3.3 Selecting Linguistic Variables

The central concept to variationist studies is the linguistic variable. Prior to Labov's pioneering study in Martha's Vineyard, it was believed that linguistic data should be free from 'inconsistencies' in order for it to be studied, and that any variation was unpredictable. Now, after 50 years of variationist research, it is agreed that variation is systematic, not random, and is inherent to language. The simplest definition of a linguistic variable would be that it is 'two or more ways of saying the same thing' (Tagliamonte, 2012, p. 4). In other words, it is an element of language that has multiple forms, but all these forms perform the same function (Tagliamonte, 2006). For example, the linguistic variable 'ing' (the suffix on present participle verbs) can be pronounced in two ways: 'waiting' or 'waitin'. So, 'ing' and 'in are two speech features of the variable 'present participle verb suffix 'ing''. Whether 'ing' or 'in' is used does not change the linguistic meaning of the word 'waiting'/'waitin'. How often different speakers in different contexts choose to say "waiting" over "waitin" or vice versa does indicate the social meanings of the speech features 'ing' or 'in', however.

Fundamental to the linguistic variable is the principle of accountability. Essentially, this is to count the number of times a speech feature had the potential to occur and the number of times it actually occurred and then compare (Tagliamonte, 2006). Extending the ing/in example, a researcher would identify all the present participle verbs in the data. So, in addition to 'waiting', whether 'ing' or 'in' was used in the words 'sitting', 'eating', and 'walking', for example, would also be examined. However, the words 'wing' or 'sting' would not be examined. This is because although they end in 'ing' they are not present participle verbs. They are a noun and a present simple verb, respectively. So, the "ing" in "wing" and "sting" is not equitable to that in "sitting" or "eating".
It is advised that the linguistic variables studied be i) frequent (so data is plentiful), ii) with adequate variation (as one form dominating and the other being rare is less likely to be due to interesting interactions between social and linguistic factors) and iii) be timely and relevant to sociolinguistic research (Tagliamonte, 2006).

Below, how sociolinguistic variables are identified for analysis (that is the thought and practical processes a researcher could take in this decision making) is considered in projects that focus on offline (3.3.1) and online (3.3.2) data. This section outlines how the backdrop on which linguistic variables are selected is 'place' whether that is offline speech data or online written data. But place is likely to be ambiguous in YouTube data. This indicates that considering strategies to define place would be a fruitful topic to explore.

3.3.1 Offline Sociolinguistic Variable Identification

As Tagliamonte states, when considering a research project "you may not know in advance which feature(s) you want to study" (2006, p. 83). Thus, work in sociolinguistics can be separated into 2 categories: 1) studies that examined a set of predetermined variables (a 'top down' approach), and 2) studies that discovered variables of interest by sifting through the data (a 'bottom-up' approach), and it is not difficult to map these approaches onto tailored and found/raw data, respectively.

To carry out the first approach, Tagliamonte advises: "The place to start is to take a long, hard look at your data" and "take notes about the things you observe" (ibid, p.79). When making such observations she points out that "you tend to notice things that are different from your own idiolect" but equally "variables will slip by without you even realising they are there" (ibid, p.78). Upon the first hints that a speech feature may be interesting, a researcher must authenticate it using the three criteria defined above: i) frequency, ii) adequate variation, and iii) be timely and relevant to sociolinguistic research. Here, Tagliamonte (2012) recommends establishing that the variable is robust (there is adequate variation throughout the data) first. Then, assessing its frequency by counting how many tokens are in an arbitrarily measured amount of data, such as a certain number of minutes of continued speech or certain number of words. "If there are not enough to warrant a study, stop and find another linguistic feature" (ibid, p.111). Finally, the literature can be surveyed and sociolinguistic theory applied to see if a study of that variable is likely to be of interest to sociolinguistics. This account, and written reports of finished research projects,

makes it appear that applying a top down approach is a straightforward way of identifying a linguistic variable. However, in reality a researcher would probably be moving back and forth between their data and the literature, possibly rejecting several, even many, variables before settling on the focus of their study.

Such an approach could be perceived as high-risk. The apparent lack of structured activity for engaging with data to identify potential variables, and clear parameters for robustness and frequency to verify them, suggests selecting variables is reliant upon researcher intuition. One strategy to use would be to find a "super token" to verify a potential variable. A super token is where a speaker uses multiple speech features of a variable in the same sentence or short period of speech. Although Tagliamonte (2012, p. 111) recommends these in regard to research dissemination (to write in papers or use in presentations), their ability to clearly indicate said variable varies within speaker, and thus suggesting it varies between speakers too, could be harnessed much earlier. Still, however, one could liken this approach to a fishing trip: throwing out a net and seeing what you can catch. Although, Tagliamonte (2006, p. 79) promises that there will be fish in the water: "variation is everywhere; you just have to notice it".

Taking the second approach ('top down'), addresses some of the criteria that a linguistic variable should meet. Here, the researcher selects the variables of interest in advance through engaging with literature, and so can ensure such a study would be timely and relevant to sociolinguistic research. Also, this approach gives researchers the opportunity to design the data collection activities to try to ensure that participants produce a sufficient number of tokens of the linguistic variable being studied. This addresses the criteria of frequency.

Regardless of the approach taken, the backdrop on which linguistic variables are selected is 'place'. The earliest studies of how speech varied focused on geography, with researchers drawing maps with lines separating where speech features were and were not used (Chambers and Trudgill, 1998). Here, the focus was on mapping the variation and not explaining why it occurred. Then, with the advent of variationism (e.g. Labov, 1966), how language patterned across macro social categories (e.g. gender, class) within specific locations became the focus,

with the size of these locations and thus their populous shrinking over time as researchers became more interested in the people that were speaking and their identities (e.g. Eckert, 1989). Still, place was resigned to be a methodological construct. Thus, as this quote for Barbara Horvath (2013, p. 8) summarises:

> "The notion of "place", a geographical category referring to the embodiment of interacting sociocultural practices in a locality, remains central to most studies of linguistic variability, although the use of the concept of place in the explanation of linguistic variability is not often invoked".

However, contemporary research has begun to take a more explicit interest in space and place, a review of which will be provided later on in this dissertation (section 6.2).

Speech is related to place in that the way we speak is dependent upon who we come into contact with, our speech patterns reflecting the social groups that we want to be affiliated with. When selecting linguistic variables, place can be likened to a filter - the imagined boundary around a location doubling as a barrier to less relevant linguistic variables and a net to catch relevant potential linguistic variables. Here, relevancy is in relation to the speakers that make up that place. In other words, defining a place simultaneously defines the speech that is likely to be familiar and socially meaningful to the speakers within that place. Most simply, 'place' provides the speech resource that a researcher can scour for fruitful linguistic variables to study. However, the methodological role and value of place seems to often go unnoticed. The need to define place is a given in offline sociolinguistic work and is one of the first, if not the first, element of a study to be defined in both research practice and reporting. Even studies that on the surface do not seem preoccupied with place have had the linguistic variable limited by geography in some way. For example, the linguistic variable examined by Kirkham and Moore (2016) in their study of a politician's speech was motivated by *British* media commentary. To conclude, because speech features are tethered to a space/place to some degree, place is a vital factor in identifying linguistic research variables although it has only just begun to explicitly be considered as an analytical construct.

3.3.2 Online Sociolinguistic Variable Identification

Currently, in the literature there are no discussions or guidance for identifying linguistic variables for sociolinguistic studies of online data. However, insights can be gained from revisiting the studies of online language reviewed in section 2.3. Just like for offline studies, the literature that examines language variation online and that uses social media as a proxy for offline communication can be separated into 2 categories: 1) studies that discovered variables of interest by sifting through the data (a 'bottom-up' approach'), and 2) studies that examined a set of predetermined variables (a 'top down' approach), the latter being much larger than the former in terms of the number of papers.

The few studies that have took the approach of discovering variables of interest by sifting through the data can be described as "Big Data" studies. Both Shoemark et al (2017) and Pavalanathan and Eisenstein (2015) used similar complex statistical modelling to sift through hundreds of millions of tweets and allow those that were likely to be of greatest interest to 'rise to the top'. For Shoemark et al (2017) these were words that were the most distinctive to tweets geo-located in Scotland, which were then manually pruned to 113 words. For Pavalanathan and Eisenstein (2015) these were the top 30 or so words found in tweets that were geo-located to one of the ten areas of the US that they had identified. Thus, the affordances of digital data allow for tonnes of written data to be systematically, reliably and efficiently considered, rather than relying on researcher intuition.

The second category are studies, where a set of variables of interest are defined in advance of data collection, is also dominated by "Big Data" studies. In some cases (e.g. Gonçalves and Sanchez, 2014; Huang *et al.*, 2016; Grieve *et al.*, 2019) these were a predetermined set of concepts and their alternations (different words with the same semantic meaning). In others, the researcher had defined a set of nonstandard spellings that reflect different ways of speaking (e.g. Eisenstein, 2015; Jones, 2015). There is even an example of looking at one word specifically: "anymore" (Strelluf, 2019). Here, the digital affordance of

searchability is paramount as the researcher can quickly assess whether the variable in question is in the data, and if so, how many times.

Importantly, throughout all these studies 'place' played a key role in variable identification. As already mentioned in 2.3.1.2, geotagging in Twitter data has been utilised by many researchers. Geotagging has been used to collect country wide data, for example across the US, Scotland and the UK (Huang *et al.*, 2016; Shoemark *et al.*, 2017; Grieve *et al.*, 2019, respectively), and even world wide data (Gonçalves and Sanchez, 2014). Then geotagging has allowed analyses to consider linguistic variation by postal code (Grieve *et al.*, 2019), city (Pavalanathan and Eisenstein, 2015; Strelluf, 2019), county (Eisenstein, 2015) and country (Gonçalves and Sanchez, 2014). In studies where the data was not geotagged, place was still explicitly related to the data in some way. In both Siebenhaar's (2006) and Androutsopoulos and Ziegler's (2004) studies of internet relay chat channels the researchers used hashtags (e.g. "#bern", "#zuerich", and "#mannheim", "#bremen", respectively) to structure their analysis by relating data to place. Even in qualitative work, where a specific variable is not chosen, where the participants are from and where they currently live is fundamental to interpreting their linguistic behaviour. Thus, the analysis is constructed through a 'here' versus 'there' distinction. For example, Chicago US versus Mexico in Christiansen (2015), and participants from Greece versus Hamburg in Androutsopoulos (2015). In other cases, the construction is 'here' versus 'elsewhere' emphasising the multinationalism of online spaces and online networks, such as Sharma's (2012) study of Nepalese students and Seargeant et al's (2012) Thai speakers. Finally, Ilbury (2019) demonstrates how the 'here' versus 'there' distinction can be a mismatch between where a user is from and where the linguistic features they are using are from. He argues "in social media, where geographically disparate users forge connections with users beyond their own physical community, the potential for non-local linguistic styles to be (stylistically) appropriated is arguably increased" (ibid, p.4). To summarise, it is evident that although online is devoid of space, place remains a vital factor in identifying and defining research variables.

3.3.3 Summary

To summarise, this section has reviewed the approaches used to identify potential linguistic variables and the criteria by which they are assessed has also been discussed, namely how frequent, robust, timely and relevant they are. In doing so, how the affordances and qualities of offline and online data impact data collection and analysis methods has highlighted the role of place in identifying potential linguistic variables. Further, unlike geo-tagged Twitter data or data from users who have moved away from where they grew up to a new country, place is likely to be ambiguous in YouTube data. This indicates that considering strategies to define place would be a fruitful topic to explore.

3.4 Statistical Analysis

As was outlined in chapter 1, while language can be analysed both qualitatively and quantitatively and mixed-method studies are common, the work in this thesis aligns itself with variationist sociolinguistics, a quantitative approach, and involve statistical analysis of some kind. Variationist Sociolinguistics examines speech *variation* and *change*, these two interests potentially requiring different statistical tests. Speech *change* and its statistical analysis is more relevant to this thesis because YouTube has the potential to provide data which allows time to be operationalised with finer granularity, and thus as a continuous variable, in comparison to current offline studies. This doesn't necessitate statistical innovation, but I argue that this prompts time's role as an antecedent to be considered more closely. How time has been considered and thus statistically analysed in sociolinguistics to date will be reviewed below.

3.4.1 Time in Sociolinguistics

When considering speech change over time there are predominantly two types of research design: i) real time and ii) apparent time.

Real time studies examine data from multiple points in time. In other words, they are longitudinal studies. There are two types of real time study: i) panel study and ii) trend survey. In a panel study, data is collected from the same participants at multiple points in time (Cukor-Avila and Bailey, 2013). An example would be The Montreal French Project. As Gillian Sankoff (2018) reports, this data collection was never intended to be a real time panel study, but two graduate students who helped collect the initial dataset in 1971 seized the opportunity to reinterview participants in 1984, and then again in 1995. In a trend survey, the second type of real time study, data is collected from the same population multiple times, but not the same people. In other words, a researcher replicates the data collection methods used in a previous study and then considers their newly collected data in relation to the older data. There are several examples of these where contemporary researchers have replicated the data collection performed in some of the first studies that examined speech variation and change. For example, Jennifer Pope returned to Marta's Vineyard in 2002, forty years after Labov's original study (Pope, Meyerhoff and Ladd, 2007), and Fowler (1986, described in detail in (Cukor-Avila and Bailey, 2013)) and Mather (2012) replicated Labov's study of Department stores in New York city 24 and 48 years later, respectively.

Real-time studies are uncommon in sociolinguistics. They require significant planning for the future, and a considerable amount of time to pass between data collection activities, much more time that the duration of research projects typically allows. Further, there are likely to be difficulties in relocating participants to be reinterviewed, participants may decline being reinterviewed, and participant demographics which need to be carefully balanced across the dataset may have changed (e.g. class) (Cieri and Yaeger-Dror, 2018). To address these barriers, a specific research design evolved: the apparent-time study. First used in Martha' Vineyard by Labov (1963), in an apparent-time study data is collected from multiple generations of speakers (e.g. older, middle-aged, young adult speakers and sometimes teenagers or even children) at relatively the same point in time. Because it is believed that a person's speech patterns are relatively fixed by young adulthood, it is inferred that, say, the older participant's speech still reflects the way they spoke 40 to 50 years ago. So, the hypothesis is that

comparing the speech of people from multiple generations will mirror real time speech change (Cukor-Avila and Bailey, 2013).

Another approach to take in real-time studies is to focus on the speech of one speaker. While most real-time work, as evidenced by the studies referred to above, concerns the speech of many speakers from different social categories (e.g. gender, class, age), there is a relatively small but notable collection of studies that each take advantage of the recordings of a single, public figure (raw data). Harrington, Palethorpe and Watson's (2000b) analysed the speech of Queen Elizabeth from 9 of her annual Christmas speeches across 36 years. Cham (2016) examined David Attenborough's speech by accessing 9 episodes from a variety of nature documentaries that he narrated over a period of 55 years. Rodríguez (2019) used recordings of 15 sermons to study the speech of pastor John Piper across 37 years. Similarly, Stanley and Renwick's (2016) data was 115 recorded sermons from the Mormon leader Tom Perry across 43 years. Shapp, LaFave and Singler (2014) accessed archival records of Supreme Court proceedings to examine the speech of Ruth Bader Ginsburg from several years in the 1970s and then 1993 until 2011. And Wallace (2006) analysed the speech of Cheryl Fernandez-Versini (née Tweedy, formerly Cole) in four recorded interviews spanning 12 years. Of course, in such work both the research question and the findings are specific to that individual and so the conclusions are less generalisable. However, according to Cieri and Yaeger-Dror (2018, p. 67), "[i]n this era of data ubiquity, while longitudinal case studies may seem less enticing than large panel studies, […] no panel study is possible without a case study component".

3.4.2 Statistical Analysis and Time in Sociolinguistics

The ways in which real-time studies of single speakers operationalise and statistically analyse the variable 'time' is of particular interest to this thesis. Generally, to date this type of research has only had a few data points for time to reflect the number of recordings (e.g. 4 in (Wallace, 2006)) or as a result of multiple recordings being coalesced to compare the speech produced in each

decade (e.g. both (Harrington, Palethorpe and Watson, 2000; Shapp, LaFave and Singler, 2014; Stanley and Renwick, 2016; Rodríguez, 2019)). Most importantly, the analysis in some of this work has compared phonetic realisations at different points in time (using tests such as ANOVAs and t-tests) but not considered time itself as a variable in the statistical analysis (e.g. (Harrington, Palethorpe and Watson, 2000; Stanley and Renwick, 2016; Rodríguez, 2019)). However, both (Stanley and Renwick, 2016; Rodríguez, 2019) are posters indicating research may be in the earlier stages.

Where time has been included as a variable in statistical analysis, not just in studies of single speakers but also other real-time and apparent-time research, it has predominantly been operationalised as a categorical variable. Interestingly, in their paper Shapp, LaFave and Singler (2014) describe the running of several models where time was operationalised in different ways, such as "Era" (comparing speech when Ginsburg was a lawyer compared to when she was a Justice), "Decade", and "Term" (a court term begins October and ends the following June). This carving out of time into different periods reflects a variety of hypotheses the researchers had for why Ginsburg's speech may change. Similarly, Cham (2016) reports operationalising time as both decade and individual year but only decade reached statistical significance. They argue that this indicates that "year on year differences in Attenborough's usage were insignificant, but clear differences could be distinguished in his speech from different periods in his life" (ibid, p.20).

The contemporary statistical approach used in sociolinguistics, including studies that consider change over time, has typically been mixed-effect regression modelling (e.g. Wallace, 2006; Shapp, LaFave and Singler, 2014; Cham, 2016). Mixed-effect regression modelling is a type of multivariate analysis. Multivariate analysis allows observations to be made on multiple variables simultaneously. As sociolinguistics investigates the impact of many social (e.g. gender, class) and linguistic factors (e.g. word type, such as adjective, noun, verb) upon the speech feature used, multivariate analysis is a common approach. The intention of a multivariate analysis is to ascertain which predictor variables (independent variables) contribute to explaining the response variable

(dependent variable). Interactions between predictor variables can also be tested to see if their cumulative effect influences the response variable (such as combining gender and class variables to see if the predictor variables of 'male working class', 'female working class', 'male middle class', and 'female middle class' can explain variation in the data). Regression modelling assesses the power that one predictor variable has over the response when all the other predictors are held constant. In other words, regression can be used to understand how the response variable changes when one of the predictor variables change, thus assessing the relationship between the two. One advantage of regression modelling is its ability to accommodate many different data types at once. The dependent variable may be dichotomous (so mixed-effects logistic regression is used) or continuous (so mixed-effects linear regression is used), and the predictors can all be continuous ('regression analysis'), categorical ('analysis of variance'), or a mix of the two ('analysis of covariance') (Baayen, 2013). So, the great diversity of variables that may be considered in sociolinguistics are accomodated for.

3.4.3 Summary

To summarise, there is a paucity of research that operationalises time as a continuous variable in sociolinguistics. This is predominantly as a result of tailored real-time studies being resource intensive and high risk in a number of ways. Even studies that have accessed found or raw data do not analyse a sufficient amount of data across enough time points to justify operationalising time as continuous.

In comparison, YouTube has the potential to provide many more data points for the variable "time". Although currently these could cover a period of 15 years maximum, with many vloggers posting videos multiple times a week there may be a pool of thousands of videos for one speaker. Such an increase in data points allows the variable "time" to be rationalised as continuous, rather than categorical. However, this less typical operalisation of time does not pose the need for statistical innovation in itself because mixed effects regression

modelling can accommodate both continuous and categorical predictor variables simultaneously. Rather, time becoming a continuous variable may prompt its role as an antecedent to be considered more closely. Also, other time related variables may become available, and thus novel kinds of research questions could be asked, as a result. This could then open sociolinguistic studies up to novel statistical analyses.

3.5 Thesis Research Aims and Questions

The overall question that this thesis will address is "how can we conduct sociolinguistic research using online public video?" with one aim being that the output of answering this question is a set of guidelines for researchers to refer to guide their practice. To address this overarching question, four key aspects that construct a sociolinguistic research method have been identified and their related literature reviewed. In doing so, why typical sociolinguistic practices do not easily transfer from studies of offline speech and language or language-online to studies of speech-online have been considered. As a result, this thesis' four research questions, that relate to i) Formulating Research Questions, ii) Research Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis, are defined below.

3.5.1 Formulating Research Questions

From reviewing sociolinguistic practices in regard to tailored data collection methods, formulating research questions for tailored, found and raw data, and research practices for online written data, it is evident that researchers would need to make significant adjustments in order to investigate YouTube data from a sociolinguistic perspective.

When attempting to formulate a research question to ask of found or raw data the researcher is in dialogue with the literature and the data's qualities to find a

middle ground. In comparison, YouTube is a near infinite and a highly heterogenous source that is difficult to search and so it is not possible for researchers to look to its data qualities to impose boundaries on what can be asked. Equally, found or raw data could be happened upon by chance and inspiration for a research question serendipitously struck. But this undirected approach could take considerable time and effort for little gain. Thus, thesis research question 1 is:

- What strategies or approaches could a researcher use to i) find sociolinguistically interesting YouTube data and ii) formulate appropriate research questions for that data?

3.5.2 Research Ethics

The ethical implications around using YouTube data, and online data in general, for research purposes are complex. Thus, practice differs considerably across the research fields that use such data, and the expectations of Institutional Research Boards that approve such practices differ also. Further, as is now acknowledged for all research that incorporates online data (Brake *et al.*, 2020), there cannot be a 'one size fits all' approach and the ethical considerations for each research project needs to be assessed individually.

At the outset, taking a sociolinguistic approach to researching YouTube data does not bring any new ethical considerations to the table. However, already pertinent ethical considerations are brought to the fore or take on greater nuance as the research focus moves to speech, namely because speech is a unique and highly personal characteristic. The main ethical considerations are the tensions between i) anonymity and credit, ii) public data and informed consent, and iii) terms of service and data collection requirements.

While some ethical guidance for conducting research using online data already exists there is a paucity in regard to online audio and video, and no guidance for researching speech in particular. Thus, thesis research question 2 is:

- What are the ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?

One aim of this thesis is to produce a resource that will guide researchers whose focus is speech in online public video in considering the ethical implications when designing their research method.

3.5.3 Selecting Linguistic Variables

Selecting linguistic variables to study is a complicated decision-making process that is partly dependent on the type of data (e.g. tailored or found/raw), its content (ensuring the variable is frequent and robust), and the current topics and theories of interest (ensuring the research is timely and relevant to sociolinguistics). Also, this decision-making is performed with 'place' as the backdrop. This may be visiting the city to recruit and interview speakers in offline studies or using geo-tagged data in online studies, for example. Place is central to sociolinguistic studies because it defines the speech that is likely to be familiar and socially meaningful to the speakers within that place and therefore fruitful to study. Thus, selecting linguistic variables to study is further complicated in YouTube data because place is ambiguous. This leads to the third thesis research question:

- What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?

3.5.4 Statistical analysis: Operalisation of time

One predicted advantage of YouTube data over other types is that time can be considered in greater granularity, and thus operationalised as a continuous rather than categorical variable. While the typical statistical approach in contemporary sociolinguistics (mixed-effects regression analysis) would be able to accommodate time as a continuous predictor variable alongside other continuous and categorical predictor variables, whether other statistical analyses would be

more appropriate or could provide different insights should be considered. Further, time becoming a continuous variable may prompt its role as an antecedent to be considered more closely, and allow other time related variables to become available for inclusion in statistical analysis. Thus, thesis research question 4 is:

- What statistical approaches could be used in studies of speech in online public video considering that time can be operationalised with greater granularity?

3.6 A note on acoustic and auditory analysis

When analysing speech the researcher has to decide whether to take an acoustic or auditory approach. The production of speech is a physical phenomenon: movement causes pressure fluctuations in the surrounding air, which travel in consecutive waves across space and time to the eardrum, which vibrates in response and thus movement is realised as sound. A variety of features of speech sounds can be measured objectively along the dimensions of time, frequency (rate of vibration), and amplitude (size of vibration) (Johnson, 2012). Measuring these aspects is referred to as 'acoustic analysis'. To conduct an acoustic analysis specialist software (such as Praat (Boersma and Weeink, 2018)) is required and the audio recording must be in a compatible format.

Acoustic analysis requires good quality recordings, "good quality" being defined in two ways. First, in terms of acoustic content. That is, ideally, with little to no echo, background or foreground noise. In comparison, auditory analysis can be performed on a fairly good quality recording that may include such features. Second, from a technical perspective. The format of the recording is fundamental to the ability to take frequency (Hz) measurements. The recording should be a WAV file and have a sampling frequency of at least 8kHz, although 16kHz and above is recommended. Lossy formats (e.g. MP3) are not appropriate because they have a lower bit depth, thus in the compression process information

is removed. In regard to sampling frequency, half the value of the sampling frequency equates to the maximum frequency that can be recorded. The most important linguistic and social information in the speech signal is below 5kHz. Thus, a 16kHz sampling frequency would provide a maximum frequency of 8kHz. However, if the length of duration of a variable is what is being studied, in other words the measurement is a unit of time (e.g. milliseconds) then lower sampling frequencies can be used (Cieri, 2011; De Decker and Nycz, 2013). Acoustic analysis allows for the measurement of differences in speech sounds that may not or cannot be consciously perceivable by humans. In contrast, auditory analysis uses labelling and notation systems to reflect the perceived differences between speech sounds. This perception is linked to linguistic or social meaning that the listener may be consciously aware of (Johnson, 2012). The most prominent example of such a labelling/notation system would be the International Phonetic Alphabet: the use of alphabetic symbols to represent the sounds of speech (see International Phonetic Association, 1999).

From the outset, I decided that I would not be conducting acoustic analyses in this research. This was based on two predictions. First, I predicted that the amount of potential research data would be significantly reduced if its acoustic content had to be suitable for an acoustic analysis. Even in the more professionally produced content (such as the videos analysed in this thesis), much YouTube data includes echo (e.g. because of filming in smaller rooms such as bedrooms), background noise (e.g. traffic outside), foreground noise (e.g. rustling from opening packages and showing the items, jingling of jewellery) as well as music overlaid post filming. Second, I predicted that scholars would have less confidence in the results of an acoustic analysis due to the lack of certainty over the recording equipment. As has already been mentioned, the diversity of data on YouTube is overwhelming including the recording equipment on which recordings are made. While some vloggers appear to use high quality equipment (see figure 3.1 for an example) in many videos the equipment is out of shot. The type of recording equipment used has a fundamental impact on the recording, and thus the measurements that can be obtained from it. Third, I predict similar concerns because of data compression as a result of uploading and downloading the video data. Very little is known

about the data compression performed by YouTube and YouTube has the authority to make alterations at will without informing users. Thus, although some work has found the impact of YouTube's data compression on acoustic analyses to be negligible (De Decker and Nycz, 2011; Whitmarsh, 2017) it can be argued these results were only valid at the time the research was performed.

Finally, although the decision to take either an acoustic or auditory analysis approach dramatically shapes the research methods, I predict little relevant insight can be gained through exploring these approaches in YouTube data. The researcher actually has little say in whether acoustic analysis is an option to consider because technological factors dictate whether a recording would be suitable. As is outlined in the section above, there are other elements of a research method whose exploration is more likely to prompt reflection on the methodological commitments on sociolinguistics and contribute to sociolinguistic theory and knowledge.

3.7 Summary

To summarise, this chapter provides overviews of four aspects of sociolinguistic research methods and considers why their typical practices do not easily transfer from studies of offline speech and language or language-online to studies of speech-online. These aspects are i) Formulating Research Questions, ii) Research Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis. As a result, the four research questions that this thesis will address have been defined. The next chapter will define the case study that will be the vehicle through which these four research method aspects will be explored using YouTube data.

Figure 3.1 Screengrab of PewDiePie clearly showing a microphone and pop shield held by an arm stand (from (PewDiePie, 2018)).

Chapter 4.

Case study Subject and Independent Variable

"The ability to understand how context, audience, and identity intersect is one of the central challenges people face in learning how to [study] social media"
(boyd, 2014:30)

This chapter describes the collection and analysis of the case study independent variable, with the first dependent variable and set of case study results reported in the next chapter. The learning and experience gained from performing the activities detailed in this chapter will contribute towards answering two of the **thesis research questions**, question 1 "What strategies or approaches could a researcher use to i) find sociolinguistically interesting YouTube data and ii) formulate appropriate research questions for that data?" and question 2 "What are the ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?".

As outlined in section 2.6, this case study will examine the one-to-many and reciprocal many-to-one relationship between a YouTuber and their viewers from a sociolinguistic perspective. First, the case study's subject, and thus data source, will be described (section 4.1) and an account of the process by which the case study data was selected and collected will be reported (section 4.2). Next, the

focus moves to the case study's independent variable – the content of YouTube comments. Here, the analysis process and the results are reported (sections 4.3 and 4.4 respectively). The limitations in regard to the case study's design are then reported (section 4.5) so they can be kept in mind through the rest of the thesis. Finally, learning and insights gained by performing the research activities that are reported in this chapter are discussed in relation to two of the thesis research questions: Thesis question 1, Formulating Research Questions, and Thesis question 2, Research Ethics (section 4.6). Finally, the chapter will be summarised (section 4.7).

4.1 Case Study Data: Zoella

For this case study haul videos made by Zoella (Sugg, no date a) were chosen. A haul video is one where the YouTuber has been shopping and presents their purchases with some description, in a show and tell style format (Jeffries, 2011). Zoella, real name Zoe Sugg, is from Wiltshire, UK. She has been posting videos since December 2009, at the age of 19, to the present day, and prior to this she blogged on her site which is still available (Sugg, no date b). Zoella signed up to a management company in December 2012 (see postal address in the information box on (Sugg, 2012)) and became the director of two companies in 2013 and another in 2014 (Companies House, 2013a, 2013b, 2014, respectively), all strong indicators of Zoella's transition from video making being a hobby to her profession and main source of income. Since then, Zoella has become one of the most followed YouTubers from the UK, with 11.5 million subscribers at the time of writing (December 2019) and in March 2016 it was estimated that she earnt £50,000 per month from YouTube views, sponsorship, and her own product ranges (Oppenheim, 2016).

Zoe's closest and dearest also make their income from social media content. Younger brother Joe Sugg's YouTube channel has 8 million subscribers (Sugg, 2020), Alfie Deyes (her boyfriend since October 2012) has 5 million subscribers (Deyes, 2020), and her close friends all have large YouTube followings: Tanya

Burr (Burr, 2020) 3.5 million, Jim Chapman (Chapman, 2020) 2.2 million, and Niomi Smart (Smart, 2020) 1.6 million, and her once best friend Louise Pentland (Pentland, 2020) 2.4 million. This collective was named "The British Crew" by American YouTuber Tyler Oakley (2014) and this term has circulated among fan content (Bueno, 2019) since. Further, as this collective was once all managed by the same company called "Gleam Futures" (2019), they were also referred to using the terms "The Gleam Team" / "Team Gleam" (Dryden, 2015; Kelly, 2017), or "Gleamers" (Collingridge, 2018).

A detailed account for the decision-making process that led to Zoella being selected and the rationale for selecting her haul videos is given in section 4.9 "Reflecting on Thesis Research Questions".

4.2 Method

This section describes the data collection and data analysis methods used.

4.2.1 Data Collection

First, the "Zoella" YouTube channel was searched using the search term "haul". All the videos where this search term was included in the title were noted (68 videos). To control for factors that may affect speech, a set of exclusion criteria for the video data was established and implemented. The criteria and the number of videos excluded (stated as "[#]") were: i) collaborative videos where the haul was being conducted with a friend [2], ii) videos where an eavesdropper is evident [4], and iii) child directed speech [1]. All these videos were excluded because the "audience" included other members (the collaborator, the eavesdropper, the child) rather than just the YouTube audience. After excluding these videos, 61 videos remained.

Next, I wrote a python script that collected all the publicly available comments and comment replies (which, for ease, will now be jointly refer to as

"comments") and their metadata for 60 of the 61 videos (because the intent is to examine the effect comments have on a YouTuber's speech, it is not necessary to analyse the comments of the final video in the dataset). The comment data and meta data was stored in chronological order in a database. This resulted in a total of 195,002 comments being collected. The collected comment data for each video was then filtered based on when they were posted. All comments posted between a video's publication (as was described by YouTube) and 1 second prior to the publication of the subsequent video on the list were filtered. This collected the comments that were posted in the interim of each video's publication, resulting in removing 38,894 comments, leaving 156,108 comments. At this point a video was removed from the corpus because no comments were posted during the time period of interest. Furthermore, one video was identified where to enter a competition the viewer had to leave a comment. This video and its comment data (34,125 comments) was also removed. This resulted in a final list of 58 videos that were posted over a period of 5 years, 9 months and 15 days, from 25/02/2011 to 10/12/2016 (see appendix 1 for details). More details in regard to the amount of video data will be given in the next chapter where speech analysis will begin to be discussed.

This resulted in a corpus of 121,983 comments posted on videos 1 to 57. The comment corpus was then examined and any of the following were removed (number of comments removed stated as "[#]"):

i)   no comment available (a blank comment box) [131],

ii)   the comment was posted by the "Zoella" user account or the "MoreZoella" user account [595],

iii)   the comment was posted by user accounts associated with Zoella's vlogger friends and family (e.g. Sprinkle of Glitter, Pointless Vlogs, Thatcher Joe, Tanya Burr, Jim Chapman, Niomi Smart, Marcus Butler) [21],

iv)   the comment was in a language other than English [5],

v)   the content was spread over several comments (e.g. only one letter was posted in a comment but over several comments this spelt out a word vertically), the comment was a repeat of a one previously posted, or was a grammar or spelling amendment of a previous comment [730],

vi) the comment was posted on YouTube while Zoella was broadcasting on YouNow and so the content of the comment was unrelated to the YouTube video [982]

vii) the comment was posted at Zoella's request, primarily found on video 16 which at the end displayed a message on the screen asking her viewers to make a certain comment to show they had watched until the end [263].

This provided a final corpus of 119,253 comments for analysis (see table 4.1 for a more detailed breakdown).

4.2.2 Comment analysis

A content analysis approach was used to analyse the comment data. The summation of content analysis into one definition is difficult, as is fully outlined in Krippendorff (2004). At its broadest, Krippendorff's definition of content analysis is that it is "a research technique for making replicable and valid inferences from texts […] to the contexts of their use" (2004, p. 18). In comparison, Kimberly Neuendorf's (2017, p. 1) is explicitly more scientific: "Content analysis is the systematic, objective, quantitative analysis of message characteristics".

The purpose of content analysis in this study is to operationalise YouTube comments in a quantitative manner in order to examine if there is a relationship between the comments and the YouTuber's speech. Thus, using the integrative model of content analysis (Neuendorf, 2017) is most appropriate. The integrative model of content analysis uses four key elements (i. source, ii. message, iii. channel and iv. receiver) defined in Shannon-Weaver's communication model (Shannon and Weaver, 1998). The integrative model of content analysis collates message-centric data with other empirical "extramessage" data from one or more of the other three elements that were selected from Shannon-Weaver's (1998) communication model. Further, the content analysis within this study is best described as Predictive Content Analysis (Neuendorf, 2017); the primary goal is to identify if a relationship between comment content and speaker's speech is

present, and thus be able to predict, to some degree, receiver (the YouTuber's) response to messages (the comments).

Table 4.1. Final comment dataset - Number of comments/comment replies analysed by video.

Video	Number of comments		Video	Number of comments
1	121		30	9365
2	265		31	4304
3	205		32	8451
4	394		33	2361
5	493		34	1510
6	342		35	609
7	479		36	554
8	559		37	214
9	371		38	320
10	460		39	500
11	982		40	400
12	1222		41	317
13	939		42	289
14	769		43	182
15	1135		44	248
16	1156		45	204
17	1913		46	250
18	2473		47	170
19	3271		48	415
20	3898		49	486
21	3177		50	325
22	5855		51	350
23	4268		52	123
24	6139		53	179
25	8532		54	102
26	6668		55	273
27	17527		56	71
28	2972		57	114
29	9982		Total	119,253

In order to gain knowledge of the commenters' perceptions of Zoella, the adjectives contained in the comments were assessed. For clarity, comments that used verbs to indicate opinion or emotion (e.g. "Love this video" / "Love you Zoe" / "Love that dress") were excluded. While this data would communicate a commenter's opinion of Zoella, the focus of this investigation is on if and how Zoella responds to comments through her speech behaviour. Hence the data analysis was only concerned with adjectives as these can be linked to speech features via their associated social qualities.

I initially surveyed approximately 10% of the comment dataset. The vast majority of comments included a positive adjective (e.g. "great", "amazing", "cool"). However, it was difficult to imagine how such descriptions were gained from or communicated through the use of speech features. These words were too generic and lacked connotations related to social qualities. However, it was evident that there was a theme of youthful femininity besides generally positive comments. To be more specific, a recurring series of adjectives that would all be appropriate to use when describing an aesthetically pleasing young girl. With gender and age being significant social categories by which speech features vary, it was hypothesised that comments containing such words could be linked to speech.

A quantified content analysis was then performed on the data. Each comment was read to see if it contained a word relating to youthful femininity. The bank of words which were deemed to be referring to youthful femininity developed pragmatically through engaging with the data. The final list of youthful femininity words was:

i)   adorable
ii)   charming
iii)   cute
iv)   girly
v)   lovely
vi)   precious
vii)   sweet

The comment was then assigned to one of three categories according to what article the youthful femininity word was describing. These three categories, that were identified prior to analysing the data, were:

i) 'Zoella' – comments that refer to Zoella explicitly. "Love you Zoe. Too cute!", for example.

ii) 'Media' - comments that refer to the media through which Zoella portrays herself (e.g. the channel, video, video thumbnail). "Your videos are so cute x", for example.

iii) 'Content' – comments that refer to the properties of items within the video. These were predominantly objects (e.g. the purchases being shown, the clothes Zoella was wearing, items situated in the background), "my fave is the cute blue dress" for example. But also included properties of Zoella herself that were physical (e.g. hair, eyes) as well as non-physical (e.g. personality, sense of humour, sense of style). "cute messy bun = style goals" and "she so girly and cute" would be examples, respectively.

Comments could be assigned under multiple categories at once (e.g. Zoella and Content, or Medium and Content) as well as be coded for multiple adjectives within and across the categories (e.g. "the pink dress is so sweet I love it! the sandals are cute too" would be "sweet" and "cute" for Content). However, as will be explained in the results section (4.3.1), in the end this double counting of comments was removed, and the multiple categories of comments were collapsed into one.

4.3 Results

First, a summary of the comments across the three categories (Zoella, Media and Content) and over time will be given. Then evidence of Zoella's own commenting on her own videos will be reported. Finally, whether there is a change over time in who is commenting will be considered. R Studio (2019) was used for all calculations and visualisations.

4.3.1 Comments

From the filtered corpus of 119,253 comments, adjectives of youthful femininity were used 4,988 times. As can be seen in table 4.2, "cute" was used the most and "charming" the least, with 123 comments using more than one youthful femininity word. When comments that used more than one youthful femininity word were just counted once, the corpus contained 4,865 comments, 4.08% of the filtered corpus. In regard to the percentage of comments, by video the mean was 4.4%, the range was 0% to 11.7%, and the standard deviation was 2.82%.

Table 4.2. Number and percentage of youthful femininity adjectives.

	Adorable	Charming	Cute	Girly	Lovely	Precious	Sweet	Total
Zoella	374	2	1012	3	257	13	133	1794
Media	31	0	209	2	202	0	7	451
Content	273	3	1845	11	587	3	21	2743
Total	**678**	**5**	**3066**	**16**	**1046**	**16**	**161**	**<u>4,988</u>**

Figure 4.1 is a line graph of the cuteness-comment data. The percentage of cuteness comments fluctuates greatly, although there appears to be a downward trend overall as indicated by the line of best fit in figure 4.2. A more specific description would be that there is a steady, steep decrease in the first three years (from 2011 until 2014) which then plateaus or stays level for about 2 years (from 2014 until 2016) and then slightly increases again (from 2016 until 2017). Two reasons are hypothesised for why the youthful femininity comments decrease over time. The first is age. The dataset spans 6 years, from when Zoella was to 20 when she was 26. It may be that it become progressively less appropriate to describe her as "cute" the older she gets. The second is that this decrease could be a consequence of the number of subscribers, and thus viewers and, most importantly, comments, increasing over time. While it is not possible to know how many subscribers she had when she uploaded each of the videos in the data set we do know that she had amassed 11.5 million subscribers by December 2019, and was one of the most followed YouTubers from the UK. Thus, there is evidence to assume that the number of comments she received increased exponentially along with the number of subscribers.
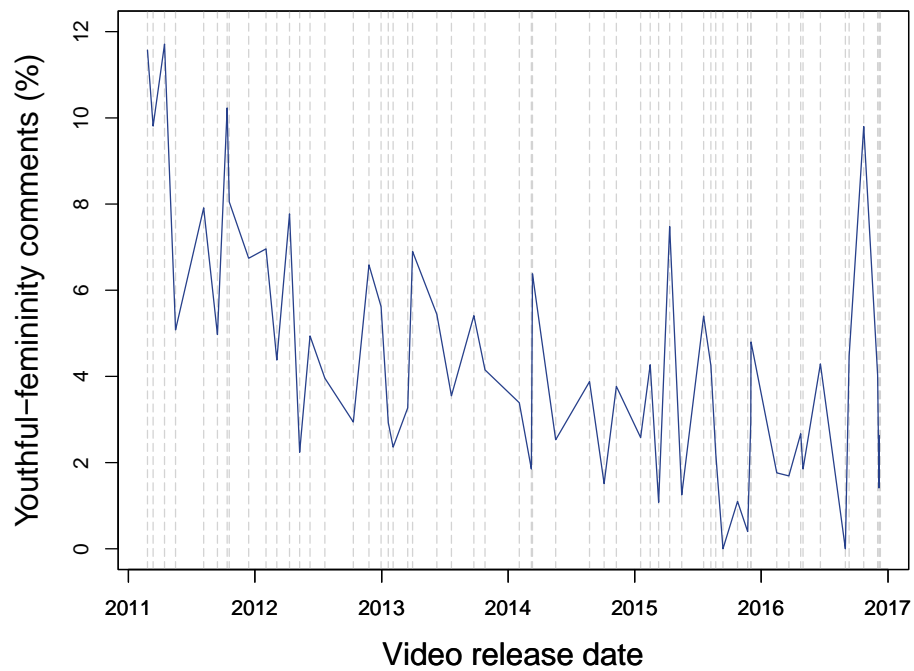
Figure 4.1. Percentage of youthful femininity comments per video
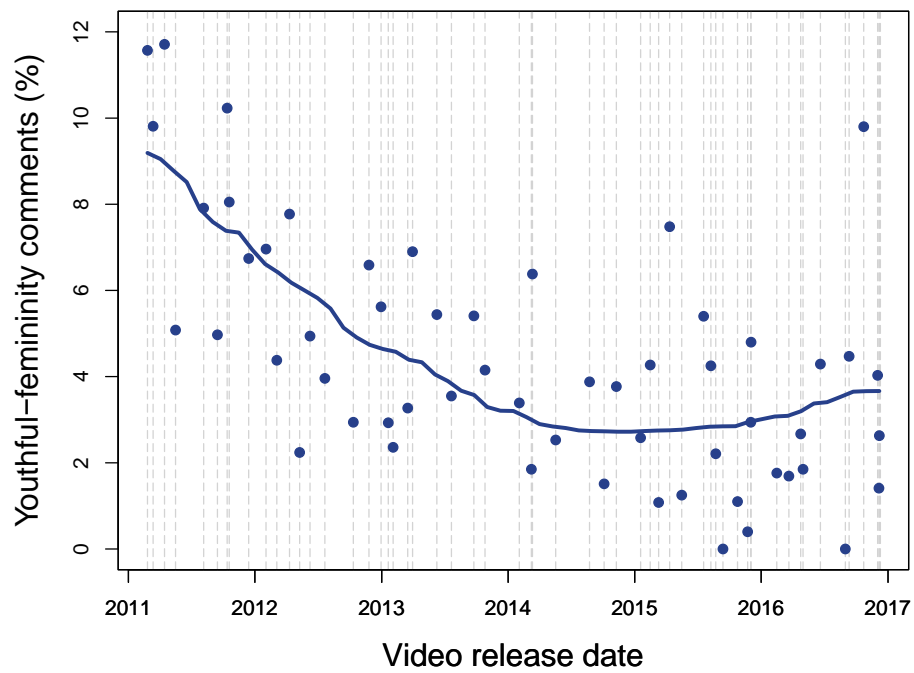
Figure 4.2. Youthful femininity comments line of best fit (fitted with 2
polynomials)

4.3.2 Zoella's comment engagement

Fundamental to the interaction that is under study in this research is that Zoella looks at the comments on these videos. Without evidence that Zoella reads the comments, any apparent relationship between the comment content and her speech could be attributed to chance. Unfortunately, we cannot be certain that she encountered all the feedback on all the videos in this dataset. However, there is clear evidence that she reads some of the comments because she replies back. This was investigated further by counting the number of comments Zoella leaves on each video in the dataset (see figure 4.3). As the line of best fit in figure 4.4 illustrates, the number of comments Zoella leaves declines over time. Between 2011 and the mid of 2013 Zoella consistently replies to the comments left on her videos. Comments are left on all videos, apart from videos 24 and 25, with 63 comments being the most left on one video (video 14). During this time period, the mean number of comments left on a video is 23. However, the number of comments she leaves substantially drops from the start of 2014. Zoella only comments on 4 videos in 2014 videos (27, 29, 31, and 33). Further, from the beginning of 2015 onward Zoella does not leave comments at all. This period in the timeline is when Zoella is a well-established, professional content creator. Thus, it is highly likely that her management team monitor the comments and distil the most important information to her. This less direct and summarised feedback may have a different impact or influence compared to her own direct engagement.

4.3.3 Commenters

There are 84,572 different users in the comment data. To gain some insight into the commenters over time, the visualisation in figure 4.5 was produced. Each vertical line corresponds to one of the videos in the dataset, and the y axis equates to a cumulative list of individual users in chronological order of when they first comment on a video.
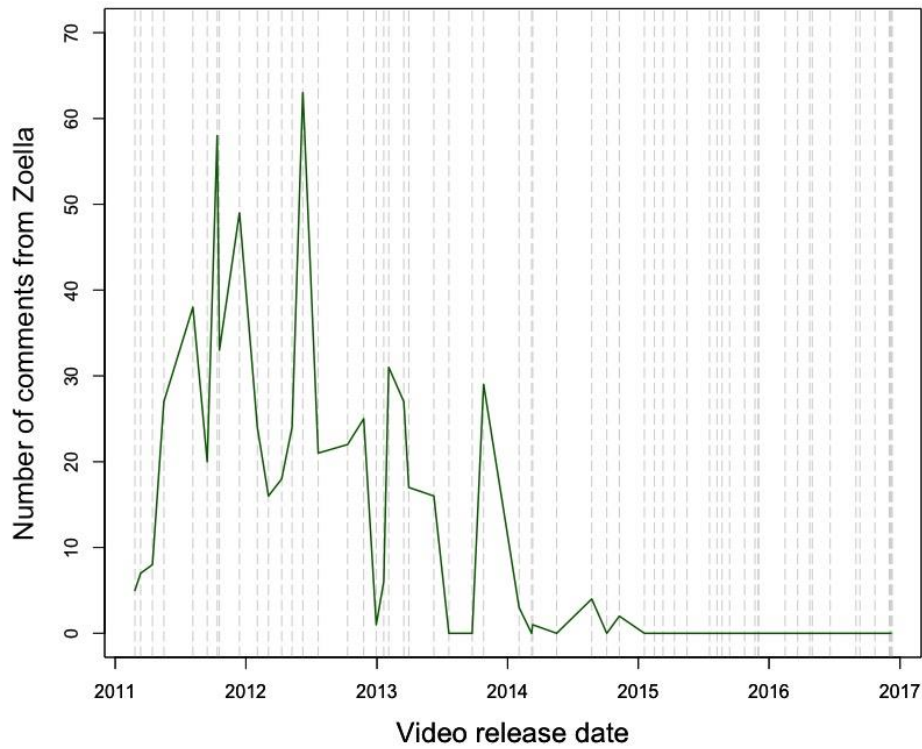
Figure 4.3. Number of comments Zoella posts to each video
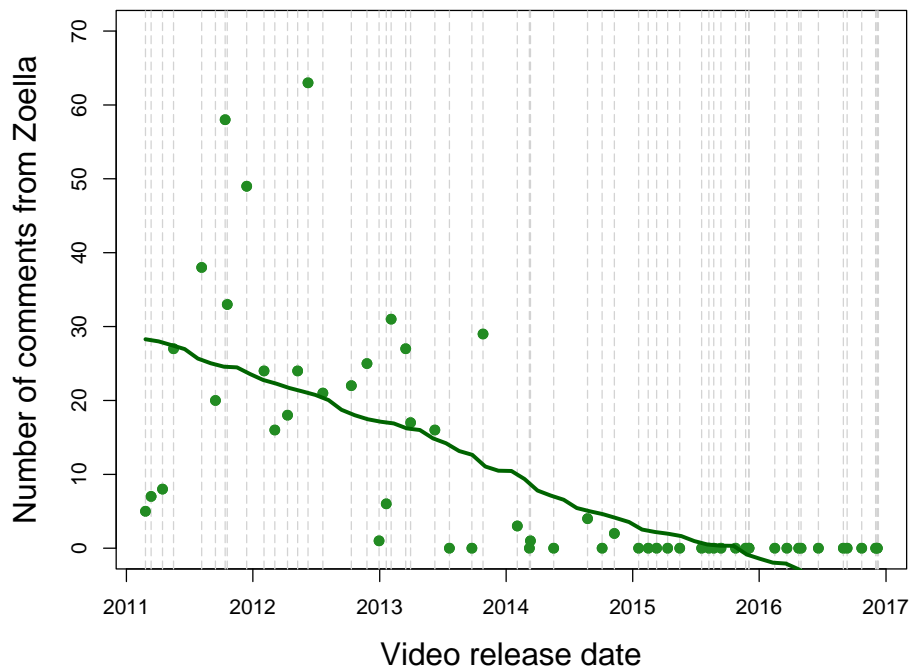
Figure 4.4. Zoella's comments line of best fit (line of best fit travels through 0 regardless of the number of polynomials used)

The first user to comment on video 1 is shown at the bottom left hand corner, and the last user to comment on video 57 (and has never commented on any other video in the dataset) is at the top right-hand corner. Thus, if a user left a comment on every video in this dataset there would be a solid black line from left to right at the very bottom of the graph. Whereas if a user commented on every video from video 29 (halfway through the dataset) onwards there would be a solid black line that starts halfway along both the y and x axis, and continues horizontally to the end of the x axis. However, as the vertical lines becoming increasingly spaced out indicates, it is evident that there is a great turnover in Zoella's commenters. In other words, most of her commenters are not loyal, and there are very few from her early videos in 2011 that comment on her latest videos. Further, there is a dramatic accrual of commenters during the middle of the timelines, between 2013 and 2014, as evidenced in the steep incline in the middle of the visualisation. This time period is when Zoella transitions from making YouTube videos as a hobby to it becoming her profession. This then stabilises for the latter part of the timeline.
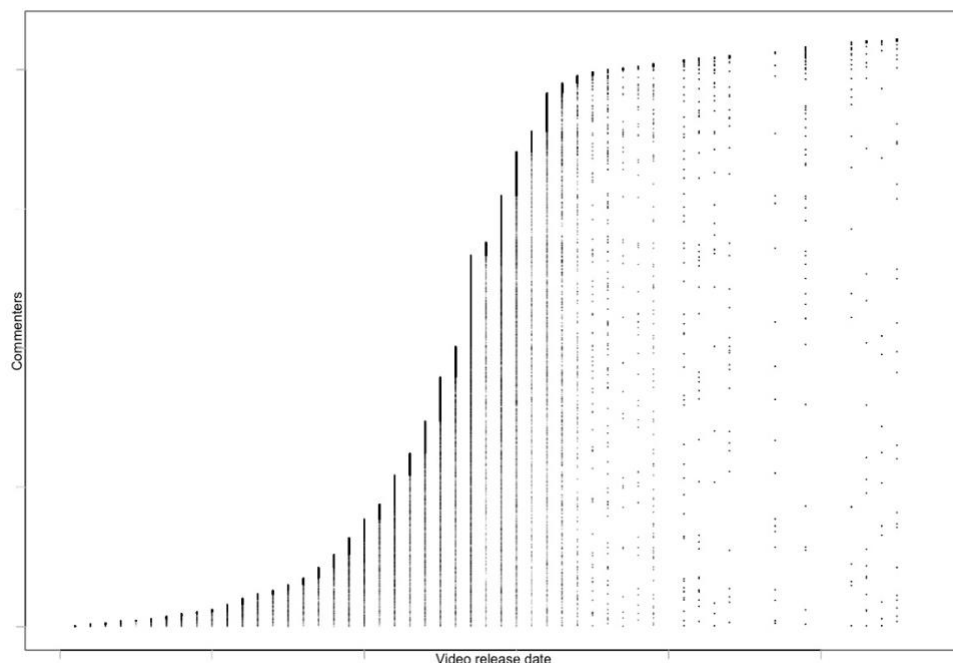
Figure 4.5 Commenters and their commenting behaviour over time

4.4 Limitations

In this case study the majority of limitations are in relation to the comment data. Some of these limitations are unique to YouTube as an interaction context, but many of the overarching issues that these limitations bring to the fore are also present in studies of real-world interactions.

4.4.1 Visual material

First, in this analysis visual material is not taken into consideration. At first this may seem absurd considering the data source is online video, and that facial expression, body posture, and gesture provide vital information to a listener. However, very rarely in variationist sociolinguistic studies is visual information integrated into data analysis, including Bell's studies that prompted his reworking of Audience Design (2001). It could be argued that this should be looked at more closely considering the vast majority of the comments about cuteness were referring to physical properties of the items shown or parts of Zoella's presentation of self (e.g. her hair), rather than directed at her as a person, and a minority of comments were directed to her speech. In response, previous variationist studies have considered physical and visual properties such as dress and hairstyle, to a minimal degree. Most importantly, these displays of identity have been inherently linked to speech, such as in Mendoza-Denton's ethnography of Latina girl gangs in North California (2008).

4.4.2 Influence of comments

It should be noted that the number or frequency of comments is not necessarily reflective of their power to influence behaviour. It may be that comment influence is logarithmic, that there is a minimum threshold before Zoella's speech is influenced or a maximum where the same degree of influence is had over Zoella's speech regardless of how many comments are left. Also, these parameters could change over time, of course. In other words, this abstraction

from the full content of each comment may obscure what influence each individual comment has and why, with the degree or strength of influence potentially varying across the comments. In the same vein, if the comments were to have an influence on Zoella's behaviour, that influence may not be consistent throughout the dataset.

4.4.3 Other sources of feedback

Further to this, a third issue is what feedback she is receiving other than the direct written comments that have been analysed. This can be broken down into two sources of feedback. First, is the feedback Zoella receives on her other videos. I have only analysed comments on specific haul videos, but Zoella posts a lot of videos in between these. An expansion of this work could therefore be to analyse the comments that are posted on videos in the interim of those that I have studied. However, the variety of activities and communicative partners in other videos might have an impact on her speech or behaviour in general. Thus, I took the strategic decision to keep the context as consistent as possible.

Second, it is possible that similar messages of 'cuteness' may be being received from feedback sources outside of YouTube, for example, when meeting fans face-to-face, and when the mass media (e.g. newspaper and magazine articles) report on Zoella's activities. It is reasonable to assume that the feedback that she would receive from fans face-to-face would reflect that which she had received through the comments. This is strikingly different in the mass media, however, as Zoella has been heavily criticised for her behaviour previously (see McGoogan, 2017 for a summary). According to Frobenius (2014, p. 61),

> "comments must be understood as one fraction of audience reactions, which, however, are geared to [videos] in that they appear in the same public, virtual space and hence are specifically designed to be recognized as reactions to them".

4.4.4 Comment curation

Finally, when interpreting the data in this case study it should be kept in mind that Zoella has some curatorial control over the comments that are seen. YouTubers are able to remove comments, and viewers are sometimes able to send private messages in addition to public comments. From the comment data it is difficult to determine the degree to which Zoella and her team curate her comments. While a minority of the comment data retrieved included blank comments, it is not stated in YouTube's API guide whether these blank comments are an acknowledgement of comments that have been removed/deleted or not. Interestingly, in the dataset retrieved there is some indication that negative comments from viewers are not removed. And, again, what always needs to be factored in when considering this context is that the YouTube interface, its functions, and Terms and Conditions are not stable entities. It is highly likely that comment posting, vetting, and removal behaviours have evolved over the six-year period under study. In fact, I would argue that the overwhelming positivity found in the comments is an advantage for operationalising this data. If defining the relative weighting or degree of effect of each positive comment is not difficult enough, integrating negative comment data complicates the analysis further. How many positive comments outweigh a negative? Therefore, while it needs to be acknowledged that Zoella may possibly be receiving negative feedback that we do not have access to, it is not possible to weave the influence of such data into the study design at this stage.

4.5 Reflecting on Thesis Research Questions

In this section I will outline the difficulties faced in designing the research method for this case study in relation to the thesis research questions, then recount the decisions made and explain the rationales for them. The different elements of the research method needed to be carefully negotiated. Of course, such decisions also need to be made in typical sociolinguistic studies, but their complexity is amplified in this online interaction context.

4.5.1 Thesis Research Question 1

Once the case study research question was formulated, it was relatively easy to find Zoella's videos and the particular videos to be the focus of the study as a result of establishing which met the desired data qualities. However, in all these regards it cannot be overstated how much my personal prior knowledge of YouTube played a role.

First, referring to 'finding' Zoella's videos is somewhat misleading. Rather my prior knowledge and experience of YouTube led me to assess Zoella's videos and YouTube career from a research perspective to see if the data may be suitable. While I cannot be certain when I first engaged with YouTube it was certainly around 2009 when I started my undergraduate degree. My typical engagement around that time was watching hair and make-up tutorial videos and this interest expanded to include hair and skin care routines, fashion and styling videos and home décor. In the past I have regularly engaged with Zoella's channel for this type of content along with many other channels. Thus, in the process of identifying Zoella as being potentially appropriate to study I also considered many other YouTubers that I have watched over the past decade. Most of these other YouTubers would have also been suitable to study, but the key factor that led to Zoella being chosen was the likelihood of there being the most comment data, because she was the most prolific amongst the candidates.

Second, my prior knowledge and experience significantly guided data selection in regard to the video content also. Hauls have been a prominent class of video since I first engaged with YouTube and, in my experience, their premise, structure and style has minimally changed over the past decade. Therefore, prior to watching Zoella's hauls as a part of this research I predicted their data qualities would make them suitable for a research study and be consistent across the data set. Typically, with a haul the activity is loosely structured but not scripted which would provide appropriate data as the speech would be produced with a reasonable degree of control but not be totally spontaneous. Also, hauls are usually a monologue with the speaker on their own and this is required for

this case study for reasons that should be plain considering the literature review in section 2.6. Further, the content of the videos dictates the linguistic content, which then has a notable influence on the speech produced. I predicted that the sentence structures and vocabulary are consistent across the videos as a result of the same activity being performed and the same sorts of items or products being shown and discussed, and this has been confirmed through analysing the data.

Finally, YouTubers usually have a typical physical set up for 'sit down' videos, probably for ease (not having to regularly set up and take down filming equipment every time they want to do a video). Thus, the physical location, down to where and how the YouTuber is sitting, can be the same across multiple videos. To summarise, haul videos were chosen because they would provide relatively consistent data in regard to the activity being performed, the structure and language associated with that activity, and the physical context where it is performed. The video content being consistent was important because this meant certain factors that may influence behaviour would be held consistent also. It was essential that the content of the videos was as similar as possible because variation within the dataset could lead to any relationship between viewer comments and YouTuber behaviour being overshadowed.

In chapter 3, YouTube was identified to be a near infinite and a highly heterogenous source. Therefore, the two main strategies for discovering raw research data were envisioned to be redundant: it is not possible to search YouTube using sociolinguistically aligned criteria for data to answer a predetermined research question, nor is it likely that data could be happened upon by chance and inspiration for a research question serendipitously struck. In the end, the approach used herein was a combination of these. Essentially, the approach I took was using my own experience and knowledge of YouTube to significantly reduce the pool of potential data, namely to specific YouTubers and specific types of videos. Then, I mapped out the data qualities that the different types of videos typically possessed, and the data qualities I needed the research data to contain bearing in mind the research question. This clearly indicated what video data and from whom would be appropriate to study. This decision-

making process was not planned in advanced or recorded in a structured way but is evident as I reflect upon my own practice.

In a similar vein, I did not plan a structured process for formulating the case study research question. As I was conducting the literature review it became evident that a study that examined the relationship between the viewers and a YouTuber would be of interest to both sociolinguistics and CMC. I had prior knowledge of the three main sociolinguistic theories of style in advance of beginning this literature review, and as one of the first computer-mediated communication related texts I read was danah boyd's "It's Complicated: The Social Lives of Networked Teens" (2014) it was not long before I found the theory of context collapse. Simultaneously, I was watching a lot of YouTube, reviewing content I was already familiar with, rediscovering content I had forgotten about, and stumbling upon new content. Unlike my previous viewing, I continually asked myself "would this be interesting data?", and "could this be good research data?". Thus, my thoughts moved between evaluating the literature and evaluating YouTube data, continually iteratively refining my definition of each to find where they met in the middle.

To summarise, an abundance of insights have been gained to inform the answer to thesis research question 1: What strategies or approaches could a researcher use to i) find sociolinguistically interesting YouTube data and ii) formulate appropriate research questions for that data?

4.5.2 Thesis Research Question 2

The decision to study this data was not solely a result of considering data qualities, the ethical implications also contributed. This case study considers two key types of data: comments and Zoella's videos. However, a discussion of the video data is reserved for the next chapter where its collection and analysis are first reported. The reflections here are limited to those that relate to studying the comment data.

From reflecting on the discussions in the literature (section 3.2), studying YouTube comment data is relatively uncomplex in regard to ethics. This is for several reasons. The first aspect to consider is the data's collection. The YouTube API is designed to allow the efficient collection of comment data and much of its metadata. In addition to this being practically straightforward it clearly indicates permission and that the collection of this data is in line with the platform's terms of service.

The second aspect was informed consent. Ethical approval from my Institution's Research Board was straightforward in this regard. The perspective took was that this was public data and so informed consent was not necessary. However, within my own decision making and reviewing my research practices, this issue was considered far more in depth. The content of the comment data is important as this can contribute to defining whether the data producer perceived the communicative context (that is leaving a YouTube comment on a particular video) as public or private, when applying Landert and Jucker's (2011) model of mass media and Helen Nissenbaum's (2004) theory of contextual integrity (see section 3.2.2). It was envisioned that the content of the comments would not be emotive, or controversial in nature because of the relatively mundane video content that is being responded to. Indeed, now that the data has been analysed it is evident that this prediction aligns with the overwhelming majority of the data. Thus, there is very little indication in the data that the commenters perceived the comments section on YouTube to be anything other than a public communicative context. Further, because the focus of the study is on what the commenters think of Zoella, if any comments did arise that contained personal information or were emotive or controversial, they would not be of interest to the study and so would be discarded. This means that any comments that may indicate the commenter perceived the communicative context to be something other than public would not have been analysed anyway. Finally, it is highly unlikely that it would be possible to gain informed consent from all the 84,572 users whose comment data is in the filtered corpus.

The third aspect was the tension between anonymity and credit. In considering this aspect I asked myself: what would be the value of anonymising or crediting

a commenter to the study and to the commenter themself? First, I concluded that crediting a commenter would give little value to the study. This impression was predominantly influenced by the knowledge that there is a great turnover in the usernames across the 6 years of data. In other words, while a handful of commenters continue to comment years after their first comment, most the commenters of the early days are not those who comment more recently. Thus, it is questionable how fruitful it would be to the study to highlight who is commenting when the commenter only appears once or twice in the dataset. I concluded that crediting the commenter would be of little benefit to them also. This comment analysis is the abstraction of a single word, sometimes two, from each comment. And now that the comment data has been fully surveyed it is evident that there is remarkable consistency in the content of the comments across the dataset. No particular comments stand out as being more informative, novel or controversial. Therefore, acknowledging one commenter over another would be inconsequential.

In addition to indicating that crediting comments would be of little value to the commenter or the study, discovering these attributes of the comment data suggests that it is well suited to the application of data "disguise" strategies (as detailed by Amy Bruckman (2002)) as an anonymisation strategy. Although it is strongly indicated that they viewed the communicative action of leaving comments as a public one, it is unlikely that commenters would have envisioned their comments being used in academic research. Thus, anonymising the data would prevent the commenters being identified and potentially perceiving that their privacy had been violated. It should, however, be noted that it would be possible to identify commenters regardless of the anonymisation strategies used if someone took the time to reengineer the data collection process detailed herein. Finally, anonymising the comment data would not be directly beneficial to the study per se, but in doing so makes a statement in regard to what I as a researcher deem to be ethical practice.

To conclude, while the ethical decision-making process in regard to this YouTube comment data is relatively uncontroversial considering prior work and current debates, it is highly relevant to thesis research question 2: "What are the

ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?". Thus, the reflections herein make a contribution towards answering this question.

4.6 Summary

To summarise, in this chapter the data collection and analysis of the main independent variable was reported. Through performing this work, two of the thesis research questions have begun to be answered. In the next chapter, the first of the case study's dependent variables will be investigated.

Chapter 5.
Uptalk

The previous chapter identified the independent variable (comments) and reported the collection and analysis of the relevant data. This chapter will report on the collection and analysis of the first of two dependent (speech) variables that are investigated in this case study. As was unpacked in section 3.3, place is a vital factor in selecting offline sociolinguistic research variables because speech features are tethered to a place to some degree. It is also evident that place is a vital factor in identifying online linguistic research variables too, with place being defined through geo-tagging, the data's topic of discussion or the social backgrounds of the users producing that data. However, it should be acknowledged that without these distinct markers many types of online data, including YouTube data, are devoid of place or place is ambiguous at least.

This chapter focuses on the first of the two novel approaches for guiding the selection of speech variables in online public video data that are explored in this thesis. This initial approach imagines the audience as "the anglosphere" – countries that are English-speaking and share common roots in British culture and history (Warren, 2016) - in an attempt to embrace the ambiguity of place. Thus, the concept of Global Linguistic Speech features (section 5.1.1) is harnessed as a strategy to identify an appropriate speech variable. Next, the speech variable chosen - uptalk - is introduced and a summary literature review will evidence why it has been selected in relation to this strategy (section 5.2). Then the **case study research question** and hypotheses will be defined based on

the findings from the comment analysis and the intention to investigate uptalk (section 5.3). Section 5.4 will detail the data collection and uptalk analysis, 5.5 will report on the statistical analysis, and 5.6 will summarise the findings.

The statistical analysis provides no evidence that there was a relationship between the two variables, indicating that the initial strategy of embracing the ambiguity of place online by defining the imagined audience as geographically broad was not successful. This will be reflected upon in relation to **thesis research question 3** ("What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?") in section 5.7.2, and how I addressed the ethical complexities surrounding collecting and analysing online public video (addressing **thesis research question 2**, "What are the ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?") will be reported in section 5.7.1. Section 5.8 will summarise the chapter.

5.1 The Anglosphere

One way in which Zoella may imagine her audience is as "the anglosphere" – countries that are English-speaking and share common roots in British culture and history such as the USA, Canada, Australia and New Zealand along with the UK (Warren, 2016). This approach frames geography broadly, literally. Thus, this imagining is in line with the essence of the phenomenon of context collapse (boyd, 2007; Wesch, 2009) – that her audience has the potential to be large and geographically dispersed, which could be imagined as multinational. Further, from a researcher's perspective, this also allows for the case study to be conducted with minimal data, and thus resource, because any further refinement of how Zoella may imagine her audience would require additional data collection, analysis and interpretation (as will be explored in the next chapter).

This, however, could be viewed as a conundrum. If speech is tethered to geography, how can Zoella attune her speech towards the multiple,

geographically dispersed listeners (that she imagines), simultaneously? Here, the concept of Global Linguistic Speech features can be harnessed to identify a speech variable that, theoretically, bridges between multiple, geographically dispersed audiences.

5.1.1 Global Linguistic Speech features

Global linguistic speech features are "linguistic innovations arising contemporaneously in highly disparate places." (Sayers, 2014b). In other words, they are a similar change in speech (or language) that occurs in more than one location at a similar time and these locations are remote to each other to some degree. To be clear, this is a different phenomenon to English and other languages spreading across the globe (and this is a research field in itself, (see Coupland, 2013)). Further, the term "global" is not to suggest that the speech feature is used everywhere "but just to highlight the sheer distances involved" (Sayers, 2014a, p. 186).

The most famous global linguistic speech feature is one of language, not speech: quotative 'be like', for example "I was like I'll do it tomorrow" or "He was like it doesn't really matter". This has emerged in the UK, the USA, Canada, and New Zealand to name a few places (see Tagliamonte and Hudson, 1999; Baird, 2001; Tagliamonte and D'Arcy, 2004; Barbieri, 2007, 2009; Blyth et al., 2018). In the domain of speech, TH-fronting to [f]/[v] (so "three" is pronounced "free" and "feather" is pronounced "feaver") and t-glottaling (so "button" is pronounced something like "bu'n" (this will be revisited in chapter 7)) are examples in the UK (Milroy, 2007). While Sayer (2014a, p. 187) states that these are not global speech features but rather "have arisen with similar speed across [a] whole countr[y]", this brings into question whether a speech feature must occur across multiple nations or landmasses to be counted as global and, if not, what's the minimum distance between two locations to categories them as "highly disparate" or distant.

Revisiting Sayer's (Sayers, 2014b) definition more critically, it is evident that attention is drawn, but also limited to, what the speech feature is ("linguistic innovations"), when ("contemporaneously") and where ("highly disparate places") it is used. However, who uses it and how it is used needs to be considered also. Returning to the most famous global linguistic speech feature, quotative 'be like' evidences that while an initial surveying may suggest that the same linguistic innovation is arising in multiple places, who is using it, and how it is used can differ. Isabelle Buchstaller (2008) found that quotative 'be like' was mainly used by young and working-class US speakers (gender did not reach statistical significance) and by young British speakers (gender and socio-economic class did not reach statistical significance). She also found that the construction 'feel like' dominated in the US (e.g 'I feel like "yay I've done it!" ') but was rare in the UK. In contrast, 'say like' was common in the UK (e.g. 'I say like "is that all?" ') but was rare in the US. Further, 'like' collocating with another verb, illustrated in both these examples through 'feel + like' and 'say + like', occurred much more frequently in the US than the UK data. Thus, "speakers participate in global trends, but do so in an idiosyncratic and locally specific manner" (Buchstaller, 2008, p. 26). Thus, regardless of which global speech feature is chosen for this study, it is unlikely to be "the same" in who uses it and how it is used across all locations. However, as is expanded on at the end of this section, the priority when choosing a global speech feature as the dependent variable for this study is that Zoella imagines it to be the same across her imagined audience.

Further, whether such 'global' changes are just superficial is highly debated. There is a theory that has begun to take hold (but still needs empirical evidence): that not all speech changes are equal. Rather, it is posited that there are some variables that are vulnerable to becoming, and some that are robust against becoming, global linguistic speech features. Lesley Milroy (2007) has suggested the terminology "off-the-shelf" and "under-the-counter" to describe these changes, respectively, that are thought to be easily picked up or requiring prolonged social contact. Note, I use the term "robust" and not "immune" as under-the-counter changes are reported in "discontinuous geographic settings" (Buchstaller and D'Arcy, 2009, p. 291).

Global linguistic speech features are somewhat of a conundrum in sociolinguistics because, as outlined in section 1.4, it is generally believed that regular face-to-face contact is a necessary pre-requisite for linguistic change. Hence, during the era when these global speech features began to be identified (in the 1990s) and began to be discussed more critically (2000's) there was some attribution to media, for the lack of another explanation. However, what role media may play has been more critically considered in recent times (see section 1.4) because, to requote Trudgill (2014, p. 216), "everyone in the British Isles would now have an American accent, or at least there would be progress in that direction." if media was a dominant force behind the spread of global linguistic speech features.

While these topics should be acknowledged, the work in this thesis does not attempt to contribute to the discussion of the mechanisms by which a speech feature may become global or what qualities of speech features make them vulnerable to globalisation. Rather, in this work I harness the concept that is the global speech feature as a strategy for identifying the variable for this case study. The prediction is that Zoella will use a global speech feature because she perceives it to be more likely to encompass a geographically dispersed imagined audience. Based upon a literature review, and an initial survey of her speech, the variable uptalk was chosen for this study.

5.2 Uptalk

Based on the results of the comment analysis and using the concept of global speech features as a variable selection strategy, uptalk was identified as a suitable dependent variable for this case study. Here, the what, where, who and how of uptalk will be outlined, evidencing the rationale for its selection.

To understand uptalk, one must first understand intonation. A simplistic definition of intonation would be that it is the melody of speech; the rises and falls in pitch. Intonation is used to communicate "*sentence-level* pragmatic

meanings in a *linguistically structured way*" (Ladd, 2008, p. 6; original emphasis). However, intonation is not the only feature used to do this, so are rhythm, volume and stress. These features in combination are referred to as prosody. An intonational phrase (IP) is a way of dividing up longer stretches of speech into smaller, linguistically meaningful chunks. The pattern of rises and falls of pitch that occur on the IP are known as the intonation contour. While these are called intonation phrases, the other prosodic features of rhythm, volume and stress also assist in defining where IPs begin and end. IPs can be made up of a series of intermediate phrases; ones which come prior to the end of the IP and so do not carry the signals to indicate that the unit of speech is ending in regard to linguistic meaning.

Returning to *what* uptalk is, a working definition by Warren (2016) is that it is a rising intonation at the end of declarative (statement) sentences. Prior to Warren's (2016) synthesis of the literature a number of terms were used that were effectively describing the same phenomenon. These included; upspeak (e.g. Bradford, 1996, 1997) Australian question intonation (eg. Bryant, 1980; Guy *et al.*, 1986), HRT which has stood for high-rise or high-rising tone (eg. McGregor, 1979; Kiesling, 2005), high-rising tune (eg. McGregor, 2005), high-rise or high-rising terminal (eg. Britain, 1992; Stanton, 2006), and the more specific high-rise or high-rising terminal declarative (HRTD) (e.g. Allan, 1986; Meyerhoff, 1992). As summarised by Di Giuacchino and Crook Jessop (2010, p. 2), 'the use of variable descriptions of uptalk by researchers makes it difficult to assign a concrete and stable definition'.

The first reason for selecting uptalk is *where* it has been observed: across the anglosphere in the USA, Canada, Australia and New Zealand, as well as the UK (Warren, 2016) and thus reflects one hypothesised way in which Zoella imagines her audience. Uptalk has been observed in speech varieties from the South of Britain, where Zoella is from. Therefore, it is highly likely that uptalk will be a speech resource that is available to her and thus should appear in the data.

In regard to *who* uses uptalk, it has a strong association with young women. The speaker being focused on in this study is a woman who is under 30. Also, it is

reported that the majority of her audience are young, mostly teenage, girls (e.g. Godwin, 2015). Therefore, it is highly likely that her viewers use uptalk. While speech is tethered to geography, the key factor that's missing online and thus makes selecting a variable more complex, it is also associated with the social demographics of communities that inhabit geographical locations. Thus, imagining an audience will include imagining their social demographics (e.g. age, gender). Thus, to assess if Zoella responds to her audience's feedback (comments) via her speech, a variable that she thinks is used by (and thus has meaning to) her imagined audience because of their imagined social demographics, as well as their imagined geography, should be selected.

But most important for this study is the meaning that uptalk has for Zoella and the meaning she believes it has for her imagined audience. To be able to identify an association between the comments and her speech the variable selected must index the social quality 'cuteness' and uptalk is rationalised as the global speech feature that would most likely do this. While no attitudinal study has examined 'cuteness', Uldall (1964) found their participants associated high-rise endings with "pleasantness" and "submissiveness", and Guy and Vonwiller's (1984) participants rated uptalkers as friendlier, more attentive, more expressive and younger. Also, in a newspaper report of the results of Bradford (1996) the title "A cute accent?" was used (Bathurst, 1996). This newspaper report is just one of many examples of uptalk garnering relatively significant attention in the general media, probably because of its negative associations. To give just a few examples, it has been described as "idiotic-sounding" (Beachcomber, 2012), "an irritating verbal tic" (Marsh, 2006), and "infuriating" (Parkin, 2006) although almost all media pieces are the opinions of journalists or readers who have sent in their comments.

One may assume that speakers would avoid using such a heavily stigmatised speech feature. However, what is stigmatised by one social group may be prestigious to another. Here, we can refer to the concepts of overt and covert prestige. According to Labov (1966), a speech feature with overt prestige is one that is widely recognised as being associated with the culturally dominant group in society, and a speech feature with covert prestige is one that is associated with

an exclusive community, often this is one that subverts mainstream cultural norms. Thus, this apparent contradiction between the negative comments found in the media of these nations and the apparent increasing use of uptalk suggests it does hold covert prestige for communities across the anglosphere. Also, considering indexical fields (Eckert, 2008) and Silverstein's (2003) notion of indexical order (section 2.1.3), it is evident that uptalk is a 2nd order speech feature or, in other words, a speech feature with $n + 1^{st}$ indexical value because such discussion in the media evidences that it is socially salient and has social meaning affiliated to it that can be activated in interaction.

Finally, a biologically defined, rather than socially defined, theory can be used to associate uptalk with 'cuteness'. The hypothesis is that intonation patterns are linked to universal notions of meaning as a result of biology; "of the parts of the human vocal system that are used linguistically, intonation responds more closely than any other to states of the organism" (Bolinger, 1978, p. 474). The sound symbolism theory, or biological code, that is relevant here is the frequency code (Ohala, 1983). The frequency code is that there is a negative correlation between pitch and speaker size, that is not body size necessarily but size of the larynx and the vocal folds it contains. Thus, the larger the speaker the lower the pitch. Hence, higher pitched voices are produced by children and thus are associated with youth and innocence, but also generally by females and thus are associated with femininity, and the accumulation of these qualities may be described as 'cuteness'.

It is important to stress that the reality of where uptalk is used, by who, how, and what it means will probably differ from what Zoella imagines. But it is unlikely that this will impact on this investigation. How accurately Zoe's approximation of her audience and their opinions of uptalk reflects reality would be a pertinent topic if the association between Zoe's speech and the content of comments was being examined, in other words how the content of Zoe's speech effects the audience's feedback. However, this is not the focus of this case study. Here, the priority is to select a speech variable that Zoella *believes* means 'cuteness' to her imagined audience.

5.3 Case Study Research question and hypothesis

As was already defined in section <mark>2.5</mark> the case study research question is:

> Does the direct written feedback received through the
> commenting function influence a YouTuber's speech?

From the comment data it is evident that the commenters describe Zoella as cute less over time. Now that uptalk has been identified as indexical of cuteness across the anglosphere and selected as the dependent variable, the hypothesis is:

> Use of uptalk positively correlates with cuteness-comments and
> decreases over time.

5.4 Method

This section details the data collection, preparation, and analysis.

5.4.1 Data Collection and Preparation

The audio for videos 2 to 58 (see appendix 1 for a full list) was downloaded. The research question is whether Zoella's speech is influenced by the comments and so video 1 provides comment data that potentially influences Zoella's speech in video 2. Thus, it was not necessary to analyse the speech in video 1. The audio was downloaded as wav files via a third party website (Unknown, no date) that now no longer exists, however there are a plethora of other similar sites that offer the same functionality.

Upon collecting the audio data, working transcripts were created. The automatic caption transcripts provided by YouTube were copied from the YouTube interface and pasted into Microsoft Word, reviewed, and corrected where necessary. The resulting transcripts were verbatim but did not include any additional notation as is used in many transcriptions such as those used in conversational analysis. Therefore, features such as the length of pauses, pace, and volume were not transcribed. The transcripts were treated as a working document of the linguistic content that would be scrutinised further once the phonetic feature of interest had been identified. One video did not have an automatic caption transcription on YouTube and so this was fully transcribed.

Next, declarative independent clauses were identified. A declarative is a statement (Bussmann, 1998a, p. 227) and an independent clause (also known as a main clause) is one that can stand on its own (Bussmann, 1998b, p. 716). Prosodic boundaries tend to coincide with the end of syntactic clauses in English so to help determine intonational phrase structure I also examined syntactic structure as recommended by Jun and Fletcher (2014, p. 502). However, it was often difficult to determine the syntactic structure because the data is monologue. Therefore, all breaths, and pauses of approximately 1 second or more were marked to assist in identifying tokens because, theoretically, IPs are separated by such pauses (Nolan, 2006, p. 438). The text prior to each of these breaths and pauses was examined along with the corresponding video data. A detailed token inclusion and exclusion criterion was devised and followed (see appendix 2). Most important to note is that only intonation contours associated with full IPs were included to avoid misidentification, and only independent clauses were included to avoid leading and trailing rising tones that can be found on dependent clauses prior to or following independent clauses (Wells, 2006, p. 69). Further, unlike some prior work, I explicitly distinguish uptalk rises from other declarative rises, such as rises associated with continuation or list intonation. These other kinds of rises were excluded from the analysis.

5.4.2 Uptalk Analysis

The audio data was opened in praat (Boersma and Weeink, 2018) and the IPs marked in a textgrid. Each IP was identified as ending in a rise or non-rise, and then rises were labelled using the Tone and Break Indices system. The Tone and Break Indices labelling system (ToBI) (Silverman et al., 1992) was selected because it has historically been one of the most dominant approaches, and has previously been used to label uptalk.

ToBI is a method for conducting intonation analysis within the autosegmental-metrical (AM) framework. Developed by Pierrehumbert (1980), the AM framework views intonation as discreet units that when used individually or in combination, in context, convey linguistic meaning. These units are perceived as prominent events in the speech stream, such as pitch accents and boundary tones. A pitch accent is associated with a prominent syllable (indicated with a "*") and can provide linguistic meaning by highlighting the word that the pitch accent is on as important (see figure 5.1 for examples of how the placement of a pitch accent can alter linguistic meaning). A boundary tone is associated with the end of a phrase (indicated with a "%") and provides linguistic information in regard to the grouping of words (see figure 5.2 for examples of how the placement of a boundary tone can alter linguistic meaning).

Each pitch accent is assigned as either being a high (H) or low (L) tone relevant to preceding tonal events, and can be bi-tonal also, e.g. L+H*, with one tone being the perceptually dominant of the two (H* in this case). The boundary tones of IPs are also made up of two tones, the first indicating the start of the boundary and the second to indicate the end (e.g. L-L%, L-H%). Combining a pitch accent and a boundary tone creates an intonation contour; a pattern of rises and falls of pitch that occur on the IP. What is "high" and "low" is based on listener perception rather than a quantitative measure of pitch. Also, what is defined as "high" or "low" is speaker specific and relative to their typical pitch range (Silverman et al., 1992). Finally, the more complicated intonation contours were collapsed into a binary grouping of rise or fall and the relative percentage of rise patterns versus fall patterns was calculated.

---

1.   Who are you going with?
                               *   %

Scenario: Who are you going with?

2.   Who are you going with?
                                *   %

Scenario: You have several options and can't decide. First you say one person, then you say another. So, what's your final decision?

3.   Who are you going with?
                                   *   %

Scenario: I don't want to know who your friends are going with. I want to know who you are going with.

4.   Who are you going with?
                                        *   %

Meaning: I know you're going with a group of people but who is your date?

Figure 5.1. Examples of differing meanings as a result of different placement of the pitch accent

---

1.   Who are you going with?   Bert?
                           *   %   *   %

Meaning: The suggestion that they might go with Bert is asked after asking who they might go with.

2.   Who are you going with, Bert?
                           *   %

Meaning: Bert is being asked who he is going with.

3.   Who? Are you going with Bert?
                          * %   *   %

Meaning: Surprise is initially expressed and then confirmation is sort.

Figure 5.2. Examples of differing meanings as a result of different placement of the boundary tone

How to operationalise this variable took careful consideration. Quantifying intonation is difficult. The same intonation contour can mean many things, and different intonation contours can mean the same thing in interaction. It is essential in variationist studies that "like-for-like" is analysed and quantified. Thus, this issue is not as a result of the data being from YouTube but is present in all studies of intonation where "natural" data is being used, that is data that is not collected through asking participants to read aloud a set of carefully constructed sentences. As was just explained, the multitude of intonation contours that were found were collapsed into just two categories; one where the pitch rises at the end and one where it falls, to reflect that the focus of this work is on uptalk as defined by Warren (2016, p. 2); "a marked rising intonation found at the end of intonation units realised on declarative utterances".

This avoids the complexities of examining different intonation contours and what they may each mean. Furthermore, through examining this data it became increasingly evident that there was little diversity in both the video content and the linguistic content, so there will probably be little variation in the interactional meanings of uptalk in this data set. I also predicted that some of the interactional roles that uptalk can play in one-to-one face to face interactions will not emerge in this data. For example, one use of uptalk is to check comprehension and to seek feedback (Warren, 2016). Here, one person is looking at a camera as they speak to an imagined audience, thus with no physical communication partner present it is unlikely that uptalk will be used in this way. Furthermore, the reduction of possible uses of uptalk should strengthen the argument that its used is motivated by social factors.

5.5 Results

First, the uptalk results will be described. Then whether there is a relationship between this dependent variable and the independent variable identified in chapter 4 (youthful femininity comments) will be examined. A more thorough reflection on the learnings from performing this work and the result gained will

be made in chapter 8, once the rest of the thesis practices and the findings made throughout the thesis have been reported.

5.5.1 Audio Data

The 57 wav files totalled to 13 hours, 50 minutes and 49 seconds and a mean length of 14 minutes 34 seconds, with the shortest file being 8 minutes 25 seconds (video 17) and longest being 24 minutes (video 4).

5.5.2 Uptalk Results

The analysis yielded 3,913 declarative clause tokens, a mean of 69 tokens per video. 1,159 of these were rises which is 30% of the total dataset, and a mean of 20 per video. Four dominant rising intonation contours were identified; low-onset low-rise (L* L-H%), low-onset high rise (L* H-H%), fall-rise (H* L-H%), and high-onset high rise (H* H-H%). As can be seen in table 5.1, L* H-H% was used the most, H* H-H% the least, and L* L-H% and H* L-H% near equally. Figures 5.3, 5.4, 5.5 and 5.6 provide an example pitch trace for each of these intonation contours. More complex contours did occur, for example those that included bi-tonal pitch accents such as L+H*, but these were relatively rare and thus have been absorbed into the most relevant of these four dominant rising intonation contours (e.g. a L+H* L-H% contour is categorised as H* L-H%).

It should be emphasised that defining what intonation contour is or is not uptalk is not the intention of this work. As was explained earlier, any contour which ends in a rising boundary tone (L-H% or H-H%) will be referred to as uptalk herein. While a sophisticated labelling system has been used, the classification scheme was reduced to whether the end of the clause was a rise or fall in pitch since the focus of this study is not to describe the phonetic form of uptalk contours or their meanings. Thus, the different kinds of contours identified are not examined separately or at the level of use in interaction and so their potential difference in meaning is not considered either. Further, it would be inappropriate to contribute to this discussion based on data from one speaker.

Table 5.1. Number and percentage of rise tokens

Contour	N (percentage %)
L* L-H%	286 (25%)
L* H-H%	424 (37%)
H* L-H%	291 (25%)
H* H-H%	158 (14%)
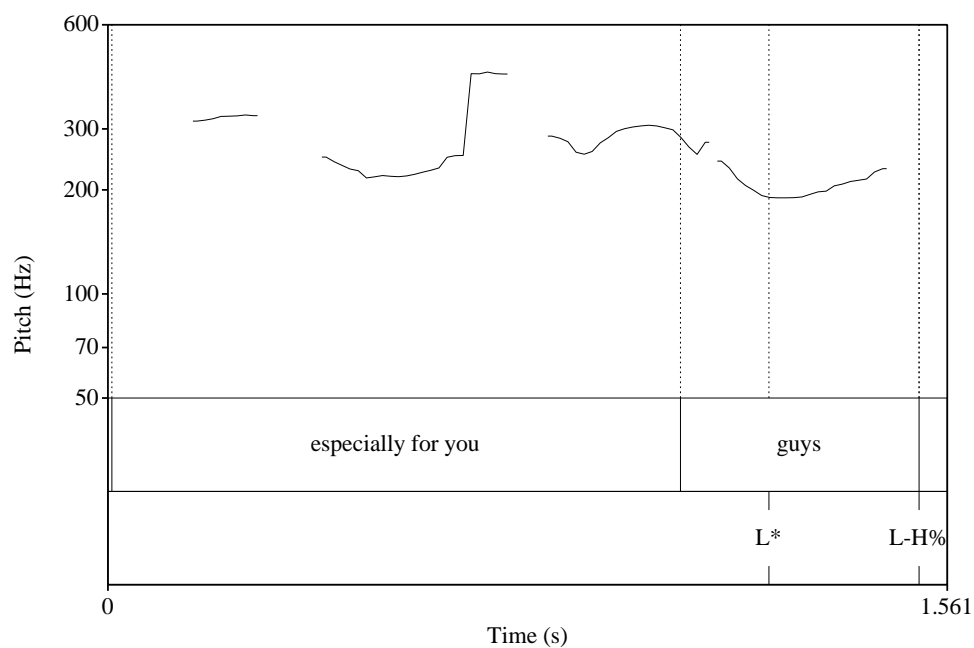
Figure 5.3. Low-onset low-rise (L* L-H%)
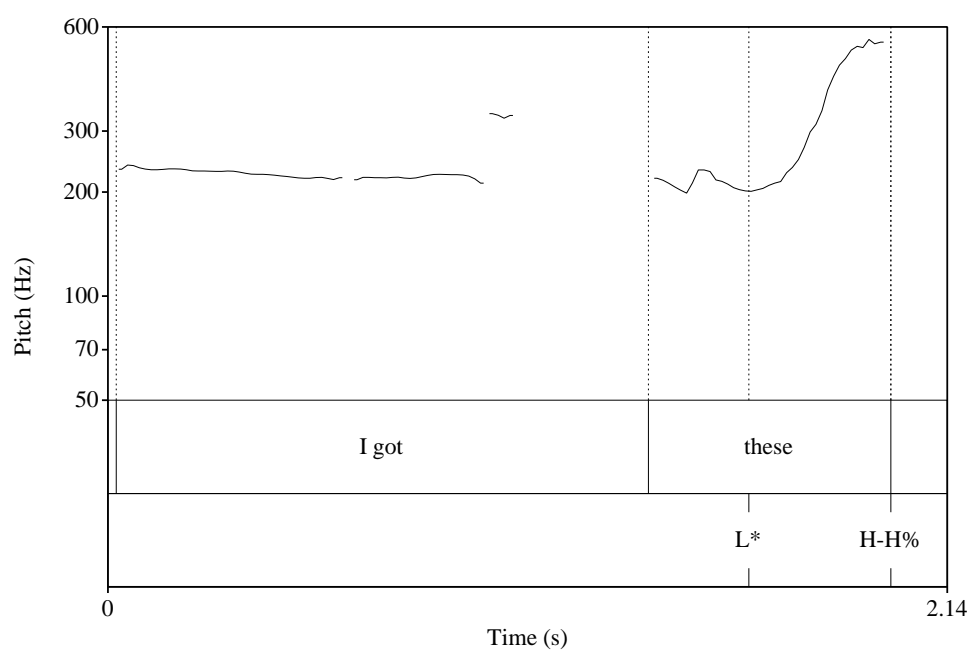
Figure 5.4. Low-onset high rise (L* H-H%)

143
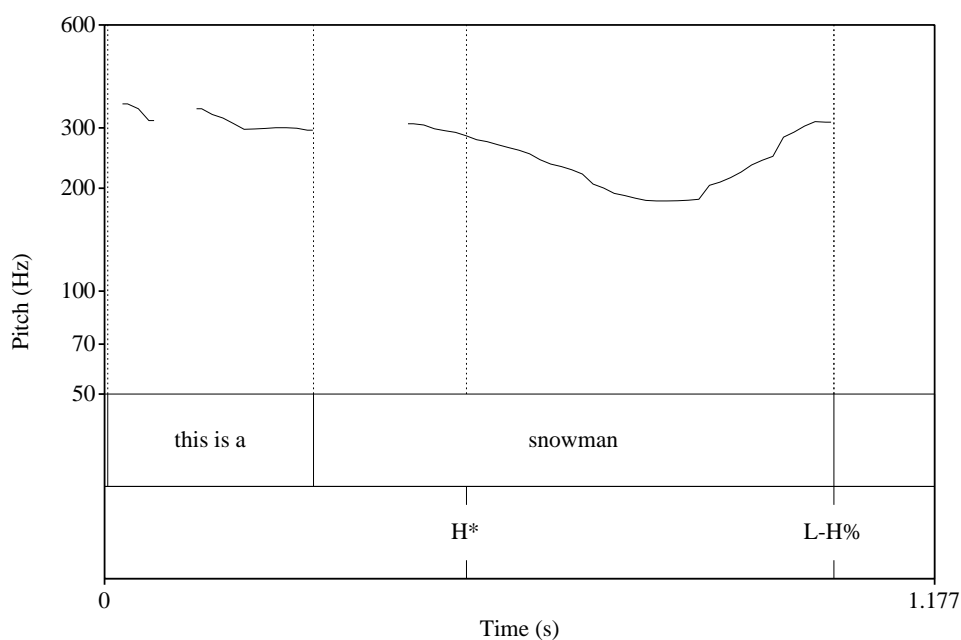
Figure 5.5. Fall-rise (H* L-H%)
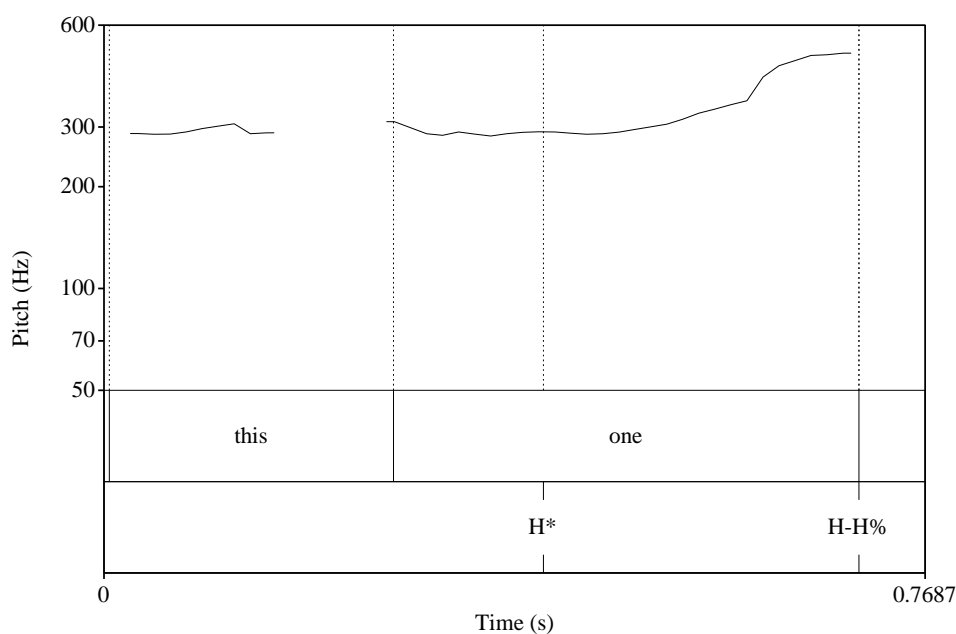
Figure 5.6. High-onset high rise (H* H-H%)

Figure 5.7 is a line graph of the percentage of rise tokens per video across time. As you can see, the percentage of rises fluctuates greatly, although there appears to be a slight downward trend overall and a short period from mid-2013 to early 2014 where the percentage of rises remains relatively low and relatively stable when this period is compared with the rest of the data. The line of best fit displayed in figure 5.8 confirms this overall trend of decreasing rise tokens over time.
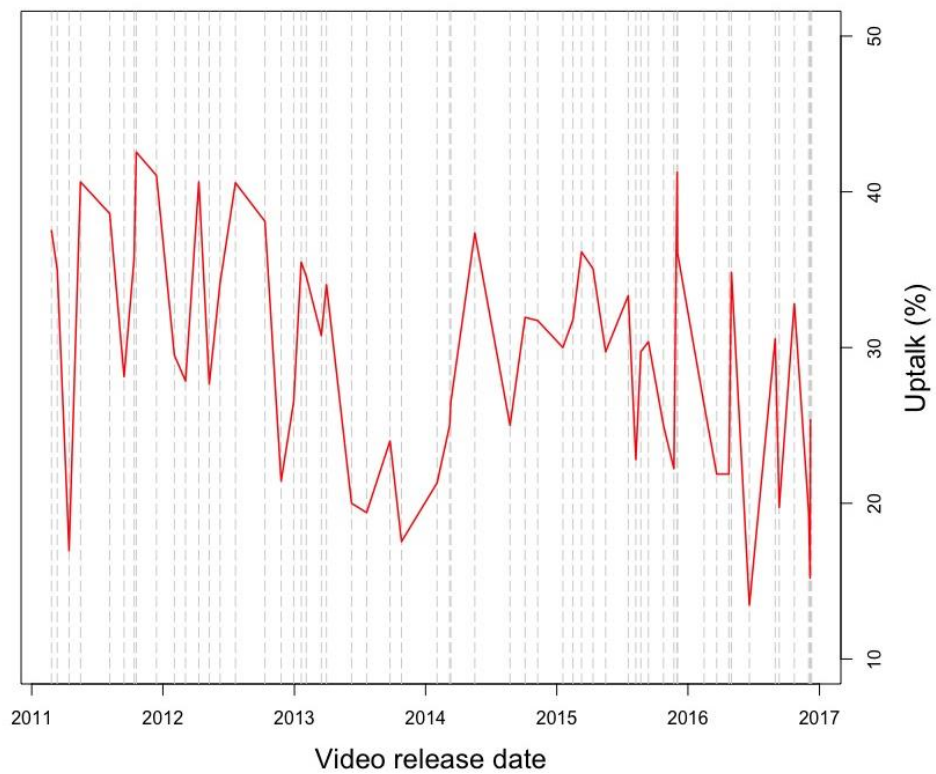
Figure 5.7. Percentage of uptalk per video
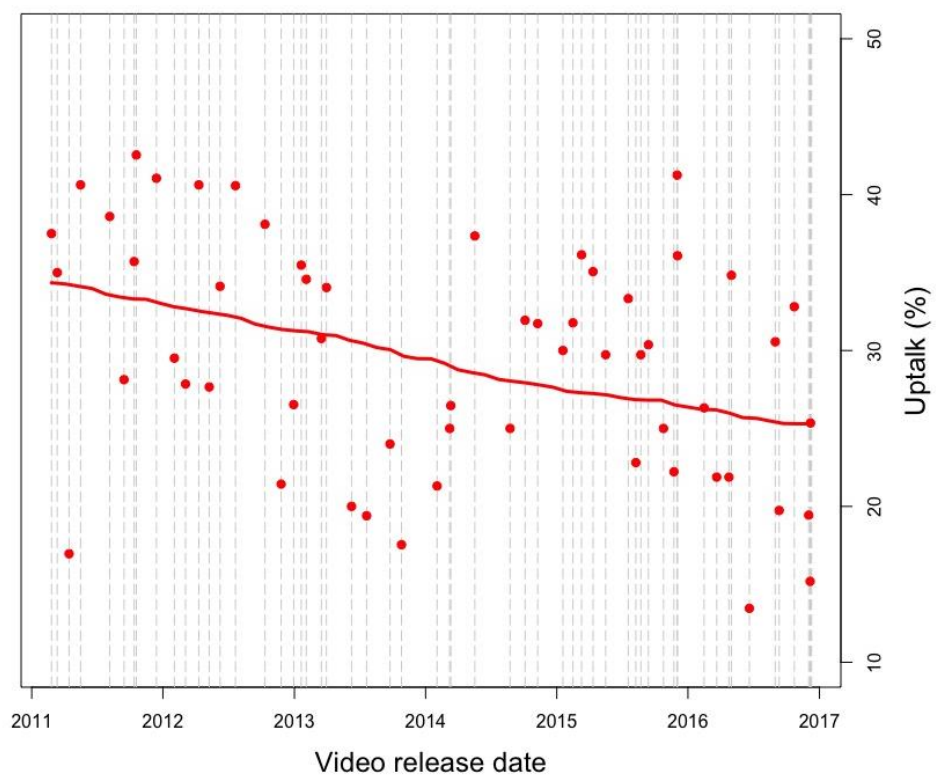
Figure 5.8. Uptalk line of best fit

5.5.3 Relationship between Comments and Uptalk

Figures 5.9 and 5.10 plot the comment data against the uptalk data. Upon visual inspection there appears to be little coordination between the two datasets in regard to their percentages (5.9) or their lines of best fit (5.10). After establishing that uptalk was normally distributed (W = 0.97433, p-value = 0.2652) and the comments was not normally distributed (W = 0. 94562, p-value = <0.05) using the Shapiro-Wilk normality test, their correlation was assessed. Spearman's rank correlation coefficient test found a slightly positive correlation that was not statistically significant (S = 24296, p-value = 0.1123, rho = 0.2126025).
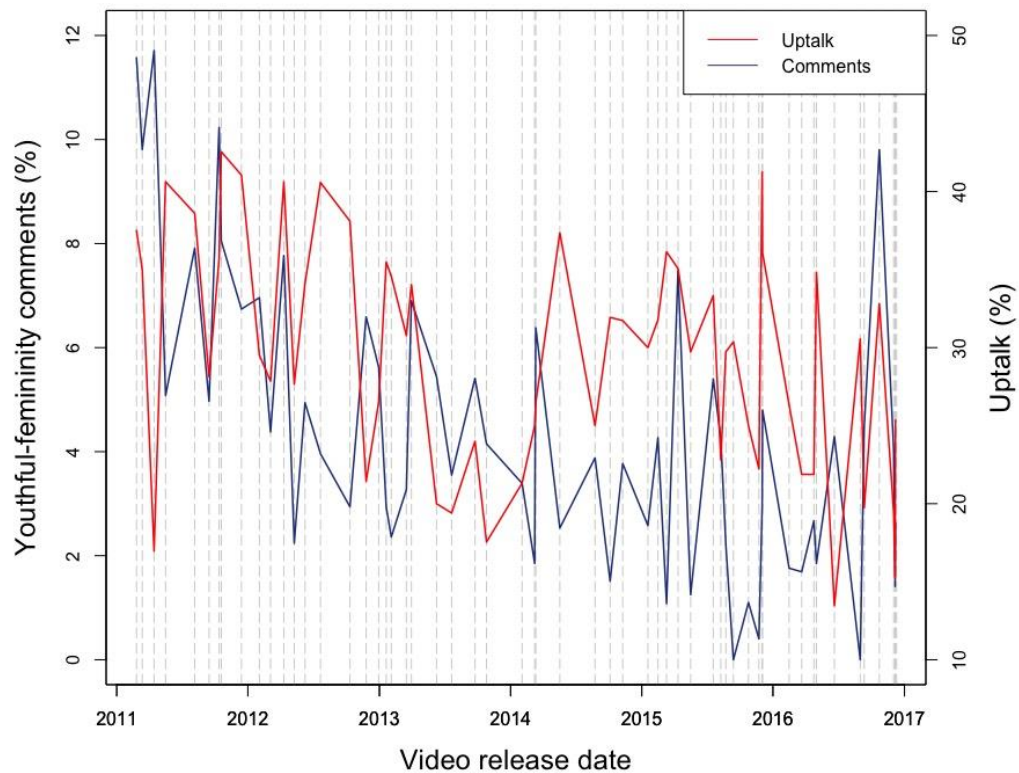
Figure 5.9. Percentage of uptalk and youthful-femininity comments per video
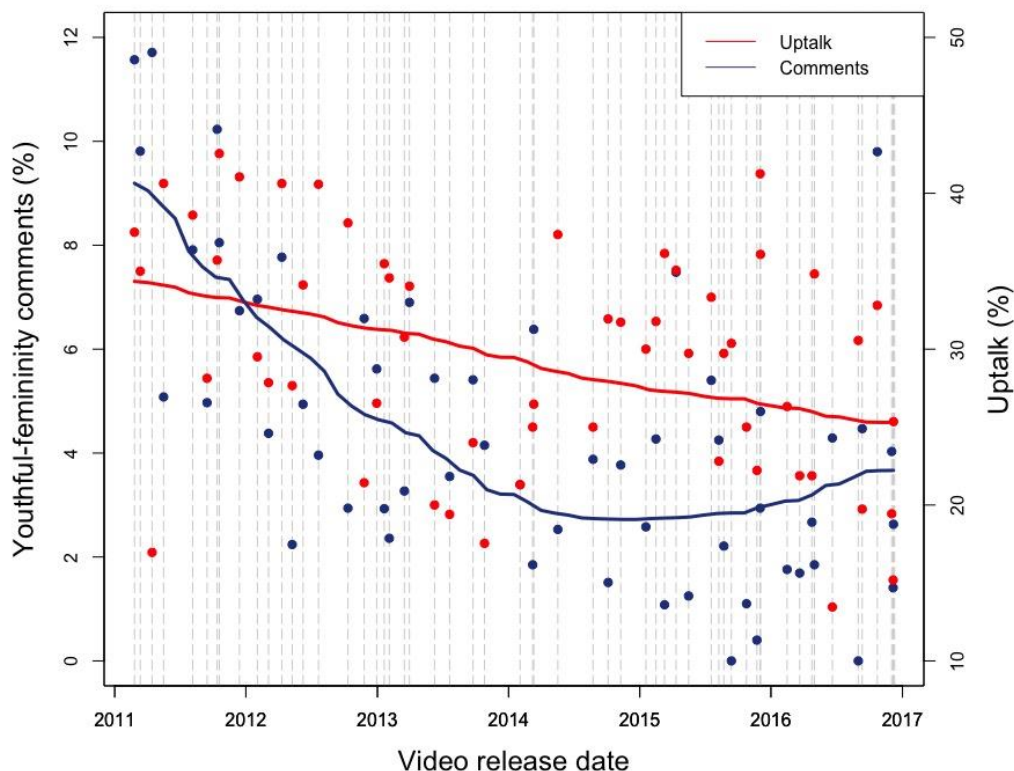
Figure 5.10. Uptalk and youthful-femininity comments lines of best fit

Next, the same statistical analysis was performed on each of the respective intonation contours. Three of the four data sets are normally distributed, do not correlate with the comment data (the individual results are displayed in table 5.2). One data set, H*H-H%, is normally distributed (W = 0.9476, p-value < 0.05) and so Pearson's product moment correlation test was used, the result of which was not statistically significant (rho = 0.063, p = 0.6395). Thus, there is no evidence to suggest that there is an association between Zoella's speech behaviour in regard to the use of uptalk and the cuteness comments.

Table 5.2. Results of Shapiro-Wilk normality test and Spearman's rank correlation coefficient test for 3 of the 4 rise contours.

	Shapiro-Wilk		Spearman's rank	
	p	**W**	**p**	**rho**
L* L-H%	0.6314	0.9836	0.0506	0.26
L* H-H%	0.1872	0.971	0.1761	0.18
H* L-H%	0.0583	0.9602	0.3527	-0.125

5.6 Discussion

The results from examining uptalk in relation to the youthful-femininity comments suggest that there is no association between these two variables (as tested by correlation). A causal relationship cannot be tested if there is no association between the two variables. Therefore, there is no evidence to suggest that the content of the comments influences Zoella's speech. Aside from this result being accurate, it may be that uptalk was not the appropriate speech feature to investigate.

First, as stated in section 5.2, no research to date has identified uptalk specifically as indexing cuteness and an argument for why this speech variable was the appropriate one to investigate was formed from multiple observations and findings. Second, Zoella's definition of her imagined audience in regard to place needed to align with the definition used in this study. These two definitions not aligning poses a barrier to selecting an appropriate linguistic research variable. In other words, while as a researcher I have defined the place of the imaginary audience as the anglosphere, and made a clear argument for studying uptalk in relation to this, Zoella may define the place of her imagined audience as one where uptalk is not used, is not prominent or salient, or doesn't index cuteness. This scenario will be unpacked further in the next section as it suggests that an alternative strategy for defining place of a YouTuber's imagined audience is required.

5.7 Reflecting on Thesis Research Questions

In this section I will describe the difficulties faced in regard to addressing the ethical issues in performing this research and in selecting the linguistic research variable. Both of these difficulties relate to the thesis research questions. I will then recount the decision-making processes I underwent and explain their outcomes. The research activities of addressing ethical issues and selecting the

linguistic research variable are fundamental to any sociolinguistic study, of course, but this online interaction context presents particular complexities that need to be considered.

5.7.1 Thesis Research Question 2

While studying YouTube comment data is relatively uncomplex in regard to ethics (as was discussed in section 4.5.2), the ethical issues surrounding studying YouTube video data are complex particularly for sociolinguistic research. The first issue is whether informed consent is required. The second issue is the tension between anonymity and credit. The third issue are the practicalities of collecting the data.

In this case study informed consent and the tension between anonymity and credit are heavily intertwined. In regard to my Institution's Research Board, they took the same view as when the comment data was considered: the YouTube videos are public data and so informed consent is not required. However, as with many forms of online data, whether the user views their data as being public or private, or rather where they believe their data sits along the continuum of public-private, is contextual (as explored in sections 3.2.2 and 3.2.3). In the case of Zoella it would be difficult to argue that she views her YouTube data as being anything but public. As described in section 4.1, Zoella has become one of the most followed YouTubers from the UK and earns a sizeable income as a result of her online activities. Her work depends on her visibility and those that see her content engaging through various means (e.g. likes, follows, and shares). In other words, her success is measured by the amount of attention she receives.

In light of this, it is highly questionable whether it would be possible to effectively anonymise Zoella as the subject of this case study. If she was to be pseudonymised in a publication it is likely that some readers would be able to identify her based on essential information that would have to be supplied to understand the context of the research (e.g. a female British vlogger whom over the past six years has gone from making videos in her bedroom for her friends to

running a multi-million lifestyle and beauty business). Further, I would argue that the ability to identify Zoella is essential to understanding the research and its findings (although it should be noted that this is upon reflecting on the whole case study including the work detailed in chapters 6 and 7, and not just the work detailed up until this point in the thesis).

However, I would like to emphasise that the sound arguments for using and crediting this data to Zoella does not equate to not needing to carefully considering the ethical implications in regard to the way this data is used. For example, the intention of the analysis was not to be critical of Zoella's behaviour or character from a moral or ethical standpoint. It is not envisaged that the research findings will have an impact upon her reputation, and the intention of this work is not to challenge her genuineness; it is clear from the literature that adjustments in one's communication style can be for a multitude of reasons, is a common occurrence, and is not equitable to deceit.

The third issue is whether it is possible to collect the data in an ethically sound manner. At the outset, it was my intention to use auditory methods of analysis. The primary reason for this was to avoid breaking YouTube's Terms of Service which state that the only method of access allowed is streaming (YouTube Great Britain, 2010, p. 5.L). Therefore, it would not possible to conduct acoustic analyses, which requires a data file to be entered into specialist software, and also to adhere to these Terms of Service without collecting a copy of the data from the person who uploaded it. However, conducting a purely auditory analysis of intonation is intensive and very time consuming. Indeed, when I discussed this approach with those who research intonation most were surprised and apprehensive on my behalf. Standard practice is to use a pitch trace in a specialist software like Praat (Boersma and Weeink, 2018) in conjunction with perceptual analysis to making labelling decisions. Further, it is possible, even easy, to download a copy of YouTube video and audio in a number of formats by using one of many third-party sites on the web.

The controversy over whether to download the data or not was amplified in this interdisciplinary context. What is considered ethically acceptable changes over

time and while change is ongoing there can be differing standards, expectations or opinions across disciplines. This was very evident when considering whether to download the data. It became evident through conversations with linguists that many were already downloading YouTube data for research purposes without an awareness of this web of legal, social and political issues. These conversations were quite uncomfortable in that some thought my cautiousness was unnecessary but also explaining the decision-making process behind my practice challenged others to reconsider their own. I agree with many of the arguments that have been made by both linguists and CMC researchers for why there should be minimal concern for downloading YouTube data for research purposes. However, there are many scenarios in which I would argue there should be great concern when downloading YouTube data. My intention in this thesis is to make initial recommendations in regard to methods and practice with the hope that it will be used by others as guidance in the future. Thus, the ethical issues surrounding data collection need to be carefully navigated.

To address this issue, I sought expert help and advice from Hugh Rhodes, Enterprise Manager and Lawyer at Northumbria University. The outcome of our discussions was that I sent an email to both YouTube and Zoella's management agency, Gleam Futures (see Gleam Futures, 2018), stating who I was, what I intended to do and why, and why we viewed this practice to be within fair use law (see (gov.uk, 2019) for a summary of exceptions to copyright within UK law). I requested that they contacted me by a specific date if they had any concerns and that I would interpret a lack of response by this date as permission to carry out my plans. I did not receive a response; therefore, I downloaded all the data I needed, reviewed my analysis up to this point with the support of the pitch trace (acoustic analysis), and then continued the analysis with the pitch trace. For additional transparency, from when I decided to make Zoella the subject of my study (March 2017) I made regular (between bi-monthly and monthly) attempts to open a dialogue with her management to gain access to additional data but, unfortunately, they have never replied to my messages. It came to light that downloading the data was the right call while I was writing the thesis. When returning to Zoella's videos on YouTube to double check so minor

details (on 01/09/20) I found videos 1 to 4 of the data set had been privatised and so were unviewable.

To conclude, the ethical decision-making process in regard to YouTube video data for the use in sociolinguistics studies is complex and controversial. The learnings gained from navigating these in the work detailed in this chapter significantly informs the answer to thesis research question 2: "What are the ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?".

5.7.2 Thesis Research Question 3

Most of the other learnings gained through performing the work that is detailed is this chapter contribute to answering thesis research question 3: "What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?". The selection of uptalk as the linguistic variable was primarily motivated by its global presence and social salience, although through engaging with the literature an additional, theoretical connection between uptalk and cuteness came to light (theory of sound symbolism (Ohala, 1983)). As identified in section 3.3, a key complexity of selecting linguistic variables in YouTube data was predicted to be that place may be ambiguous. Therefore, the rationale behind selecting a global linguistic speech feature was that Zoella may try to 'catch' as much of her viewership as possible in terms of geography by using the lowest common denominator speech feature (to adopt the term used by (Androutsopoulos, 2014, p. 66) and (Gil-Lopez *et al.*, 2018, p. 127)).

However, as stated above, the findings of this study provide little insight into whether Zoella's speech changes as a result of the feedback her viewers leave in the comments. It may be that the definition of the imagined audience used to motivate selecting uptalk as the linguistic variable (the anglosphere) does not align with Zoella's definition of the imagined audience. Therefore, a change of tact may be necessary. Rather than trying to be inclusive of the actual viewership

by making the imagined viewership geographically broad, an alternative approach would be to attempt to gain insight into Zoella's definition of the place of her imagined audience. This will be explored in the next chapter.

5.8 Summary

To conclude, in this chapter the first dependent speech variable was identified – uptalk. Selecting this variable was guided by the concepts of the anglosphere and global linguistic speech features. Then the audio data was collected, analysed, and uptalk's relationship with the independent variable, the comments, tested. The statistical analysis provided no evidence that there was a relationship between the two variables. Through performing this work, the answer to one of the **thesis research questions** was developed further (thesis research question 2 that focuses on ethics), and another begun to be answered (thesis research question 3 that focuses on selecting linguistic variables).

This result indicates that the initial strategy of embracing the ambiguity of place online by defining the imagined audience as geographically broad was not successful. In the next chapter, I report on an ethnographic analysis of Zoella's videos in an attempt to gain insights into how she imagines her audience and thus define place more specifically.

Chapter 6.

Online Ethnography

> *"Technology complicates our metaphors of space and place"*
> (Marwick and boyd, 2010, p. 115)
> and
> *"[T]he imagined audience becomes visible when it influences*
> *the information [vloggers] choose to broadcast"*
> (Marwick and boyd, 2010, p. 130)

In this chapter I will report the findings of my online ethnography, namely observations in regard to Zoella developing as a vlogger over time, and the place of Zoella's imagined audience that is co-created between Zoella and her commenters. In the previous chapter, the rationale for choosing a global speech feature (Sayers, 2014b) was to be as geographically inclusive of Zoella's potential actual audience as possible. Thus, this framing was motivated by the physical *space* that the potential actual audience inhabits, albeit this was narrowed to the socially defined space of 'the anglosphere' for pragmatic reasons. In contrast, an alternative strategy would be to estimate Zoella's conceptualisation of the *place* of the imagined audience. Both parties share the YouTube interface and the data it contains, such as her videos and the comments. Thus, ethnography was utilised.

In this chapter I review the literature related to celebrification (Zoella's development as a vlogger over time, section 6.1) to assist in the collection and

interpretation of ethnographic data. I also review the literature related to space and place within sociolinguistics (section 6.2) in order to argue that estimating Zoella's conceptualisation of the *place* of the imagined audience is a viable strategy for selecting a speech variable to study. Third, I introduce ethnography (section 6.3) and how it has been implemented in sociolinguistics (6.3.1) and online (6.3.2) before pointing to the most relevant ethnographies related to online video (6.3.3). Based on this knowledge, how ethnography is applied within this thesis is defined (6.4) and I describe how I collected the data (section 6.4.1) and how it's reported herein (section 6.4.2). Next, the main findings in regard to Zoella's development as a vlogger (section 6.6) and what the imagined audience's place is and how it is co-created by Zoella and her commenters (section 6.7) is reported. Finally, in section 6.8 the thesis research questions are reflected upon and section 6.9 summarises the chapter. This chapter addresses thesis question 3: "What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?".

This chapter has reported on the findings from online ethnography observations in regard to celebrification and the co-creation of the imagined audiences' place by Zoella and her commenters through a variety of resources. As is evidenced throughout the data reported in this chapter, many elements of Zoella's videos and the comments they receive change over time as she transitions from microcelebrity to A List vlogger. Notably, the most apparent imagined audiences' place (Britain and American) is relatively stable throughout. However, there is a change in the relative amplitude of these two contrasting imagined audiences' with 'the Americans' becoming increasingly salient in the comments and Zoella's acknowledgement of an American audience also increasing. This provides clear direction for which speech feature to choose to be studied next, which is defined and explored in the next chapter.

6.1 Celebrification

Anne Jerslev (2016) unpacked how the term 'celebrification' has been used in a variety of ways, including before the advent of social media. She argues that the term refers to sets of communicative and cultural practices, their form depending on the media in question and thus the interactional affordances through which celebrification is being performed. Contrasting traditional celebrity (e.g. TV, film) with microcelebrity (e.g. YouTuber) by temporality brings this to the fore. Rather than communication being indirect, delayed and scarce as in celebrity, microcelebrity's temporality is founded on permanent updating and the qualities of immediacy and instantaneity. In tandem, microcelebrity does not conform to the schism of celebrity (ordinary anonymity converting to celebrity at a specific moment in time) but is a continuum of "celebrifying" oneself. Alice Marwick (2013) describes celebrifying as "thinking of oneself as a celebrity, and treating others accordingly" (p. 115) and "the presentation of oneself as a celebrity regardless of who is paying attention" (p. 114). Thus, "there has been a move away from celebrity as a being to a doing" (Jerslev, 2016, p. 5236). However, the continuum of microcelebrity is vast regardless of how it is measured. Thus, Sophie Bishop (Bishop, 2018) uses the term "A List Vlogger" to distinguish an elite group. These vloggers "film in domestic space, in their bedrooms, but in practice are often represented by high profile digital talent agencies" and have "significant capital on the platform, but also appear in fashion magazines, offer merchandise, makeup lines, and hold book deals" (ibid, p.8). Crystal Abidin (2015) also points to these markers of traditional celebrity (e.g. brand deals, management by talent agents, coverage in tabloids) as indicating ascension from microcelebrity.

In the context of YouTube, the filming (potentially with editing) and uploading of videos are the vehicle through which much celebrification is performed. This involves many layers of decision making: what type or genre of video, with what kind of content, structure and activities; filmed where, when, with what equipment, with what resources, with whom, with what camera angle and lighting; if editing is to be used and if so in which software, what kind of editing

effects, with or without opening title; when to upload, what the thumbnail looks like, what the title should be, what content to include in the information box; to name just a few. Another activity is likely to be the monitoring of and, potentially, replying to comments. These media-dependent practices provide rich data. Thus, in addition to communicative and cultural practices that reflect temporality, this ethnography also makes observations that relate to materials - the selection and manipulation of a multitude of resources both digital and physical – through which the continuum of microcelebrity can be observed.

Since the focus of this study is whether Zoella's speech changes in response to the comments left on her videos over time, it would be naïve to not consider other changes over time, namely her transition from microcelebrity to A-List YouTuber, as factors. Further, some thought should be given to whether aspects of this transition or of celebrification more generally influence how the imagined audiences' place is co-cocreated and what place this might be, because celebrification and co-creation of place are performed through the same media content.

6.2 Space and Place in Sociolinguistics

The literature review reported in chapter 2 evidences that studies of language variation and change in CMC to date have either known where the data was coming from (such as Twitter's geolocation service or by knowing the social history of the participants) or the data is explicitly related to a location in some way (such as forums dedicated to cities or regions). Thus, as was unpacked in 3.3.3 specifically, ambiguity of place is theorised to be a key barrier to selecting speech variables within this study. In this section, relevant literature from sociolinguistics is reviewed to gain insight into how this research field currently conceptualises space and place. As will be revealed, not only are speech features tethered to some degree to a space/place but the definition of place both shapes and is shaped by speech variation and change.

Speech is related to space in that the way we speak is dependent upon who we come into contact with; our speech patterns reflecting the social groups that we want to be affiliated with. Thus, the physical space does not directly influence speech but does so through the people that inhabit that space. Historically, with travel being very difficult for most, a lack of movement lead to geographically restricted accents and the ability to carve out natural (rivers, bogs, mountain ranges) boundaries between them. As travel became easier and more affordable, the first linguists to look at how language varies sought out Non-mobile, Old, Rural, Male participants (NORMs): people whose contact with different ways of speaking was minimal. It was thought that NORMS used the truest or most accurate way of speaking for the area, untainted with other ways of speaking. Researchers drew maps with isoglosses: lines where on one side a certain speech feature was used, and on the other side a different speech feature was used (see figure 6.1). This is known almost interchangeably as Dialectology or Dialect Geography (Chambers and Trudgill, 1998).

Figure 6.1. /j/ dropping in Eastern England (following any consonant) (Chambers and Trudgill, 1998, p. 74).

In the advent of variationism, there was a shift towards considering how macro social structures played a role in language variation. In other words, how the speech found in a space (typically a city or town) varied across social categories such as age, gender, and social class (e.g. Trudgill, 1974). However, the potential influence of the relationships between space and the social structures within the space was overlooked in most of the first wave. Thus, in regard to Dialect Geography and first-wave Variationism, to quote Britain (2013, p. 475) "the former asocially quantifying space, and the latter aspatially quantifying society".

The second wave of variationism saw space reclaim a role. Rather than assessing macrosocial categories, the focus zoomed in, on neighbourhoods for example, and how the inhabitants interacted within them. Milroy and Milroy's (1985) social network analysis found that speakers who interacted with the same people in multiple spaces (living in the same area, working together, spending their spare time together) tended to use local, nonstandard speech features. Thus, a lack of mobility and coming into contact with the same people regularly maintained the use of local, nonstandard speech. Consequently, it is argued that those with open, less dense networks (those that are more mobile and come into contact with a greater variety of speakers) are conduits for linguistic change.

Although threads of interest are woven throughout the first and second waves of variationism, recently there has been a more explicit interest in the role of space. Most notable is Britain (2013) transferring the concept of spatiality from the field of human geography to sociolinguistics. Britain (ibid) argues for spatiality to re-examine the roles of contact and mobility in language variation and change, although it appears to be useful to many more sociolinguistics interests, as will be explained below. Spatiality constitutes three interlinked and co-dependent components: i) Euclidean space - "the objective, geometric, socially divorced space of mathematics and physics"; ii) Social Space – "the space shaped by social organisation and human agency, by the human manipulation of the landscape, by the creation of a built environment and by the relationship of these to the way the state spatially organises and controls at a political level"; and iii) Perceived space – "how civil society perceives its immediate and not so immediate environments –important given the way people's environmental

perceptions and attitudes construct and are constructed by everyday practice"
(Britain, 2013, p. 472). Finally, Britain (ibid) adds:

> "Geometric space is appropriated and thus made social through human
> settlement, but social space can never be entirely free of the physical
> friction of distance. And our perceptions and value systems associated
> with our surroundings, although deeply affected by both social and
> Euclidean space, can in themselves affect the way space is later
> appropriated and colonised."

Thus, spatiality is not fixed but constantly in flux, and also these three
conceptualisations of space interact with one another.

As the definition above states, spatiality constitutes three interlinked and co-
dependent components. Thus, when considering how researchers have used
spatiality in their studies, it is not that Euclidean, Social or Perceived space can
be truly separated from one another. It is more likely that one or more of the
three is prioritised somehow in the research question or analysis arguably
resulting in some correlations between research interest and the kind of spatiality
that is primarily considered.

In regard to Euclidean space, Dialectology/Dialect Geography has seen a
resurgence as technological innovations have allowed its practices to evolve
beyond cartography. In addition to using a variety of new data collection and
presentation techniques, Dialectologists now integrate sociolinguistic
frameworks by considering inter-speaker variation (across gender and social
class for example), and changes over time. In regard to Perceived space,
Perceptual Dialectology/Perceptual Dialect Geography has surfaced. According
to Montgomery (2017, p. 153), "The aim of perceptual dialectology […] is to
gather data relating to non-linguists' perception of the dialect landscape" by
asking participants to draw isoglosses on maps of where accent boundaries exist
and then examining the result of aggregating many of these maps, for example.

In regard to Social Space most work has considered how changes have affected
identity and thus speech. For example, Sali Tagliamonte (2017) considered the

impact of the built environment changing (urbanisation and its repercussions) in York. She found that the use of 'in' for 'ing' in nouns was increasing across the younger generations. She attributes this, along with the patterning of two grammatical features, to maintaining a range of nonstandard speech features in order to signal specific local identities. Carmen Llamas (2007) also found that changing Social Space also impacted speaker's productions, but this time along a political dimension. Her study focused on the English town of Middlesbrough which had been reassigned to three different jurisdictions within just 30 years: from being in the North Ridings of Yorkshire, to County Borough of Teeside in 1968, as part of a new county named 'Cleveland' in 1974 and then as its own authority in 1996. She found generational differences in speech that corresponded with these changes in political affiliations and thus the labels that participants used to describe themselves (answers to "What accent would you say you had?").

Another direction in sociolinguistics is to consider *place* rather than *space*. Reed (2020) recently introduced Agnew's (2002, p. 16) components of place to sociolinguistics: i) Locale "or setting in which everyday life is most concentrated for a group of people", ii) Location – "or node that links the place to both wider networks and the territorial ambit it is embedded in" such as a city, region or nation and thus the groups of people that inhabit them, and iii) Sense of Place – "or symbolic identification with a place as distinctive and constitutive of a personal identity and as per of personal interests". Transferring this theorisation from the field of political geography, he argues that "a speaker's relationship to place [that is, Agnew's (2002) Sense of Place,] is crucial to understanding language variation" (Reed, 2020, p. 7). In other words, how a speaker feels towards a region, for example, can influence their usage of features that index that regional background.

Reed (2020) reveals how many variationists have unknowingly considered speakers' Sense of Place in their work. In the first variationist study, Labov (1963, p. 305) describes the fishermen's feelings towards Martha's Vineyard as "the ever present conviction that the island belongs to them" in the face of seasonal 'invasion' from mainlanders. In a key second wave study, Eckert

(1989) found speech features patterned across the groups known as the "Jocks" and the "Burnouts". These local identities reflected how they felt about the area (whether they were school-oriented or urban-oriented, respectively). In his work, Reed (2020, p. 8) operationalises Sense of Place as "rootedness": "the relative strength of […] local place-based attachment, where certain individuals may have a stronger connection to particular place". In Reed's (ibid) study, quantifying rootedness allowed the speech pattern across three sisters to be explained.

Comparing Britain's and Reed's work exemplifies how distinguishing space and place has only just begun in sociolinguistics. It is not that previous work had conflated these concepts or not clearly delineated them in error, but space and place are often synonymous and separating them may not have been necessary or fruitful in previous work. This is aptly illustrated by Montgomery and Moore's (2017) edited volume entitled "Language and a Sense of Place". Although the same term is used as Agnew (2002), the concept of a "Sense of Place" in this volume is far more inclusive. Rather, along with not wanting to present one unified approach to language variation and change, it appears that Montgomery and Moore embrace many different definitions and uses of space and place.

Further, it is also evident from reviewing the literature that considering how the relation between (distinguished) space and place in sociolinguistic research has not begun. It is reasonable to argue that for a speaker having a Sense of Place (Agnew, 2002) is constant, albeit that what that Sense of Place is (which could be rationalised as Reed's (2020) rootedness) is in flux. Thus, it may be that spatiality provides a framework for examining language variation and change in relation to shifts in space, and Sense of Place provides a lens through which this spatiality framework can be considered. Returning to one of the examples above used to illustrate spatiality, Sali Tagliamonte (2017) states "The comments from the York interviews support the development of positive affect related to place" (p.32) and that "[t]ogether with the results from the three variables, such comments converge in suggesting a particular allegiance to the city" (p.33). In other words, the changing of the Social Space through the built environment also manipulated Sense of Place and thus rootedness ("positive affect", "allegiance")

for its younger generation of residents. This then influenced their speech. Hence, for variationism the untangling of space and place may provide useful analytical devices that can structure the examination and explanation of data.

To summarise, the statement that speech is related to space still holds although contemporary research has seen significant developments in questioning why and how this comes about. Further, state of the art research has just begun to introduce considering speech in relation to place. However, these two concepts are yet to be untangled, and then brought back together, in sociolinguistics.

6.3 Ethnography

Androutosopoulous (2006a, p. 423) argues that "If […] a sociolinguistic approach to CMC takes online communities and discourse as its starting point rather than the medium and its modes, ethnography seems an indispensable part of both quantitative and non- quantitative approaches". Further, in reference to context collapse Piia Varis (p.58) states that "[s]uch contextual complexities potentially shape people's communicative practices and need to be ethnographically established". As outlined in 4.1.1, a key "contextual complexity" (ibid) is the ambiguity of who the audience is. danah boyd (2007) found that users take cues from the social media environment to imagine their audience, cues that Marwick and boyd (2010, p. 130) found to include "linguistic, cultural, and identity markers". In regard to content creators, such as Zoella, Marwick and boyd (ibid) also state that "the imagined audience becomes visible when it influences the information […] users choose to broadcast" as they "conceal or reveal information based on who they imagine to be listening". Based on these recommendations and the need to gain insight into Zoella's and her commenters' experience of the YouTube environment to tune in to these cues and markers, ethnography was rationalised as an appropriate approach.

Of course, through analytical reports (see YouTube Great Britain, 2020) it is likely that Zoella will have some knowledge of where her actual audience is

viewing from. However, I am not party to that information and in future projects researchers may not be either. Plus, while these analytics may make a contribution, it is unlikely that this data alone will equate to Zoella's conceptualisation of her imaginary audience and their place. Further, as has been outlined above, I would argue that the place of the imaginary audience is cocreated through action: the action of commenters revealing information about themselves and the action of Zoella adjusting her content and interactions to acknowledge this information. The actual audience (as detailed in analytical reports) is not one and the same as the commenters, Zoella's imaginary audience nor the fellow audience members that the actual audience imagines. Hence, the employment of ethnography.

The aim of an ethnography is to understand the social world through the experiences of the research participants. Thus, ethnography is rooted in interpretivism, a research philosophy that "views individuals as actors in the social world rather than focusing on the way they are acted upon by social structures and external factors" (O'Reilly, 2009, p. 119). Ethnography is a methodology, not a method, and hence can take a diverse range of forms, although its transition to and inclusion of digital and online technologies required a rethinking of the emphases that are made and the assumptions that underpin its more traditional form, as will be discussed in this section. But these variations, both on- and off- line, share 5 key principles in regard to what ethnographers do, as Hammersley and Atkinson (2007, p. 3) outline:

i) People's actions are studied in everyday 'natural' contexts, rather than under conditions set up by the researcher (e.g. experimental, highly structured interview). The contexts being studied are known as 'the field site' and the researcher spending time collecting data in these contexts is known as being 'in the field' (O'Reilly, 2009).
ii) Data are gathered from a range of sources, but participant observation is often the central component.
iii) Ethnography is made up of a family of data collections methods and which methods are used is decided upon in response to the data and experiences gained during fieldwork. Thus, rather than using a set of fixed, predetermined methods,

what methods may be most appropriate emerges as the ethnography is performed (iterative-inductive research (O'Reilly, 2009, p. 3)).

iv) To facilitate a detailed, in-depth study, the focus is usually small-scale.

v) The analysis focuses on human behaviours and practices, interpreting their meaning, function and consequences within a local and, potentially wider, context.

In this section, the use of ethnography in linguistics, specifically sociolinguistics (6.3.1), and online (6.3.2) is reviewed and the most relevant previous ethnographies of online video highlighted (6.3.3).

6.3.1 Ethnography in Sociolinguistics

One way in which ethnography can vary in its employment in research is whether it is the only methodology used and thus the predominant analysis, or whether it is one of several, potentially many, methodologies used and thus its role is to guide the analytical decisions in regard to the other data collected and/or support the interpretation of their results. The latter is the predominant way in which ethnography has been employed in sociolinguistics.

As was explained in section 3.1.2, the first variationist sociolinguistic study (Martha's Vineyard, (Labov, 1963)) included an ethnographic element. Through ethnographic interviews Labov found that some inhabitants of the island had negative feelings toward the mainlanders who visited every summer and other inhabitants did not, and the speaker's stance in this regard influenced their pronunciation of vowels. Although the knowledge gathered from the ethnographic interviews was crucial to understanding the speech patterns of the people who lived on the island, the usefulness of ethnography was initially overlooked and most subsequent work from other researchers focused on objective social categories (e.g. age, class, gender). In the 80s the potential of ethnography to sociolinguistics re-emerged, a key catalyst being (Eckert 1989), as already mentioned in section 6.2.1. Penelope Eckert's ethnographic observations informed her of the social groups at a suburban Detroit Highschool,

and interpretation of the speech data in relation to these social groups revealed an explanatory pattern. Since then, ethnography has become a staple methodology in sociolinguistic research, mostly used to explore linguistic variation in identity construction (e.g. Lawson, 2011; Alam and Stuart-Smith, 2015) and, more recently, the links between linguistic variation in identity construction and place (e.g. Burland, 2017; Snell, 2017).

6.3.2 Ethnography online

Transitioning ethnography to digital technologies has resulted in its appearance in a broader range of disciplines and thus it being called many different names such as those listed by Piia Varis (2016, p. 55): digital ethnography, virtual ethnography, cyberethnography, discourse-centre online ethnography, internet ethnography, ethnography on the internet, ethnography of virtual spaces, ethnographic research on the internet, internet-related ethnography, and netnography. While the work conducted under these terms is united in their interest in computer mediated communication (CMC), "[t]his is basically where the commonalities end; so diverse is the field – if such a field can even clearly be identified" (ibid). In addition to the differing disciplinary foundations onto which an ethnography may be built, this diversity is partly a consequence of the diversity of technology at a given time and its continual evolution. Thus, approaching CMC with this methodology in mind amplifies the need for reflexivity in the methods performed.

Having said this, in 2009 Robinson and Schulz (2009) identified three phases of ethnography in CMC under which work can be categorised. They refer to these as the i) Pioneering, ii) Legitimizing, and iii) Multi-modal phases. The pioneering phase saw ethnographers engaging with early adopters of online technologies and "conceptualiz[ing] new media as offering a space of identity play and deception" (ibid, 686). With an increase in users and the types of activities conducted online expanding, in the second, legitimizing phase ethnographers' topics of interest evolved and through their work found cohesion in online and offline identities and interactions. Thus, a key difference between

the first two phases is 'where' the ethnography took place and how they viewed the relationship between 'online' and 'offline' worlds. The first phase took a segmentalist perspective, that online and offline were clearly delineated, and thus limited their work to online contexts only, predominantly virtual worlds. Whereas the second phase took an integrationist perspective, endorsing "a vision of the cyberfield as part of a flow between online and offline realities" (ibid) and so regularly included both online and offline research activities and data. In the third phase, ethnographers have focused on how to utilise and manage multiple modes of interaction, i.e. visual and aural material as data as well as text.

While this is an accurate record of literature trends and thus development of online ethnographic practices and theory at the time, thinking about more contemporary work in this way is somewhat unhelpful. "Phases" suggests that prior work is superseded by subsequent work, inferencing that the latter is superior to the former. In fact, work that can be categorised under the first two "phases" has endured but also developed in theory and practice in order to continue to ask justifiable and relevant questions. Thus, it would now be more appropriate to consider Robinson and Schulz's (2009) first two phases as evolved into two broad approaches with the analysis of multi-modal data (supposedly consigned to the third phase) being fully incorporated into both.

Evidently, when transitioning ethnography online "the concept of the field site is brought into question" (Hine, 2000, p. 64) and the issue of 'where' an online ethnography can or should be performed is heavily debated. Christine Hine's (ibid) third of her ten principles of virtual ethnography is that mediated interaction should be thought of as mobile rather than multi-sited and thus, as the fourth principle emphasises, "The object of ethnographic enquiry can usefully be reshaped by concentrating on flow and connectivity rather than location and boundary as the organizing principle" (ibid). This thinking is reflected in Kozinet's (2010) Netnography approach, which focuses on online community and its movement across the web, and also Postill and Pink's (2012) rationalisation of social media as a "messy web" of field sites, where the researcher is "carried" through the web by various features of interconnectivity (e.g. hyperlinks in blog posts, Twitter hashtags). Further, Hine (2000) suggested

that ethnographers can follow a particular event and reaction to it, her example being the discussion of a criminal trial. In contrast, Boellstorff and colleagues (2012) argue the legitimacy of focusing on one site in an online or digital ethnography, referring to Marcus' "strategically situated (single-site) ethnography" (1995, p. 110). Therein, Marcus (ibid) argued that there is a difference between ethnography that is genuinely single-sited, as was the initial convention in traditional ethnography, and an ethnography that is a "foreshortened multi-sited project" (ibid):

> "what goes on within a particular locale in which research is conducted is often calibrated with its implication for what goes on in another related locale, or other locales, even though the other locales may not be within the frame of the research design or resulting ethnography" (ibid).

Thus, not moving around does not necessarily make an ethnography single-sited.

Last in regard to the 'where' of ethnography, Christine Hine's (2000, p. 64) second principle of virtual ethnography is: "Cyberspace is not to be thought of as a space detached from any connections to 'real life' and face-to-face interaction". Similarly, Postill and Pink (2012, p. 3) state "social media practices cannot be defined as phenomena that take place exclusively online". Indeed, from surveying the literature there seems to be an agreement that a dichotomy of online/offline or digital/material is false.

To end this section, I diverge from 'where' to consider 'when' instead. Temporality is discussed far less in the ethnographic literature, presumably because prior to the advent of digital technologies all research activities had to be conducted in real-time. Now online, it is possible to retrieve time-stamped data. Kozinets (2010) uses such 'archival data' – data created before the research began – as one part of his netnographic approach. In reporting on a case study about online news discussion boards, Hine (2000) outlined the main benefit of using such data. In the real-world the ethnographer has to instantly make decisions about what data to collect and how, resulting in some selectivity and shaping of the research during the collection process. But online, ethnographers

are not under the same constraints and pressures because all interactions and other meta-data are automatically recorded. Therefore, this detailed data can be reviewed, reconsidered and refined within the interests of the ethnography. Thus, "It appears that ethnography can be time-shifted so that the ethnographer's engagement can occur after the events with which they engage happened for participants. Ethnographer and participants no longer need to share the same time frame" (Hine, 2000, p. 23). However, Hine (ibid) also points out that the ethnographer not being present in real-time during events means the experience of the participants cannot be observed or understood to the same degree because of the very methods by which such archival data is collected.

In relation, Postill (2017) recently aired frustrations in regard to temporality in ethnographic writing, particularly those that consider media and social change. He argues that writings favour "present continuism" thus "conflating the recent past, the present and the near future in a fuzzy 'now'" (ibid, p.22). Instead, he posits reporting times and dates along with observations rather than leaving them in an unspecified continuum. It is not to say that experience is linear nor that ethnographic writings should be diary-like; analysing and understanding ethnographic observations requires reflection as well as time to collect relevant experiences and other materials. But clarifying when and in what order in ethnographic writing acknowledges the processual nature of social change and ethnography itself. This seems particularly pertinent for online studies considering that online platforms continually evolve but also because many of them are asynchronous in nature.

6.3.3 Ethnography and online video

To date, there is a notable body of ethnographic work on online video, with the following being examples of work that are most relevant to this thesis. Arguably one of the first ethnographies of online video was Theresa Senft's (Senft, 2008) *Camgirls*: "women who broadcast themselves over the Web for the general public, while trying to cultivate a measure of celebrity in the process" (ibid, p.1). Conducting fieldwork in the US in the early 2000's, Senft coined the term

'micro-celebrity' (discussed above in section 6.1) and reported the variety of self-presentation strategies used.

A decade later, Florencia García-Rapp's (2017b) thesis focused on one beauty vlogger in particular, Bubz, generating a wealth of insights. These include: the presence of two spheres of influence (community and commercial) and how Bubz creates different types of content to build her value in each of these spheres (García-Rapp, 2016); the tension between managing one's self-presentation to balance the professional behaviours required from YouTube to adhere to their guidelines and function within its business model, with the audiences' want to watch an aspirational yet relatable vlogger (García-Rapp and Roca-Cuberes, 2017); how the markers of authenticity are community-specific and the role authenticity plays in legitimizing a vlogger's position of celebrity and maintains this status (García-Rapp, 2017a).

In contrast, Sophie Bishop (Bishop, 2018) investigated many beauty vloggers and the wider context of the vlogging industry through a feminist political economy lens. Through online ethnography, immersing herself in the content of many beauty vloggers including Zoella, offline ethnography attending industry events, interviewing beauty bloggers, their managers and other stakeholders, and analysing ancillary media, she identified the practices and labours of beauty vloggers and how these reify already existing, offline social inequalities. For example, a key element to beauty blogging success ('authenticity') is both classed and raced; the 'A List' beauty vloggers of the UK are overwhelmingly white and middle class (or at least aligning with middle class values and performing a middle-class persona). Florencia García-Rapp's and Sophie Bishop's work are very valuable resources for this project. Their rich descriptions and detailed insights of data that is highly relevant and phenomenally similar to this case study provide greater understanding of the very complex context within which the observations herein need to be interpreted.

6.4 Ethnography in this thesis

The work herein necessitates reflexivity towards the particularities of conducting research on YouTube with sociolinguistic intentions, and so diverts from the typical methodological responses to several of the principles of online ethnography described above, predominantly in regard to 'where' and 'when'. Coincidently, these reflect the most debated aspects in regard to transitioning traditional ethnography into the digital realm because of "the capacities of the Internet to restructure social relations in time and space" (Hine, 2000, p. 11). In addition to the literature above informing methodological decision making, I took guidance from the practices of Florencia García-Rapp (García-Rapp, 2019). Further, it should be emphasised that in this thesis ethnography is not used in its major form and is not the predominant analysis, but is one of several methodologies and it's employed to guide an analytical decision (what speech variable to choose) and support the interpretation of that variable's results, specifically.

First, this ethnography is not just limited to being online or platform specific (YouTube), but centres specific videos. This is pragmatic in regard to the research question: to gain insight into Zoella's conceptualisation of her imagined audiences' place through interactions that are bound to the media where the dependent variable data, her speech, is present. However, observations made from other videos and websites associated with Zoella are used to provide context for and assist in the interpretation of the haul videos and their surrounding interactions.

Second, in regard to the comments, the ethnography is restricted to the data that had already been collected for the content analysis (detailed in chapters 4). This, again, is pragmatic in regard to the research question. Once the video is uploaded it cannot be edited and thus represents a specific moment in time. Comments, however, can be posted and deleted as long as the video is online. Thus, collecting and analysing a random sample or all available comments would

collapse the linear timeline (video, comments, video) and the potential cause and effect relationship between these that is being investigated.

Finally, I have chosen to refer to the work presented in this chapter as an "online ethnography". Using this term more clearly indicates that the data collection and analysis is limited to being online but is also broad enough to not ally itself with any of the terms stated in section 6.3.2 and the methodologies that may be associated with them.

6.4.1 Data Collection

It is important to note that I had already engaged with the data significantly before approaching it ethnographically. I had watched many of Zoella's and other beauty vloggers' videos (e.g. Louise Pentland (Pentland, 2020), Tanya Burr (Burr, 2020), Fleur De Force (De Force, 2020)) when considering who to select as the subject of the study.

Further, I watched many of Zoella's videos in the process of defining the dataset. Then, I had watched each video in the dataset at least four times: 1) to check and edit YouTube's automated transcript, 2) to add pauses, breaths and edits into the transcript, 3) to identify declaratives phrases, 4) to perform the ToBI coding. I had also engaged with sources external to YouTube (e.g. newspaper pieces, magazine articles, blogposts). Finally, I read all the comments collected in order to conduct the content analysis. Thus, by the time I came to engage with the data from an ethnographic stance explicitly I had already immersed myself in the research site and collected a lot of knowledge of Zoella and her vlogging life, and many observations of her behaviour beyond her speech, as well as collected a lot of knowledge of her viewers. Much of this had been captured in notes in various documents, unsystematically. I would argue that, retrospectively, this can be considered akin to an ethnographer's initial immersion in a field site, although admittedly I mostly became acquainted with the data through a specific set of actions (analyses methods in chapter 4 and 5) for a particular purpose (answer an already formed research question). However, I was still 'boot-

strapping', building understanding of Zoella, her vlogging and commenters incrementally as I figured out what it was I wanted to know about (see Hine, 2015, p. 25) or, rather, what I needed to know about was revealed through the result of the comment and uptalk analysis.

Therefore, I collated the notes that I had already made into one research journal. This process prompted my remembering of other relevant observations that I had not recorded, and so I reflected on these and added them. I then grouped similar observations together. Many of these were followed with questions marks: 'In which video did this happen? What exactly was said?' And so, the next step was to revisit the data to reconfirm my observations so far and further evidence them by assigning them to specific videos and comments. I also collected illustrative screen grabs to enrich my notes. These were mostly of the YouTube interface, focusing on the video window. Also, although I had unwittingly collected a wealth of data, I felt it was important to perform some data collection with ethnographic purpose. Thus, as I was embellishing my notes I also gathered new observations, assigning them to already established categories as well as creating new categories, through a constant comparison approach (Parry, 2011).

6.4.2 Data Navigation

To help navigate the data, the timeline will be split into two periods representing different phases in Zoella's development: i) Microcelebrity, and ii) A List. These periods were defined through identifying multiple markers that Bishop (Bishop, 2018) and Abidin (2015) state differentiate a microcelebrity from an A Lister.

The Microcelebrity dataset is from videos 1 to 17, posted between 25/02/2011 and 25/11/2012. Video 1 is the fourth video she ever posted, thus the dataset includes the earliest Zoella videos. The A List dataset are videos 18 until the end of the dataset (video 58, posted 10/12/2016). The start of this period was defined by the first occurrence of one of Bishop's (Bishop, 2018) and Abidin's (2015) markers: Zoella signing to the digital talent agency Gleam Futures (2019). It is not clear the exact date that this occurred, but a video posted on the 9th of

December 2012 stated the Gleam office as Zoella's postal address in the information box (Sugg, 2012), and so the first haul video posted after this (video 18 on the 30/12/2012) will be defined as her transition to a A List vlogger.

After this, all the events that signal A List vlogger occur during this dataset. In regard to indications that Zoella is receiving significant capital from her activities, 'Zoe Sugg Limited' (Companies House, 2013a) and 'Zoella Products Limited' (Companies House, 2013b) were incorporated with Zoe Sugg as Sole Director (incorporation dates 12/02/2013 and 25/11/2013 respectively), as well as 'Crew Live Limited' being incorporated with Zoe Sugg as one of the Directors (28/05/2014) along with eight other YouTubers, all of whom are signed with Gleam Futures, and Dominic Smales, Gleam Futures' CEO (Companies House, 2014). Three more companies ('ZS Lifestyle Limited' (Companies House, 2016c), 'ZS Beauty Limited' (Companies House, 2016b), and 'Pippin Productions Limited' (Companies House, 2016a)) are incorporated in July 2016 with her as Director. Zoella also releases a book series (Sugg, 2014a, 2015d, 2016c), homeware products (reported in (Boyden, 2016) for example), and multiple lines of beauty products (as Zoella shows in videos such as (Sugg, 2015c, 2015e, 2016b)). She appears in print media (UK Vogue (Sheffield, 2014)) and on terrestrial television multiple times (e.g. The Great Comic Relief Bake Off (Love Productions, 2015)) as well as at YouTube conventions in the USA and UK (Playlist Live 2013 in the USA (as she vlogs in (Sugg, 2015a)), and Summer in The City (as she vlogs in (Sugg, 2013c)).

But these ventures were not without their battles[4]. Soon after launching her first book it was revealed that she used a ghost writer, which wasn't in keeping with her statement on the book's cover: "My dream has been to write a book, and I can't believe it's come true. Girl Online is my first novel and I'm so excited for you to read it." (Sugg, 2014a). As Lucy Hunter Johnston summarised in the heading of her Independent piece, "Yes, using a ghostwriter matters when your

---

[4] While there have been far more damaging scandals in regard to Zoe's commercial empire (e.g. the Zoella advent calendar (Wilkinson, 2017) and the Hello World event (Singh, 2017)) these occurred from 2017 onward, beyond the timeline defined in this thesis, and thus are not considered.

whole brand is built on being authentic" (2014). Authenticity is also an important issue in the haul videos that are the focus of this thesis, as will be explained in section 6.5.1.

Re-examining these datasets to consider communicative practices, they reflect the observations that Anne Jerslev (2016) made of Zoella and celebrification. For example, in regard to Zoella appearing in mainstream media Jerslev states that "the broadening of her field of operation to include more traditional media changed the temporality of instantaneity constructed on YouTube" (ibid, p.5235) to the temporality of delay, indirect and scarcity.

6.4.3 Reporting Comment data

It is important to clarify how the comment data is referred to throughout this thesis. As was unpacked in 3.2.2, there are many issues with using publicly available data at all stages of research. Here, the crux of the matter is that it would be impractical to request informed consent from Zoella's commenters, but direct quotes can easily be searched for and thus deanonymize commenters. Thus, the strategy herein is twofold. First, whether a comment was placed on a video during the Microcelebrity or A List period will be stated, but not which specific video. Further, the comments will not be directly quoted but minimally rephrased to prevent them being found through searching. These rephrasings will not be placed in quotation marks but in italics to differentiate them from the main body text. In this rephrasing strategy, placeholders will be used where possible. For example, if a comment was "I love you so much Zoe!!! Sending hugs and kisses from Arizona. Love, Scarlett x" a rephrasing using placeholders would be: *Love you Zoe! Hugs & kisses from [US state]. [commenter name] xxx.* A double space either side of a forward slash (" / ") will be used to separate each comment when in a list or series. Finally, when discussing the interaction between comments and the content of specific videos (such as what Zoella says) the video number will not be given for the same reasons given above.

6.5 Vlogger development

In following Zoella's journey from microcelebrity to A List vlogger, a great deal of change is evident yet there is also consistency throughout her videos. The change in a number of elements reflects her becoming an A List vlogger: i) qualities of the videos and the practices in making them, ii) the brands and products that she features, and iii) the content of comments and commenting behaviour of both Zoella and her viewers. Each of these will be described more thoroughly below. Of course, while these aspects and those reported in the next section are described separately they cannot be truly untangled.

6.5.1 Video qualities and practices

Zoella's video making practices developed significantly over the six years. In the earliest videos we witness technical difficulties (e.g. light flare (see fig. 6.2) and glitching (see fig. 6.3), and equipment resourcing issues (e.g. having to share the camera with her Dad who took up most of the memory card [video 2], laptop breaking and unable to replace [video 10]). Further, her lack of understanding of the impact of editing (music on in the background which becomes discontinuous post-editing [video 4]) and of copyright (music removed by YouTube [video 7]) is also evident. She is open and apologetic about her lack of expertise ("I'm actually filming on my Canon SLR today, so I mean if this even works I I have no idea what I'm doing […] I think it's in HD I really don't know. I'm sorry" [video 6]) and restricted resources ("Sorry it's taken me a while to make another video but I do finally have a laptop" [video 10]). She also uses her efforts to rectify these issues as a demonstration of labour for her viewers ("Hope you like the quality of this video guys! It was my first time using imovie to edit, so it took me a while to get to grips with everything :)" [video 11, description box].

Later on, these difficulties are often a result of elements that are beyond her control (e.g. "[sigh] never film videos on cloudy days" [video 53]). Or, where they are because of things she can control, she attempts to create self-deprecating humour around them in a "I should know better" kind of way (e.g. "let me just

check my mike's on" [video 48]). These changes in resources and skills evidence her development and investment, financial but also time and effort, in her craft.

Zoella mostly films her videos in the now stereotypical beauty vlogger setting: a bedroom (in parent's house [videos 1, 2, 4, 12, 14, 15, 17 to 26], in her Brighton flat [video 28 to 32], and Brighton house [video 35 to 57, apart from 41 and 51]). However, the exact shot that becomes her convention isn't used until video 35: In the foreground, she is sat on the end of the bed, the camera directly in front of her, brightly lit and we are able to see her from mid-waist upward. She is flanked by two bedside tables at the head of the bed in the background, with decorative items placed on these, and with soft (often fairy) lighting or bright natural light from windows on her right-hand side (see fig 6.4). From then on, deviation from this style is minimal in the dataset (apart from video 36 and 51 being from a different angle, and video 41 being in her living room) and the rest of her videos, with the only variation being the decoration in the background (e.g. the items on the bedside tables, items hung above the bed head, the bedding) changing to reflect seasonality (see fig 6.5).

Zoella was an early pioneer of beauty vlogging so for most of the videos in the microcelebrity dataset there was not yet the expectation of the setting being a bedroom. Thus, she expresses frustration in experimenting with locations ("Today I am filming in a different location, again. Um, I can't find anywhere in my house that I like to film" [video 6]) and the restrictions or qualities that locations imposes on the video ("I look exceptionally pale in this video, but it's just the light washing me out, i'm not ill... haha" [video 9, description box], "Apologies for the light. I have one very bright yellow light up here shining on my face and I have some window light shining here on my chest" [video 13]). Her struggles and attempts to "make do" with what she has, again, communicates labour to her viewers ("oh I'm cutting my head out of screen" [video 3], wobbling on her chair [video 13], difficulties in ensuring larger items are in shot [video 5] and looking in the viewfinder (e.g. [video 2])). Not only does this process of discovering a practical but also visually pleasing set up further evidence her investment in making videos, but consistent use of one shot and location means this element becomes incorporated into her visual branding.
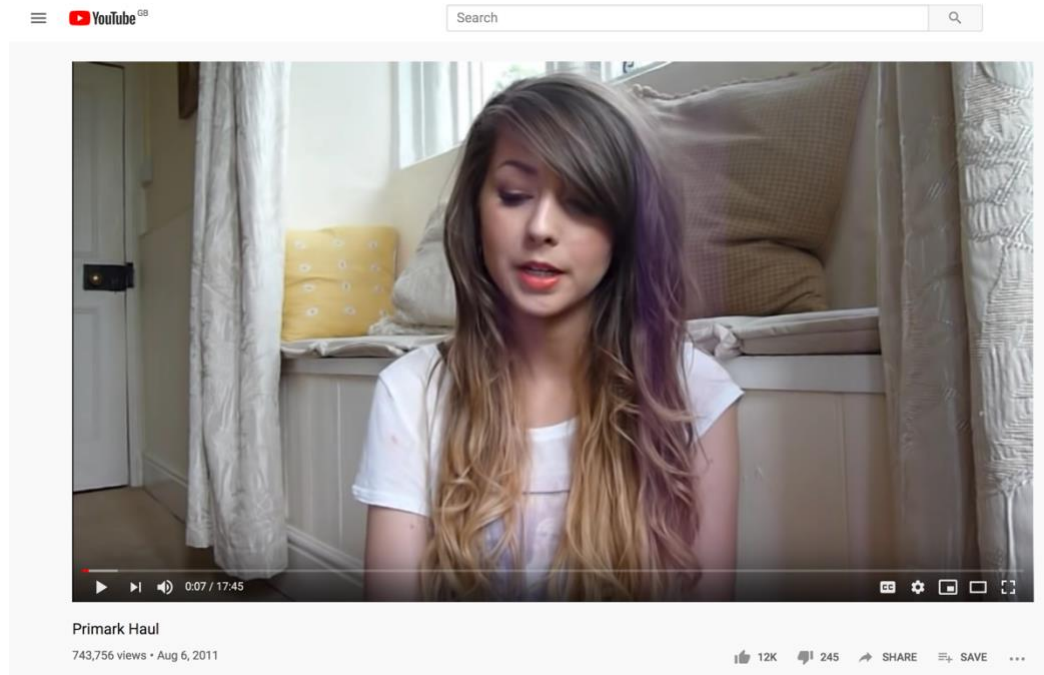
Figure 6.2. Purple streak as a result of light flare [video 5].
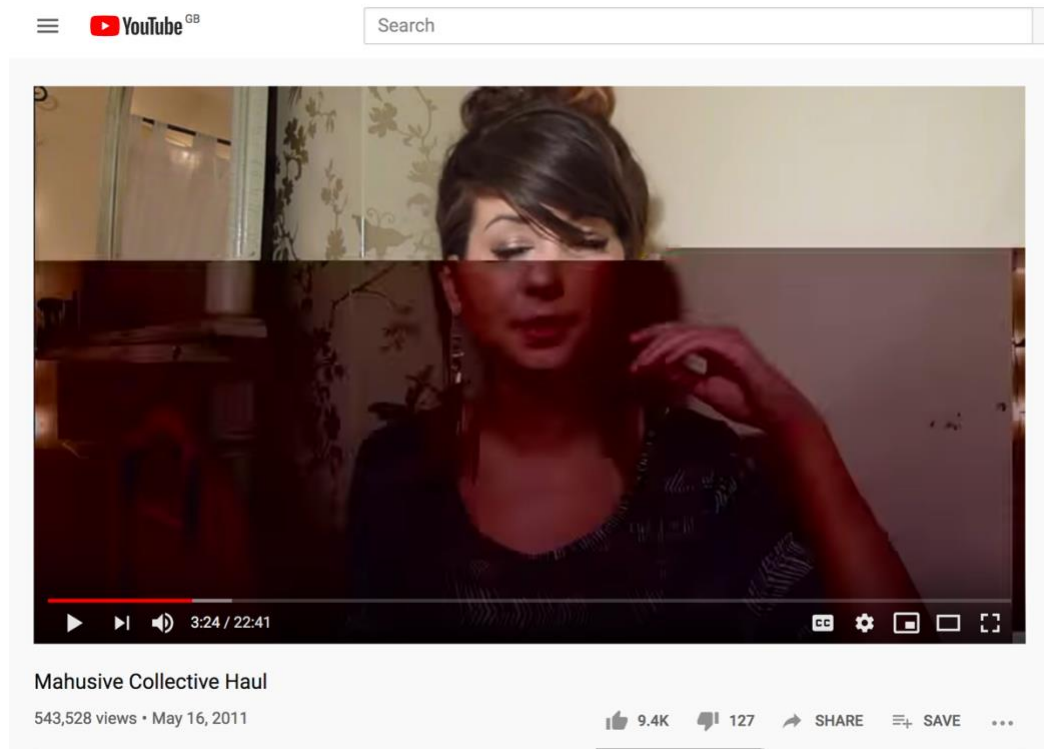
Figure 6.3. Example of a glitch in [video 4].

Figure 6.4. The first video where Zoella's conventional setting and shot is used [video 35].
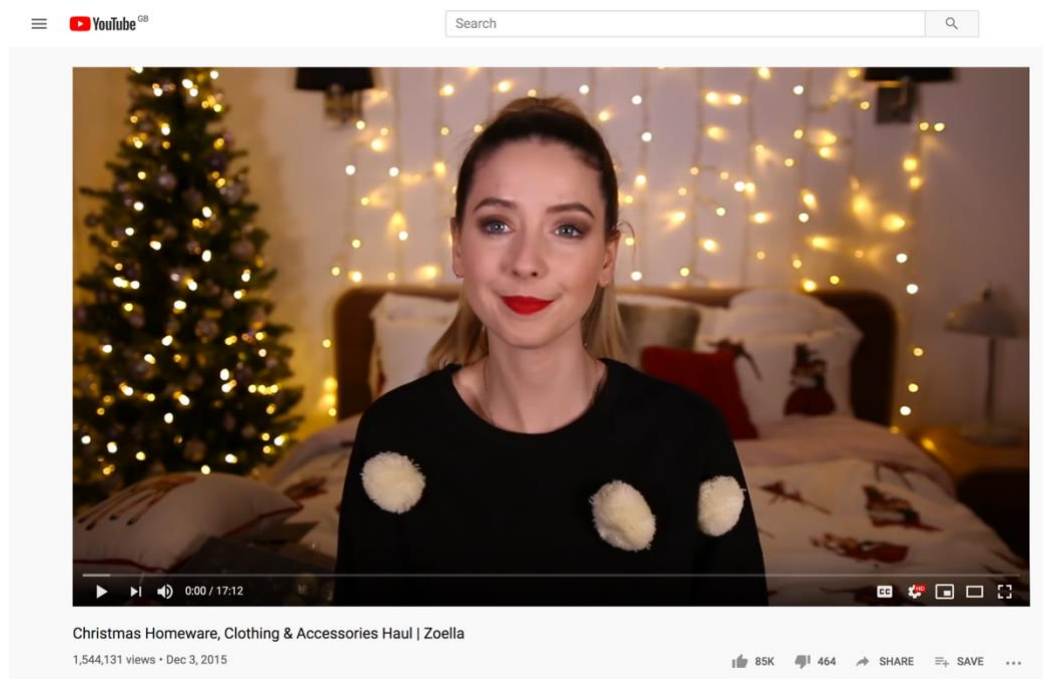
Figure 6.5. Example of Zoella's staple setting changing to reflect the seasons [video 46].

The most notable difference between the Microcelebrity and A List data is the degree and type of post-production used (editing, adding music etc). Zoella's "Vlogmas" videos (where she posts one video per day of December in celebration of Christmas) begin with an animated scene with jingle akin to a theme tune of a television programme (e.g. figure 6.6). While there had been an introductory scene, sometimes with music, on some of Zoella's videos before these were of a much lower quality: a slide, sometimes pixelated, with text that would appear and/or disappear using a standard animation or slide transition (e.g. figure 6.7). It is more difficult to ascertain the degree of change based on the video thumbnails; initially, the thumbnail had to be one of three shots that a YouTube algorithm had randomly selected. Then, when the YouTube Partner scheme was introduced (which was by invite only at first), it was possible for any image to be the thumbnail, as long as it accurately represented the content of the video. Nevertheless, contrasting figures 6.8 and 6.9 illustrates the development in the content and style of the thumbnail image of Zoella's videos.

One of the most consistent aspects of Zoella's videos is the inclusion of bloopers: short clips of mistakes or unplanned happenings, often humorous, that were captured in the process of making the video. These include dropping things ([video 26]), being disturbed by her brother ([video 2, 14 and 15]) or her dogs (e.g. [videos 50, 51 and 53]), hitting herself in the face ("never face palm when you're wearing chunky rings" [video 1]), and in almost every video she leaves in moments where she is stumbling over her words (e.g. "I just think it looks so spring and so, so spring? It just looks so spring!" [video 30], "to have anywhere. Ba. Bluh. make English." [video 41], "Today I am going to be doing a Bath and Body [wɔːks], Body [wɔːks]?" [video 52]) or making exaggerated facial expressions (e.g. [video 25, 35, and 47]). In the Microcelebrity period and a few videos beyond these are collated at the end of the video after she says goodbye, whereas in the A List period they are peppered throughout, keeping them at their original moment in time. This retaining of mistakes and mishaps is designed to give Zoella 'authenticity'. According to Sophie Bishop (Bishop, 2018) authenticity within beauty vlogging can be defined in several ways, the most relevant here being content that is: "apparently un-edited and even un-mediated" (ibid, p. 186).

Figure 6.6. Animated introductory scene for Vlogmas 2015 [video 45].

Figure 6.7. Introductory scene [video 11]
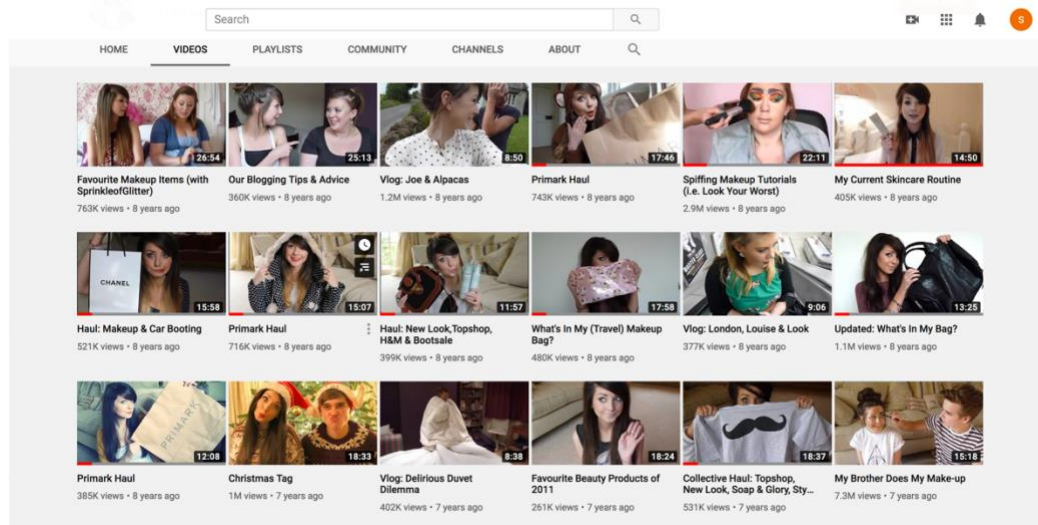
Figure 6.8. Thumbnails for videos uploaded between July 2011 and February 2012. Videos 5 (top row, 4th from left), 6 (1st in middle row), 7 (2nd in middle row), 8 (3rd in middle row), 9 (bottom left), and 10 (5th from left, bottom row) from the haul dataset are depicted (Zoella, 2019).
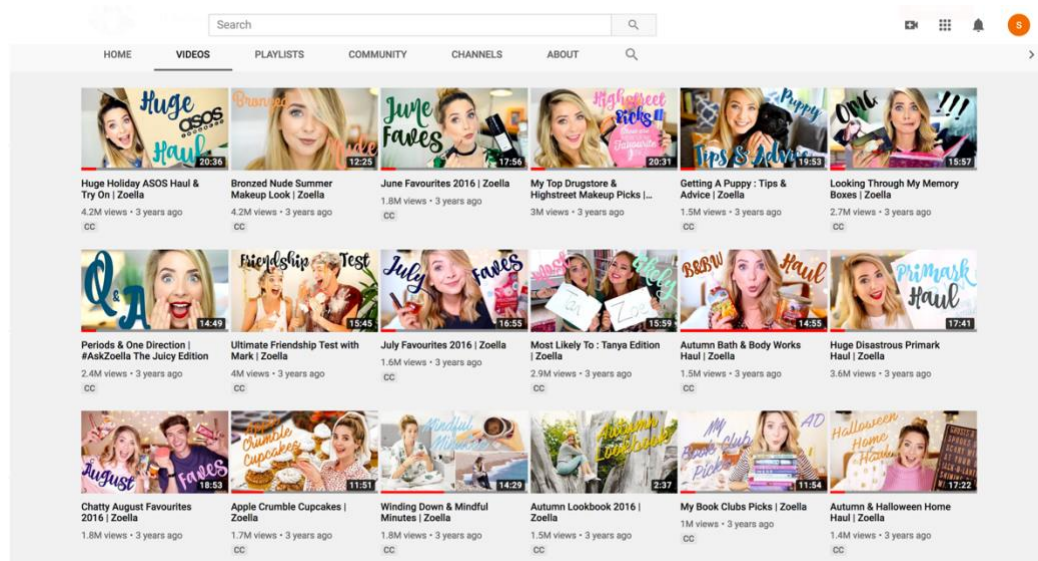
Figure 6.9. Thumbnails for videos uploaded between June 2016 and October 2016. Videos 51 (top right), 52 (middle row, 5th from left), 53 (middle row, 6th from left), and 54 (bottom right) from the haul dataset are depicted (Zoella, 2019).

6.5.2 Brands and their products

A shop that dominates in the Microcelebrity period is Primark; in the first 12 months of the dataset, 7 out of the 10 videos is centred on Primark alone [videos 1, 5, 7, 9] or feature Primark items [videos 2, 4, 11]. So much so, that Zoella jokes: "It's me. The crazy Primark freak, that goes to Primark all the time and is Primark obsessed." [video 2]. Primark is a high street shop that originated in Dublin in the 1969 and has continually expanded, particularly over the last decade. It stocks men's, women's, and children's clothes and accessories ranges as well as homeware and make-up. It is known for its low prices and high pace of stock rotation, making very little profit per item and so is dependent upon sale volume. Thus, it is the "poster child" (Moore, 2019) of the "fast fashion" movement which has now come under criticism for ethical controversies and environmental impact (Butler, 2019; Hinsliff, 2019; Onita, 2019).

It can be argued that Primark featuring less and less is symbolic of her transition into an A List vlogger; as her professionalism, and thus income, increases the featuring of Primark decreases ("It has been a long time since I have done a Primark haul" [video 36], "Today I am gonna be doing a video I haven't done on this channel in so long" [video 53]). From video 18 (which coincides with the beginning of the period defined as "A List" in this dataset), there is a gap of 1 year and 6 months before Primark is mentioned again (in video 30, a collective haul) and of 2 years and 3 months before a Primark haul is produced [video 36]. Zoella says the reason for this is her move to Brighton: "the Primark in Brighton isn't my favourite. Um, I used to live near Bristol and their Primark is my absolute favourite" [video 36], "then when I moved to Brighton and discovered that the Primark in Brighton was not as good as the Primark in Bristol I was most disappointed […] any time I go in there I never come out with anything" [video 53].

As Zoella's YouTube career continues, she found other brands to haul from. Topshop could also be viewed as emblematic of her ascension to A List vlogger with its screen time increasing over time. Topshop is a high street shop and

brand that stocks women and men's (Topman) clothes and accessories. It aspires to be a brand that bridges the high street with designer fashion as exemplified by them regularly showing at London Fashion Week (see Cochrane, 2017) and releasing capsule collections from designers (see Topshop, 2018 for examples). Thus, its fashion prestige and prices are noticeably greater than Primark's. In the Microcelebrity period, video 3 includes some TopShop makeup (which has a lower price point) and a few clothes items are mentioned in video 8, 10 and 16. However, their presentation always comes with a caveat in regard to the cost: "I don't normally shop in Topshop because I go to places like Primark […] um when I do go in Topshop very occasionally I get a little bit like choked on the prices. I'm just like "Oh my gosh, why is this so expensive" " [video 3], "I also bought this which I love but I just think is a bit too expensive and I might take it back" [video 8], "That's quite a lot for a hat actually" [video 16], or the purchases were made using a gift card [video 10].

Apart from Primark Hauls, most early videos include a range of brands, rather than the video being dedicated to the products from one store, as is evident from their titles. For example, "Haul: Topshop, New Look, H&M & Superdrug" [video 3] and "Collective Haul: Topshop, New Look, Soap & Glory, Style Compare, Orange Circle & Vintage" [video 10]. A noticeable diversion from this early norm also coincides with the first sponsored video; "Topshop Haul & £500 giveaway" [video 20]. Soon after two other sponsored videos are posted (with the companies Wantworthy [video 21], and Very [video 23]) all within the first year of signing to the digital talent management, Gleam. This transition is noticed by some commenters, and not responded to positively, however: *I'm a fan of the old Zoella vids, not this video. When you were talking about [company name] it was awkward listening to you cause it's like you had to say [company name] so many times to gain the sponsorship?* and *[company name] just used you to get subscribers*.

One theme that is consistent throughout Zoella's videos is "excess". We regularly see her communicate excess, prior to showing the individual items purchased, through the large bags or boxes that contain her purchases. She describes their size ("I'm sat here with a box big enough for me to curl up in"

[video 52]), makes heaving and straining noises when she picks up bags [video 32], and clearly displays them during the video (see figures 6.10 and 6.11) and in the thumbnail (see figures 6.12 and 6.13). It should be noted that one would expect this somewhat with haul videos as the concept is to display all the purchases that have been made, with fewer purchases leading to a shorter video. However, in the Microcelebrity dataset it is typical of Zoella to mitigate this excess by stating the items had been purchased across multiple shopping trips: "It may seem like I've bought a lot but this is over about a month maybe more" [video 4], "This is a collection of stuff that I've sort of collected over the last month or so" [video 8], "I haven't bought all of this all in one go. This is over numerous amounts of times in Topshop" [video 16]. In comparison, most of the explanations used in the A List period are that this excess was accidental, that she got carried away which resulted in an unintentionally large number of purchases: "Showing you my "accidental-basket-slip" purchases from Boots. ;)" [video 19, description box], "I seem to have indulged a little (A LOT) in bath time treats ;)" [video 26, description box], "I accidentally fell into Boots & they MADE me buy things when I was in there. Jokes." [video 28, description box], "Went shopping didn't I! Ooopsie" [video 31, description box], "I did a little haul. Oops" [video 31]. In one moment, she even attempts to justify her indulgences by comparing them to other, more socially stigmatised addictions: "I like candles a lot, and that's ok, because some people really like gambling and buying candles is isn't like that. You know? Two very different things" [video 42]. In another, she implies this excess allows her to be generous to others: "I am never gonna get through all these and I am aware of that. So, I think some of these will be finding their way into people's stockings this Christmas" [video 56].

While the actual value of Zoella's hauls and the affluency of many of the brands featured increases, applying a strategy for explaining or mitigating abundance is consistent throughout. Sophie Bishop (Bishop, 2018) found a similar need to mitigate or minimise excess in make-up application videos.
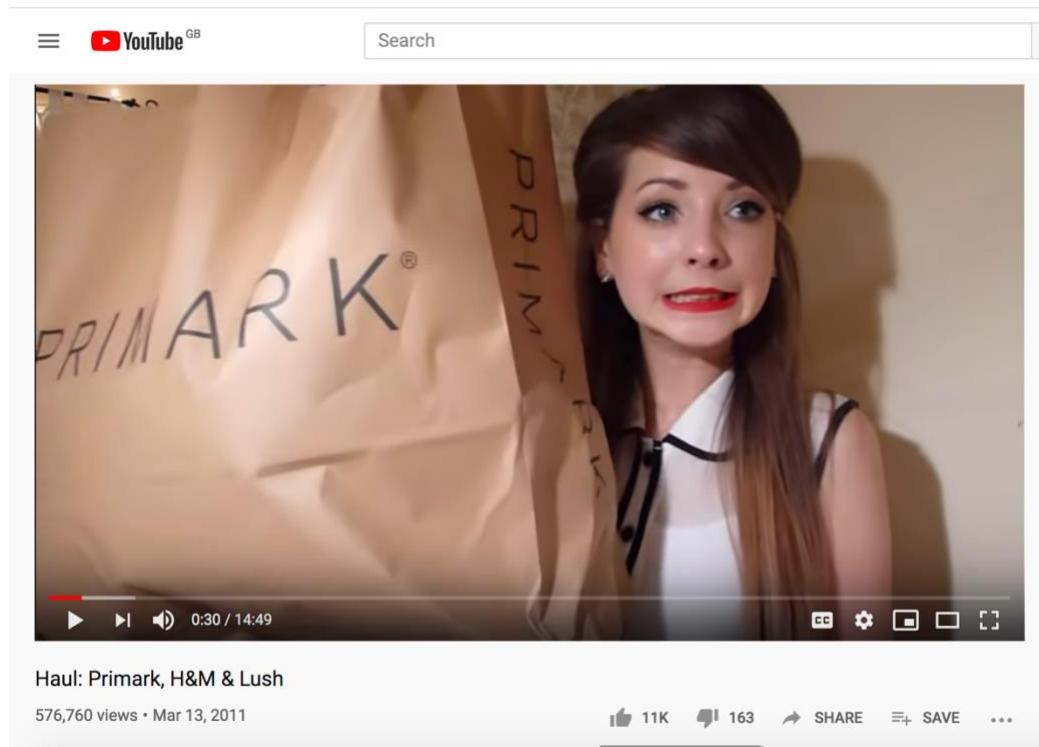
Figure 6.10. Zoella holds up a bag filled with hauled items [video 2]
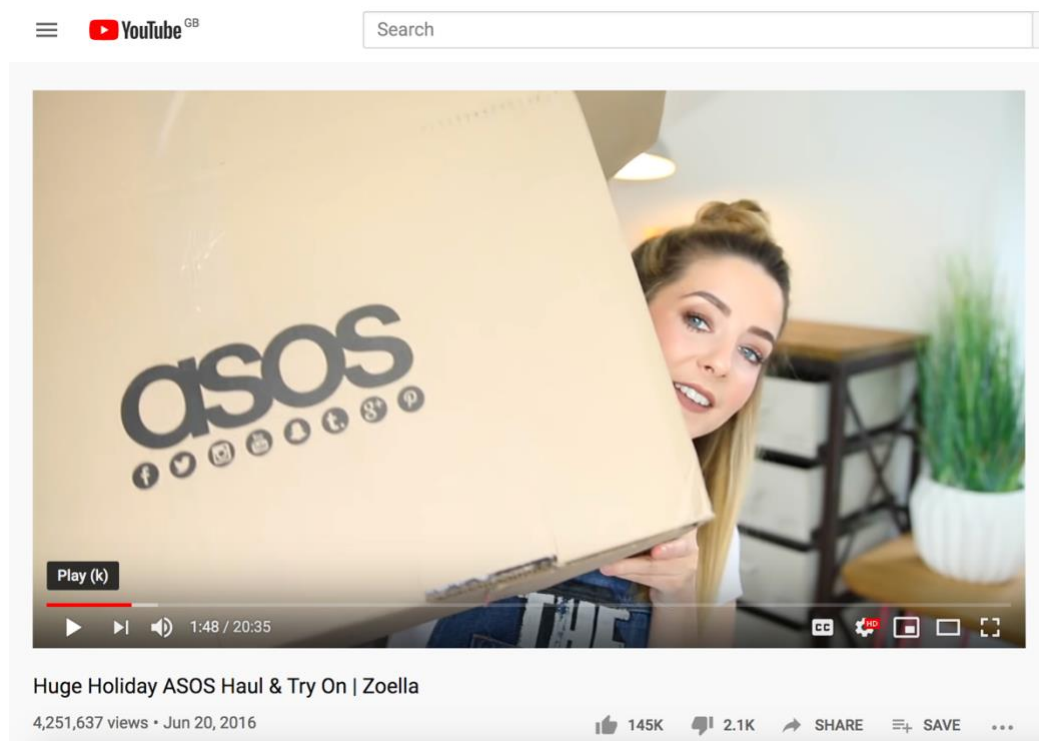
Figure 6.11. Zoella holds up the large box that her purchases were delivered in

[video 51]

Figure 6.12. Thumbnail image for "Huge Winter ASOS Haul" [video 57]
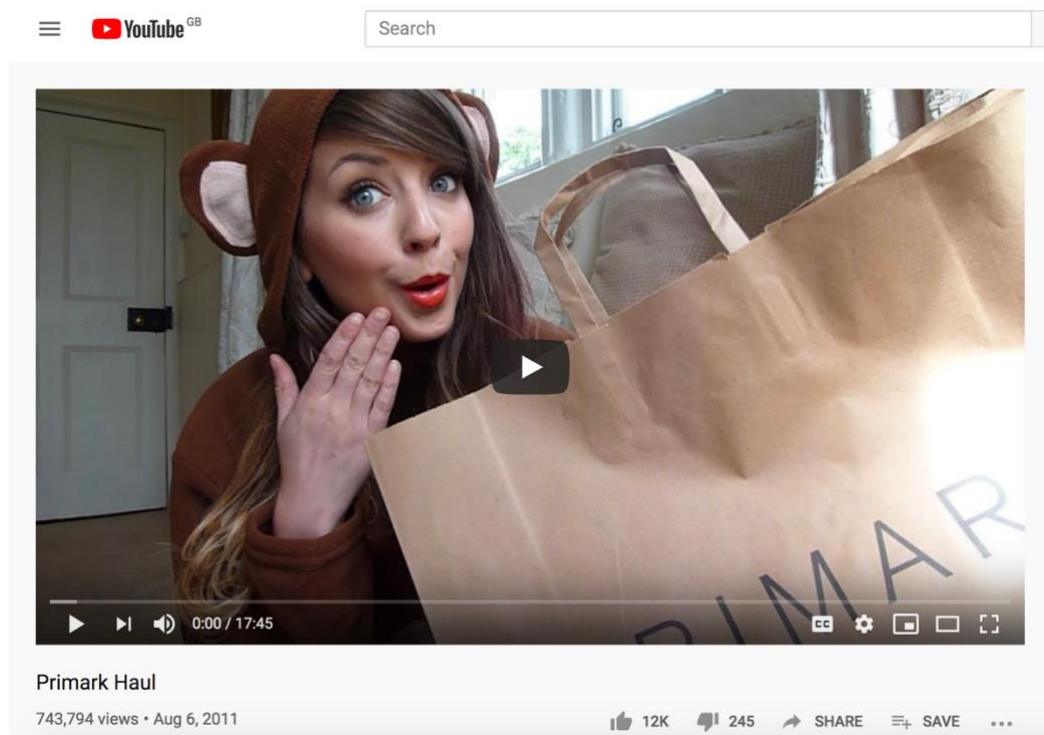
Figure 6.13. Thumbnail image for "Primark Haul" [video 5]

6.5.3 Comments and commenting

During the Microcelebrity period, Zoella regularly responds to her commenters, thanking them for their engagement (e.g. *@user Aww that made me smile :) thanks so much!* and *@user Thanks for the tip – I'll go check it out!*). Mostly, her commenters ask questions. These are often about the products she has shown, such as how and where she wears them (e.g. *@user I've worn them with skirts and tights and I am falling in love with wearing them with cute ankle socks that peep through*) as well as other elements in the video such as how she did her hair and makeup (e.g. *@user It's not extensions, it's a dyeing technique called ombre* and *@user There's a tutorial onmy blog :)*) and even how she created her video (e.g. sourcing music: *@user it just came with [video editing software]*). At others, they can be critical of her practice and in response Zoella defends herself (e.g. *@user I can make whatever videos I like and I like to show the things I've worked hard to be able to buy. That's not advertising. Now buzz off, I'm not interested!*) or is critical back (e.g. *@user at least my videos don't send people to sleep. That's why I have subscribers :)* and *@user what a snob! Where should I be shopping then, Miss Designer?*). However, when such comments are more like constructive criticism or polite requests she responds positively (e.g. when a commenter points out she hasn't responded to a social media request: *@user You're right *slaps wrist* I'm sorry! I'll get to it soon x*). Finally, a minority of comments are even abusive - she retaliates to this abuse (e.g. *@user not sure why you're watching my videos… hmm *pervert alert**) and defends her right to retaliate when this is also criticised (e.g. *@user I think I have the right to retaliate to someone saying they hope I die a horrible death , no?*)[5]

In addition to viewers commenting, in the Microcelebrity period Zoella's YouTube friends also do, notably Barbara Rossi (Rossi, 2020), Tanya Burr (Burr, 2020), and Louise Pentland (Pentland, 2020). These may be in direct response to the content of Zoella's videos (e.g. *I need those belts in my life! And that necklace. Everything you bought is lush! :D xoxo*) or more like messages

---

[5] It is evident that Zoella has reported the abusive comments to YouTube and they have been removed.

(e.g. *Chummy! I LOVE you and I MISS you! Can't wait for our London trip! LOVE xxxxx LOVE xxxxx LOVE xxxxx*). Further, Zoella's friends support her when she receives criticism or abuse. For example, a commenter says that Zoella should not be purchasing products or beauty treatments at a time when others are suffering because of a natural disaster. In addition to Zoella's response (akin to *@user don't see how me turning down a free [beauty treatment] has any effect on [natural disaster]*) Louise Pentland also contributes: *@user Actually, Zoe won and booked this treatment well before [the natural disaster], not that that's your business. If you care so much about [natural disaster], stop watching Youtube hauls and go do something about it!*

However, her engagement with the comments gradually decreases. As was discussed in section 4.3.2, from the start of 2013 there is a notable reduction in Zoella responding to comments on her videos (coinciding with her signing to talent management company, Gleam). There's a further reduction from the start of 2014 and by video 35 in mid 2015 (middle of the A List period) Zoella has ceased responding and does not comment again throughout the rest of the dataset. This change in commenting behaviours aligns with Anne Jerslev's (2016) description of the communicative practices that differentiate microcelebrity and celebrity: Zoella's commenting in regard to temporality transitions from embodying immediacy and instantaneity to scarcity. This lack of engagement and the disappointment it causes her commenters is palpable (e.g. *Please respond – it would make my year!* / *@zoella280390 [Asks question]? Wish you'd reply :(*).

The content of the comments from her viewers also changes over time. First, the descriptions of how her commenters feel about her intensifies. Rather than: *I love you Zoe!* they become: *Zoe, I can't express how much I love you. You make me smile and bring light into my life. I wish we could be friends, sometimes I imagine it. Love you forever xxxxxxxxxxxxx*. Further, the YouTube commenting practice 'Under 301 club' begins to be used. It is believed that YouTube is programmed so that if a new video is receiving a high volume of traffic upon its release a view count of 301 will be displayed until a time when YouTube's software is able to process the actual number of views (PaulApproves, 2011).

Hence, the comment *301 club!* or similar is posted to state the achievement of being one of the first to view the video, the behaviour of a true fan. Other similar commenting practices begin to appear, such as watching the video before a *thumbnail appears (e.g. THUMBNAIL SQUAD!* and *So early there's no thumbnail :p*) being the first or one of the first to comment (e.g. *FIRST* or *almost first :)*) with the cut off value for this achievement varying (e.g. 100, 500, 1,000) as it is being set by the commenter themself.

Finally, comments that are critical of Zoella and her practices, questioning whether Zoella deserves her success, increase in the later portion of the A List period: *No offence, but I just don't get the hype around her. / I know this won't be a popular opinion but Zoe you need to tell us more about the products and be more precise. You just take one out say where you got it form and then say 'I'm looking forward to trying that' or 'that looks interesting'. Like, is it actually any good? Just want to let you know so you can improve. / How come she is so famous? I don't understand. She just sits infront fo the camera and chats. No intro. No conclusion. Not the sort of effort YouTubers like [Youtuber A] and [YouTuber B] put into creating their videos. So, how come she won the [Industry Award]?* With her success increasing, Zoella becomes more vulnerable to criticism.

6.6 Co-creating Place

One intention for conducting the online ethnography was to explore what place may be attributed to Zoella's imaginary audiences and how this is established. In the data a British imaginary audience and an American imaginary audience are the most apparent, their dual prominence emphasised as a result of Britishness and Americaness being continually contrasted. It is reasonable to assume that this contrasting, and indeed the contrasting of Zoella's audiences from places around the world, is prompted by Zoella's Britishness. In regard to contrasting Zoella's Britishness with Americanness, this is most emphasised when she attends the US based YouTube conventions Playlist Live (Sugg, 2013a, 2014b,

2015b) and VidCon (Sugg, 2013b, 2014c). Also, it is at this event that an American YouTuber, Tyler Oakley, coined the term "The British Crew" (2014) to refer to Zoella and her small network of fellow YouTubers that are all managed by Gleam.

While these prominent moments in Zoella's YouTube career maximally antithesise Americanness with Britishness, there are other elements that are more pervasive, present in most or all videos and their comments, and thus are a more continual force in shaping the imagined audiences' place. In addition to her commenters stating where they are from (see 6.6.1 Zoella's Commenters), the discussion of cultural events and practices (6.6.2), vocabulary (6.6.3), Zoella's speech (6.6.4), and brands and products (6.6.5) contribute to the co-creation of place. Note, in discussing these elements in separate sections I am not suggesting that they can be fully untangled or that they do not have influence over each other. It is through considering these other resources that it becomes clear that both Zoella's commenters from the US, and Zoella's awareness of and accommodations for an American audience in her content increases over time.

6.6.1 Zoella's Commenters

The commenters regularly state where they are from. As well as being from the UK (which commenters communicate through stating their home city or region, (e.g. *I'm for the Midlands* and *I'm in north London)* although defining oneself as British rarely happens outside of discussions with other non-British commenters about Britishness), we know that Zoella's commenters are from (in alphabetical order) Argentina, Australia, Austria, Belgium, Brazil, Canada, Chile, Colombia, Cyprus, Egypt, France, Germany, Greece, Guyana, Holland, India, Indonesia, Ireland, Italy, Kuwait, Malaysia, Malta, Mexico, New Zealand, Norway, Panama, Paraguay, the Philippines, Qatar, Romania, Saudi Arabia, Spain, Sweden, Transylvania and United Arab Emirates. Zoella acknowledges awareness of an international audience as early as video 12: "for all of you living international, please send me some sun".

As a result, for a portion of Zoella's audience English is an additional language (e.g. *just been watching videos in my language then watch Zoe and I'm like WTF why is she speaking English? haha*) and state so in their comments (e.g. *Love your video. First time I meet you. I'm [nationality] (sorry, my english not so good and I do mistakes)* and *Your speech is so good! They help me to improve my listening. I live in [country] but soon i'm going to [foreign city] for to learn more english AND YOUR VIDEOS HELP ME SO MUCH! Kisses x*) and a small minority of comments are written in languages that are not English[6]. Some even say that watching Zoella's videos is an educational experience for them (e.g. *I'm studying English and your videos have helped me improve my accent so much*).

The most vocal group of commenters are from the U.S. They complain that they cannot source the items Zoella has hauled (e.g. *Urgh! Why can't we have Bourjois in the US?*) and visit the same shops (e.g. *We need Boots in America!*), or about American products in general (e.g. *American fashion is so boring! But I love the fashion in the UK!*). They compare Zoella to American YouTubers (*[Youtuber name] is an AMERICAN VERSION OF YOU!!!* / *You're like a British [Youtuber name]*), request currency (e.g. *500 pounds is what in the states?*) and sizing (e.g. *Is UK and USA the same?* / *Is a US 2 a 0 or a 2 in the UK?*) conversions, state economic differences (*I want to shop in the UK! The dollar is worth less than the pound yet everything is so cheap*), or defend their spellings to other commenters (*Um, you are wrong, actually. I'm from the states for us it's [spelling]*), in addition to the kinds of contributions described below.

6.6.2 Cultural events and practices

Much of the discussion amongst the commenters involves those that are British explaining or even defending aspects of British culture to (mostly) American commenters. These discussions are triggered by the content of Zoella's videos. For example, when Zoella suggested that some products had been inspired by Bonfire night American commenters asked *Is the 4th of July the American*

---

[6] These were filtered out during the content analysis, and even if they were included it would be inappropriate to attempt to provide a rephrased comment in another language as a monolingual British English speaker.

equivalent of UK Bonfire night? and *I'm from the US, so excuse my ignorance, but can someone explain what bonfire night is please?* Some British commenters helpfully offered explanations, with differing degrees of accuracy, for example:

> *By the way Americans, a Catholic man called Guy Fawkes and a group of his followers tried to kill the King (James 1ˢᵗ maybe?) by blowing up the houses of parliament with gun powder. Their plan failed and so we celebrate by burning home-made Guy Fawkes' on bonfires and having fireworks! Not a detailed explanation but hope you get the jist :)*

Although Zoella did not explain what Bonfire night is, she did acknowledge that this celebration may be UK centric and that some viewers may not know about it, in the video in question:

> "I don't know if this is just something that we celebrate here in the uk. I think it might be because I mentioned it in a previous video and everyone was like 'ya what now? I don't know what that is' "

Some British commenters confirm Zoella's guess, although in a patronising tone: *We only celebrate bonfire night in the uk cause Guy Fawkes didn't try to blow up everyone's parliament in the world! Lol* / *LOL aw little Zoe, of course bonfire night is only celebrated here in the UK. It's about guy folks trying to blow up the houses of parlememt. the houses of parlememt are in London xox.* However, commenters from nations other than Britain that celebrate Guy Fawkes, now or previously, are also keen to stake claim to the occasion (e.g. *[country] celebrates Guy Fox too!* / *We have guy faux in [country]* / *guy fox night (is that how you spell it?) used to be celebrated in [country] when my grandma was a little!).*

Another example is in one video Zoella says:

> "this next thing is also from homesense and it is a little egg tray with hearts. are you an eggs in the fridge person or an eggs out of the fridge person? that's the question that I would like to know the answer to because everyone has a different thing like some people put ketchup in the fridge and some people ke put ketchup in the cupboard"

which results in her commenters comparing egg storing practices: *always eggs out of the fridge :)* / *I'm from the States, so definitely an eggs in the fridge person* / *hang on.. do Brits keep their eggs out of the fridge? Weird... and I though British culture could no longer surprise me!* / *In America, (and maybe other countries?) we have keep eggs in the fridge because of how we clean them. How anyone can keep their eggs OR their ketchup out of the fridge is disgusting to me!* / *When I stayed Europe I found that they keep them on the side in a bowl! US eggs are washed differently so we have to refrigerate ours.*

But the topic of discussion that most reveals the diversity of Zoella's international audience is the stereotypically British topic of the weather. This is prompted by a video that Zoella filmed during an unusually hot summer:

> "I'm gonna try and film this video really quickly because as we speak it's about thirty two degrees Celsius[7] here in the UK and most people in the United Kingdom don't have air conditioning in their houses because this doesn't usually happen and I am currently melting under the light and the heat of my bedroom because I can't open the window because you'll hear the traffic. sorry if by the end of it I have melted"

As is found throughout the dataset, the most comments (at least the most comments that include the location of the commenter), were from American viewers, e.g. *I'm in 90+ degrees fahrenheit with high humidity. Talk about melting. [US state] living...* / *it is 100+ farenheight in [US state] which is about 40+ degrees Celsius.* / *Its about 30+ degrees C* / *88 degrees F in [US State]. Homes don't have air conditioning here either. So I feel your pain!* / *No air conditioning? Wow. To me that's so weird! It's 100 fahrenheit in [US city] today* / *In the [US coast] (where I live) it has gotten up to degrees F which is about 38 degrees Celsius. It's really humid too -.-* / *its 110+ degrees in [US States] :(.* However, the comments left about this topic on this video illustrate the

---

[7] An additional layer of interpretation in regard to these comments is the apparent confusion between Celsius and Fahrenheit for many of her commenters, particularly those from the USA - *32 CELSIUS IS hot! That's like 90 degree FAHRENHEIT. People like: well it's 50 degrees here EVERYDAY! You're wrong. 50 Celsius is like 120 Fahrenheit*

international diversity of Zoella's audience far more clearly than the comments left on her other videos. Comments like *In [country name] it's like [x] degrees today / usually / in the summer* are common on this video, with the countries referred to covering most of those stated in section 6.6.1, from Austria to India, Norway to Qatar, Egypt to the Philippines.

Most of Zoella's commenters feel 32 degrees celsius isn't very hot, and so many of their comments have a condescending tone. Examples include: *Girl, you have no idea! In [country] it sometimes reaches 50 or more degrees. / Haha! UK people. 32 degrees? In [US state] that would be cardigan weather. / when people in Europe complain about coping with 32 degrees without aircon and you have no aircon and it's over 40 – awkward / It's funny you're complaining cause in [US city] its over 36 all year. Yay Murica! Haha not really / yeah that's not "melting" temperature in the states / and the succinct 32? Bitch please.* Here, in addition to the Americans, Australian commenters are very vocal (e.g. *32! Thats nothing in Australia / 32 degrees? Bschhhh In Australia you'd think 32 was cold / Aw poor Brits and their 'heat wave'! haha In Australia 32 is just a nice day!*) so much so this is explicitly addressed by one commenter: *Laughing at all the Aussies like "it's 30 degrees everyday here!!" have you heard of the Middle East?*

In response, British commenters (and possibly those of other nationalities) point out the ignorance of these comments:

> *Guys, understand this: Britain's weather is boring. It's cloudy and grey. So if it gets really hot or cold we fucking freak out. We close roads and schools cause of an inch of snow and in a "heat wave" we ban hose-pipes and put out weather alerts. Why do you think we talk about the weather so much?! Haha!*

/

> *Please shut up with the "omg 32 isn't even hot". It is in the uk, we're not used to it so we're gonna complain! we're used to rain! and we don't care about how hot it is in your country! Haha!*

/

> *In the uk, we haven't had a proper summer in years. We never get hot weather! So we will struggle suddenly going from like 13 degrees to 32*

with no aircon. Shows how much foreigners (americans) know about Britain.

/

Everyone commenting "you think 32 is hot ha" yeah maybe in australia. she lives in the uk so 32 is very hot. Who cares bout you living in the desert in 50 degree heat, she is complaining bout Britain cause it's usually cold.

As one commenter summarises *Jeez. This has turned into a contest for who lives in the hottest place.*

6.6.3 Vocabulary

From engaging with Zoella's videos, while it appears that she makes minimal adjustments to her vocabulary to be inclusive to an American audience or at least an audience that uses American English, these adjustments are salient. These are restricted to using 'fall' as well as 'autumn' ("all their autumn or fall as you say in America" [video 52], "autumn slash fall candle haul" [video 42], "it is autumn or fall as you may call it" [video 25]) and clarifying that one shop is called "TK maxx or TJ maxx if you're in America" [video 27 and 38]. However, it is left to her commenters to explain that the sister store to TJ/TK Maxx is called "Home Sense" in the UK and "Home Goods" in the USA": *Home goods is American home sense.* / *Home sense is the British equivalent of home goods.*

She also uses the word 'drug store' for pharmacy or chemist [video 19]. However, this is challenged by her commenters (e.g. *Zoella, you're British, why are you saying drugstore when it's pharmacy? Just wondering ...* / *"drugstore"? you mean pharmacy, surely. this isn't the USA...* / and *she said drug store but she is english.*). This reaction may be because the title of the video uses the word "drugstore" rather than a word that would be more in keeping with British vocabulary; this is different to other video titles where a similar opportunity to use American terminology is passed for using the British word (e.g. "Autumn & Winter Fashion Haul" [video 25], "Bath & Body Works Autumn Candle Haul" [video 42], "Autumn Bath & Body Works Haul" [video 52], and "Autumn & Halloween Home Haul" [video 54]). It may be that this move is interpreted as

adopting the vocabulary as her own and trying to be more American, rather than merely trying to accommodate an American audience as the phrases "fall as you say in America" [video 52], as you may call it" [video 25], if you're in America" [video 27 and 38].

Her commenters view Zoella's use of British vocabulary endearing (e.g. *the words you have for things and the way you speak is adorable. Love from the USA. / I like this video and your fancy terms for things :)*), as well as seek clarification on many more words, and often find the differing vocabulary humorous. Further, this clarification takes the form of stating their British/American pairings more often than not. Examples are:

i)   jumper/sweater (*We americans call jumpers sweaters so it's really weird to hear you say different.* And *I was like, what the hell is a jumper? until she pulled out the sweater. American problems...*)

ii)   nappies/diapers (*nappies are diapers, for all the americans asking!*)

iii)   playsuit/romper (*A playsuit is a romper in the US*)

iv)   colander/sifter (*I'm America and I know what you call a sifter we call a coldander.* And *Fun fact: collanders are strainers in the states just saying!*)

v)   holiday/vacation (*British people say "go on holiday" for vacation (what we say in America), but what about national holidays? what do them call them?*)

vi)   snoods/infinity scarf (*Thought I'd let you know that snoods are infinity scarves in the states*)

vii)   batwings/Dolman sleeves (*Bat wings? Ha ha! We call them Dolman Sleeves in the US*)

viii) fringe / bangs (*Dear not-British peeps, a fringe is just bangs lol*)

ix)   joggers / pyjama pants, and

x)   dungarees / overalls (*Joggers and dungarees instead of pajama pants and overall. Love British lingo! So funny compared to the American*)

xi)   wellington boots / rainboots (*those rainboots are so cute! When i visited england I found funny that you call them wellies lol you english are so cute.*)

Further, words that differ in their meaning across the US and the UK, often with humorous consequences, are also discussed. For example: *lol dork is American slang for dick* / *where I'm from (America) dork means penis* / *haha homely is a synonym for ugly in the states,* and *Haha just wanna let you knoew that 'flashed' means something completely different in american english ... :)* . Commenters also request and supply clarification on words that are alien to some. Sometimes the location of the commenter is unknown (e.g. *I need help with the British words! What's high street? Expensive or affordable brands?* and *What's a gap year? (I've heard British people talk about it)*) and when the location of the queries is known they are almost always from American viewers (e.g. *For those in the US, [brand name] is a medicine for children in the UK.* and *Hi, what's [food item A] and how is it different from [food item B]? I'm american and we just have [food item B]*).

The vocabulary that triggers the most discussion is 'chucky pig'. This is prompted by Zoella trying to describe a section of quilting on a pair of pleather trousers:

> "they've got […] like what I like to call um not armadillo yeah Chucky pig what do you guys call Chucky pigs? because here in the UK I think we all call them the same thing it's those little like woodlouse okay maybe we don't all call them Chucky pigs maybe that's just me and they also have the Chucky pig part here which is just like the lower part of the trousers"

A lot of discussion in the comments ensues, with many contributing the alternative terms 'pill bug' and 'rollie pollies' / 'roly polys' and other spelling variations (e.g. *I think what she is talking about is a pill bug?? I call them rolly pollys!*. *In the states they're called pill bugs...* *I'm not sure if they are what you're talking about but they're "woodlice" technically and we call them rollie-pollies in America.* *Rolie polly's! That's what American's call them but I like you're name better* and *Chucky Pigs = Rollie Pollies for americans*). However, rather than just national vocabulary differences coming to light so do regional ones: *Haha chucky pigs is deffo just you... For the rest of britain they're Woodlice!* *Im from the UK too and i had no idea what a chucky pig was*, and *My teacher calls them chukky pigs!! I think it's a west-country thing?*).

6.6.4 Zoella's speech

Zoella is very regularly praised for the way she speaks with her commenters expressing both a love and jealousy for the way she speaks (e.g. *I love the way you talk / I wish I had your accent*), as well as saying things like *I'll watch every video, even if they're boring, just to hear you speak.* Her accent is described as; *adorable, amazing, awesome, beautiful, the best, brilliant, charming, cool, cute, eloquent, fabulous, gorgeous, hot, lovely, nice, perfect, precious, sexy, sophisticated, stunning,* and *sweet.* Very, very, few commenters disagree with this positive sentiment. However, her speech is described as *weird* by one commenter, another says they *can't stand* her accent, and another says that they mocked her accent while watching. A few comments indicate difficulties in understanding her (e.g. *what was that last thing she said? I couldn't understand it.* and *that bit where she's talking about boots, and socks and a what? I don't get what she said!*), complain about her pace (e.g. *why do you talk so fast? is it on purpose?* and *omg you're talking so slow and dragging out your sentences in this video.*) and a couple question why she is speaking how she is (e.g. *why is she talking like that? Her viewers are not 4 year olds!* and *she should act her age. she's talking to us like we're kids*). Her accent is even described as one that *doesn't exist naturally*.

Zoella's speech is, of course, a reflection of where she is from. Her commenters often refer to "Britishness", aligning her speech with British celebrities (e.g. the singer Cher Lloyd, and actress Emma Watson), fictional characters (e.g. Alice in Wonderland, Arya from Game of Thrones) and other British YouTubers (e.g. charlieissocoollike). Most expressions of aspiration seem to come from American viewers who explicitly contrast British and American accents (e.g. *I'm gonna go uni in the UK to try to catch your accent rather than my [US state] one. wish i had a british accent. american english sucks.* and *I'm from the states but I love your accent more than mine!*). The suggestion that she *do a video in an American accent. That would be so funny!* because *Americans do british accents for fun all the time (I'm sure we don't do them properly haha!) Do you guys do the same?* provides further evidence of the saliency of these ways of speaking.

Further, her British accent is seen as prestigious not just on an international scale (e.g. British vs American) but a regional one: *you speak proper queen's English. I love it :)* / *Your English is better than half the British population...* / *I love your nice Southern Accent.* / *I've heard lots of other british accents before but yours stands out* / *you speak English properly, not like Youtubers who speak like commoners!* / *i like your accent more than the average british accent*. However, it is not clear to all commenters that Zoella has a British accent (e.g. *hey are you american or english??* / *I didn't realise u was british I thought u was from American haha* / *am I the only one suprosed by your accent? You look so American!*) or what kind of British accent (e.g. *from where in England is her accent? I'm pretty sure it's south England... I'm guessing Paddington, London???* / *I love your accent are you from Leeds? :)*). This is a rarity, but a reminder that a variety of language and accent experience, and therefore indexical fields, will be overlapping within Zoella's viewers.

Zoella's commenters' interest in her speech is exemplified by their picking up on the pronunciation of specific words (e.g. *the way she says top shop – I love it!* / *I love the way she prononses 'chocolate'* / *OMG i love your accent 'VITIMIN' haha* / *PrinTTT lol i wish i had your accent <3* / *The way you say "again". Omg, I want your accent Zoey! :D* / *her accent is so strong when she says "crop tops" haha love it* / *holm-sense! wish I had your accent* / *how you say jaguar why dont i have a cool accent* / *she pronounced hooray as "hoo-ra". I love zoe! I love her British accent*) and even specific sounds (e.g. *your Ts are so pronounced ha.* / *I almost always miss my t's at the end of words. It sounds horrible! But your voice is so clear*). Sometimes the commenters explain that they are entertained by her pronunciation by referring to their own: *I'm from the states so love how you say garage! haha so cool!* / *zeeebra! That is how americans say it, lol :) I love your accent <3* / *Being from the states, the way she says "massage" is sooo funny* / *In Wales we say "Preemark" but like that you say Primark.* / *british/american people say jaguar differently? I had no idea!* / *Tutti Frutti in your british accent is so cute <3 love from Italy.* /. *I'm from Newcastle and hate the Geordie accent – I want yours so much!* /. *I feel like I dont have an accent cause I'm from the States but I probably do have one to you. Sorry for rambling! Short story is I love your accent.*

Further, a few commenters home in on the regional aspects of her accent: *you get more west country when you speak to your brother. / your accents so prominent when you said 'answer it!' at the end hahaha / your west country accent really comes through when you say leggings – love it!*

6.6.4.1 H&M

In several hauls, Zoella purchases items from H&M (a high street clothing and accessories store (H&M, 2020)) and her pronunciation of this brand name is greatly discussed amongst her commenters. As she explains:

> "a lot of you don't like the way I say [heɪtʃ]. I'm not the only one that says it like this though. I thought I was weird because I was saying [heɪtʃ] but everybody around me says [heɪtʃ] too so maybe it's just a cultural thing? But for the benefit of everybody I'll say "I went to [heɪtʃ] and M and I went to [eɪtʃ] and M"[8]

Examples of the comments that she is referring to include *I'm used to saying "Eich and M" so "Hei-ch and M" sounds so weird to me* and *Haytch & M? Never heard it like that before,* but for some commenters the idea that there are different ways of pronouncing "H" is new to them (e.g. *I say "h" the same as you...isn't that normal? Lol*). Some align with Zoella's pronunciation (e.g. *i say heych too its a "h" sound not a "a" sound. / I like how u say H cause I say it the same / you say "H" correctly, like me! x*), many others suggest that their pronunciation is correct and not Zoella's (e.g. *it's pronounced ay-ch actually :D / 'H' isn't pronounce heytch is eych! If that makes sense? / you can say it how you're comfortable with but I think it's ACH and EM*), and some suggest the pronunciation is interchangeable (e.g. *Don't worry, Zoey. For me sometimes it comes out as "Haych" and m or "aych" and em lol! anyone else do this?*).
In regard to the reason for differing pronunciations, some commenters state that it is accent related (e.g. *how you pronunce 'H' depends on where you live and the accent of English you speak, dunno if someone has said that already.* and *It's not*

---

[8] To provide an explanation that is not reliant on the International Phonetic Alphabet: Say the words "hate" and "ate" and think about the difference at the start. You should be producing an extra sound in "hate". Zoella is talking about the same difference in the two ways that the word that refers to the letter "H" can be pronounced.

cultural. Its about the accent of the people in the region where you live :D) and thus this discussion emphasises Zoella's British-Englishness: *Why do brits say "hay-ch and m" for H&M lol, so funny / [other British Youtuber] say 'H' the same as you :D maybe it's just how british people are most comfortable saying it? / she says "haitch & em" because of her accent! Other British gurus say it like that too. So telling her off for saying it like that is just being rude.* Further, in the vast majority of cases commenters contrast her British pronunciation with that found in the US (e.g. @*user hay-ch-uh is right, americans say ay-ch-uh / I say 'H' how you do. its just how english people say it as I keep seeing Americans saying it the other way <3 / I think that a lot of British people pronounce the letter H as "haych". In the USA we say " aych " / i'm from the states and i had never heard anyone pronounce it "heych" til now - we just say it differently in the States! anyway, love your accent! / I never knew that Americans pronounced H as 'aiych'. I assumed everyone pronounced it as 'haich'.*) and discussion of "H" varying across other British English accents is minimal (e.g. *I'm for the Midlands and I say "H" not "ACH"! / I say H like aitch not haytch and I'm in north London*), emphasising the international rather than national.

However, this internationalism is minimally inclusive of Sweden – where H&M originated from, and thus provides an indication of how it should be pronounced. *To be picky, it's a Swedish store so it is not pronounced either of those ways, but say it however you like! :D / In Sweden we don't even say "H AND M", we just say HM lol so I find it so funny that you get told off!*

6.6.5 Brands and products

Initially, the majority of items Zoella included in her hauls had been sourced by shopping near her parent's home in Wiltshire (e.g. [videos 3 and 11]) or online (e.g. [videos 14 and 16]), with purchases made when visiting London [video 2] being a novelty. With a focus on the high street, it would be possible for viewers to go and purchase the same items themselves. One of her justifications for her habit of Primark Haul videos was "since you can't all see what stock Primark have in their shop online because they don't have a website it only seems right

that any time I go into the shop and buy some things that I share them with you" [video 18]). Notably, some items were purchased from ambiguous locations and would be near impossible for viewers to source if they wanted one the same. For example, Zoella makes purchases from car boots [videos 6, 8 and 13] and house clearance shops [video 10], promoting them as great places to find bargains. She also wears items in her videos that she didn't get from the high street and itemises them in the video's description box (e.g. "Earrings - Random Boutique (sorry)" [video 1]). The inclusion of such locations and items suggests a national focus in the early days.

The brands in Zoella's videos increasingly represent internationalism, reaching a pinnacle at video 42 and again at videos 52 and 56. Here, she presents hauls of "Bath and Bodyworks" products; a brand that does not sell in or ship to the UK. In her first Bath and Bodyworks Haul her explanation for how she sources these products is: "my management Gleam actually have an office in LA […] So, um, I got my order delivered there to the Gleam office and they then forwarded it onto me" [video 42]. Obviously, few of her viewers would be able to replicate this strategy emphasising a disparity between her and them. Then in a later Bath and Bodyworks Haul she hires a company to forward packages from their address in the USA to her "which you can do. There are multiple ones that you can choose you just have to kind of search it and find a good one" [video 56]. The choice of brand in these three videos indicates a shift towards creating content that is intended for American viewers, even at the potential expense of 'home' British viewers. Further, in one of these videos, she repositions herself as less British, not necessarily more US, but certainly blurs 'here' being the UK: "I am sorry if you cannot get hold of Bath and Bodywork or if you're watching this in the UK" [video 56].

Videos that focus on shopping in the USA do exist before the Bath and BodyWorks hauls, however they are very much framed as "holiday shopping" with souvenir purchases from tourist attractions (e.g. Universal Studios, Disney World [video 22]), products that are different to those in the UK ("in the UK these are a lot more expensive than they are in the US" [video 22]) or are not available or are from stores that are not in the UK ("I went in here because we do

not have a sephora here. Do we? Do we now? Do we have one now? I don't think we do" [video 35], "this is very difficult to find in the UK" and "I think this is an american thing because i haven't seen it in the UK" [both video 22]). Her American viewers find this entertaining: *it makes me laugh how she's surprised with the flavored gum when it's so normal to us Americans / Every time you were like "we don't have these" I was like uh yes we do. then I remember that I live in the US hahaha.*

Throughout the dataset her US audience is salient because Zoella is not able to fully cater to them and their location. This is evident in American commenters regularly expressing disappointment that they cannot shop in the same places as Zoella (e.g. *dammit, i wish I lived in Britain, i want to go shopping in new look / I want Boots to come to the US! / Is Primark the British version of Forever21?*) or purchase the same products easily (e.g. *Oh why don't we have Bourjois makeup in America :(/ Omg I love the cardigan! US doesn't have primark though / I love the white blouse but very doesn't ship to the states*) including Zoella's own lines of products (e.g. *why isn't zoe beauty in the states? sad face / Will Zoella beauty be coming to America soon*?). So, they take the opportunity to express joy when Zoella covets products that they can easily source (e.g. *I'm in America so have pop tarts all around me, but I hardly ever eat them! haha / You talking about [cereal] is funny because I'm from America & I eat them all the time lol*). This leads many commenters to suggest an exchanges (*I'll send you American pop tarts if you send me [sweets] / How you feel about [cereal] is how I feel about the Primark tights. I need to become a Youtuber, get English fans, and get them to send them to me! / Can you Brits trade your Primark for our American sweets?*) and request that Zoella accommodates to them more (e.g. *Can u haul in stores that are in American too? I love your taste in clothes but I live in the US and we don't have those stores here / Do a British makeup giveaway for people that don't live in the uk! We can't get sleek or berry m products in America*). In response, one of Zoella's commenters argues that *you should expect British products when you watch a British YouTuber.*
The incongruity of Zoella's location with a portion of her audience is most evident in the comments on four videos, all of which are sponsored and involve some sort of giveaway (Topshop Haul & £500 Giveaway [video 20], Huge

Collective Haul & Giveaway [video 21], Very Haul & Giveaway [video 23], and Boohoo Haul & £500 Giveaway [video 29]). In all four instances, one must reside in the UK to be able to enter the giveaway contest. This causes a lot of confusion and disappointment amongst her US commenters: *What if I live in the States? / Damn I'm in America :(/ Does the giveaway work worldwide? Like, for Americans? / It's only open to people living in the UK. Being American sucks sometimes :'(/ I can't enter cause there are no topshops in the US / do a give away in dollars – just for us Americans! / I went to the very giveaway but it was for the UK only. Us Americans can shop online too ;)* .

6.7 Reflecting on Thesis Research Questions

In this chapter, progress has been made in answering thesis research question 3: "What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?". A review of literature across multiple disciplines has built a cohesive argument for using ethnographic methods online in order to guide sociolinguistics research practice. In this case study, through ethnography a clear direction and rationale for speech variables that could be investigated was generated. Further, while it is argued that a key strength of ethnography is how its range of methods can be applied reflexively in response to the particular context under study, in order to conduct an analysis where the insights can usefully contribute to the research question ethnography was applied somewhat atypically in regard to 'where' and 'when', although some precedent has been set by Florencia García-Rapp (2019). Thus, in addition to confirming that ethnography can be used as a strategy for identifying speech variables, some knowledge of how ethnography can be adapted in order to respond to an online context, where spatiality and temporality are unhelpfully complicated, has also been gained.

6.8 Summary

This chapter has reported on the findings from online ethnography observations in regard to celebrification and the co-creation of the imagined audiences' place by Zoella and her commenters through a variety of resources. As is evidenced throughout the data reported in this chapter, many elements of Zoella's videos and the comments they receive change over time as she transitions from microcelebrity to A List vlogger. Notably, the most apparent imagined audiences' place (Britain and American) is relatively stable throughout. However, there is a change in the relative amplitude of these two contrasting imagined audiences' with 'the Americans' becoming increasingly salient in the comments and Zoella's acknowledgement of an American audience also increasing. This provides clear direction for which speech feature to choose to be studied next, which is defined and explored in the next chapter.

Chapter 7.

Word medial /t/

This chapter reports on the collection and analysis of the second of the two dependent (speech) variables that are investigated in the Zoella case study. The ethnographic observations in the previous chapter have indicated the place of Zoella's imagined audience is constructed by contrasting Britishness and Americanness, and so I reason that the speech feature to investigate should be one that does this also. This variable is referred to herein as word medial /t/. It is not possible to understand the realisation of word medial /t/ without understanding phonology and phonetics and so these concepts will be explained first (7.1.1). The three main allophones of word medial /t/ (the ways in which word medial /t/ may be pronounced, i.e. [t], [ʔ], and [ɾ]) will be stated along with a summary literature review of their prevalence in American English and British English (7.1.2) and their potential positive and negative connotations (7.1.3). These literature reviews helped define the research questions and hypotheses as detailed in section 7.2. Then, the data collection and analysis methods will be reported (section 7.3) as well as the results and the study's limitations (7.4). The statistical analysis indicates that Zoella's comment engagement and her status as an amateur YouTuber moderate the effect the comments have upon her speech. The more Zoella sees what the comments say, the more influence they have over how she speaks, but only when she is an amateur and not when she is a professional. After a discussion of the findings (7.5), the relevant thesis question will be reflected upon (7.6), and the chapter summarised (section 7.7).

This chapter addresses **thesis question 4**, 'What statistical approaches could be used when considering the variable time in sociolinguistic studies of online public video?'.

7.1 Word medial /t/

Word medial /t/ was selected as the dependent variable for this study. First, foundational knowledge of phonology will be introduced (section 7.1.1). Then the three accents of importance to this study (General American, Received Pronunciation and Southern Standard British) will be introduced (7.1.2.1) and their use of [t], [ʔ], and [ɾ] described (7.1.2.2). Finally, the social connotations attached to each of these speech sounds will be discussed (7.1.2.3). It is here that the rationale for selecting this variable, as a result of the ethnographic observations reported in the previous chapter, will be revealed.

7.1.1 Phonology, Phonemes and Allophones

Phonology is the study of speech sound systems (Ogden, 2017). Each language has their own phonological system of how speech sounds pattern and function. First, the phonological system defines what speech sounds are valid to use. For example, in Xhosa, one of the official languages of South Africa, click consonants are valid speech sounds[9] but they are not to be used in British English (Ladefoged, 2005). Further, speech sounds that are valid may only be so at certain syllable or word positions. So, while the [ŋ] sound is allowed at the end of syllables in British English (e.g. "sing", "walking") it is not allowed at the start (Ladefoged, 2005). A phonological system also dictates that a speech sound can or cannot be used with certain types of speech sounds preceding or proceeding. For example, in British English one cannot put a /s/ before /b/ /d/ or /g/ at the beginning of a word but one can put /s/ before /p/ (e.g. 'speak'), /t/ (e.g. 'stop') and /k/ (e.g. 'skill'). These factors - syllable position, word position,

---

[9] See (Ladefoged, 2005) supplementary material:
http://www.phonetics.ucla.edu/course/chapter6/xhosa/xhosa.html

preceding and proceeding speech sounds - make up what is referred to as a phonological context. So, a phonological system comprises that language's valid speech sounds and each speech sound's valid and invalid phonological contexts. Another important aspect of a phonological system is which of the valid speech sounds lead to a change in meaning and which do not. The term 'phoneme' is used to refer to speech sounds that can lead to a change in meaning and are indicated using '/ /'. Take the syllable 'ba'. Adding different speech sounds on the end gives you different English words, and thus different meanings; adding a /p/ for 'bap', /t/ for 'bat', /d/ for 'bad', /k/ for 'back', /g/ for 'bag', /n/ for 'ban', /ŋ/ for 'bang', /θ/ for 'bath', and /ʃ/ for 'bash'. Thus, /p/, /t/, /d/, /k/, /g/, /n/, /ŋ/, /θ/ and /ʃ/ are different phonemes in English. Each phoneme in a phonological system will have allophones – a number of speech sounds that if produced in that phoneme's place will change how the word sounds but not what it means. These are indicated using '[ ]'. Take the [x] sound in Scouse. A Scouser may pronounce a <k> at the end of a word (e.g. "back" and "dock"[10]) as [x] (Watson, 2007). This is produced by raising the back of the tongue to the roof of the mouth as one would for [k] but not letting them touch. The air being pushed through this narrow gap makes a hiss-like noise. Returning to the concept of allophones, regardless of whether 'back' is pronounced [bak] or [bax] and 'dock' is pronounced [dɒk] or [dɒx] the word's meaning is the same: 'back' is still a direction or place, and "dock" is still a place where boats moor. Sociolinguistic research often operationalises a phoneme as the speech variable and its allophones as the variants of the variable.

While the use of a phoneme's allophones does not alter linguistic meaning, it can be indicative of social information. As was explained in section 2.6, a person's speech is not always consistent; the same person may speak differently in different situations. One way in which intraspeaker variation can come about is through the use of allophones. For instance, a speaker's use of allophones can reflect how formal or informal they perceive the conversation to be, and thus

---

[10] See "37-back.wav" and "38-dock.wav" in the "Consonants" folder of the (Watson, 2007) supplementary materials: https://www.cambridge.org/core/journals/journal-of-the-international-phonetic-association/article/liverpool-english/992DEF3999B0F2F20952870B188A77A5#fndtn-supplementary-materials

their speech also indicates this perception to their listener. This is probably a change in the relative use of allophones rather than switching categorically from one allophone to another upon the situation changing.

The phoneme /t/, when word medial, has two main allophonic realisations in Zoella's accent of British English ([t] and [ʔ]) and its stereotypical realisation in certain word medial contexts in American English is [ɾ].

7.1.2 /t/ in American English and British English

Here, the three accents of importance to this study are introduced (7.1.2.1), their use of [t], [ʔ], and [ɾ] described (7.1.2.2) and the social connotations attached to each of these speech sounds is discussed (7.1.2.3).

7.1.2.1 General American, Received Pronunciation, and Standard Southern British English

Guided by the observation that the content of British Zoella's YouTube videos increasingly accommodates an American audience over time, the speech feature to examine must be one that maximally contrasts American English and British English. Of course, just like in the UK, America has a diversity of accents. However, if taking the perspective of a non-expert Brit, and thus someone with little awareness of such accent diversity, the accent to refer to when thinking of North America would be General American (GA). As Wells (1982, p. 118) explains "'General American' is a term that has been applied to the two-thirds of the American population who do not have a recognizably local accent".

For the UK, one could argue the equivalent would be the national standard (a concept that was unpicked in section 2.1). At the time of Well's writing this was Received Pronunciation, also known as RP: "RP is associated with England, though not with any particular locality within England" (Wells, 1982, p. 117). It should be noted that since Well's writing RP has become less prestigious and aspirational. The social revolutions of the latter half of the twentieth century that

saw fights for greater economic equality (e.g. the Equal Pay Act of 1970) and self-government (decolonisation), and the superiority of the upper classes undermined partly as a result of greater press freedoms (e.g. political scandals such as the Profumo Affair) saw RP's shine dull. In other words, "the social foundations on which RP stood collapsed" (Lindsey, 2019). Now, "RP speakers are perceived, as soon as they start speaking, as haughty and unfriendly by non-RP speakers" (Trudgill, 2000, p. 195).

With RP on the wane, Wells (1982, p. 118) predicted that

> "by the end of the [20th] century […] some new non-localizable but more democratic standard may have arisen from the ashes of RP: if so, it seems likely to be based on popular London English".

This seems to have come to fruition in the form of 'Standard Southern British English' (SSBE), described as the "modern equivalent" of RP by the International Phonetic Association (1999).

British English listeners are likely to describe RP or SSBE speakers as people who do not have accents, or at least they cannot be sure where the speaker is from by their accent. But American listeners, who are unlikely to be able to differentiate between the two, would view RP / SSBE as a typical British accent (see Wells, 1982). The same can be said for GA: American listeners would take the view that these speakers do not have an accent, but British English listeners would think that they are speaking a typical American accent.

7.1.2.2 Phonological context of 'word-medial /t/'

When considering the phonological systems of GA, SSBE and RP, it became apparent that the realisation of the /t/ phoneme when in specific phonological contexts would be an appropriate variable to investigate. Herein, this phonological context will be referred to as 'word-medial' for ease but requires a finer definition. A speech sound being 'word-medial' means it is in the middle of

a word. However, many other speech sounds can flank the word-medial one creating many different phonological contexts.

The consonant /t/ can be, for example:

- Word-medial pre-consonantal – after a vowel and before certain consonants either within the same syllable (e.g. 'bats') or at the end of a syllable in a multi-syllabic word (e.g. 'witness')

- Word-medial post-consonant - after certain consonants and before a vowel, e.g. 'shelter', 'winter'

- Word-medial pre-syllabic nasal – after a vowel and before a nasal (e.g. /m/ /n/) that effectively replaces the following vowel, e.g. 'bottom', 'button'.

- Word-medial pre-syllabic /l/ - after a vowel and before a /l/ that effectively replaces the following vowel, e.g. 'bottle'.

- Word-medial intervocalic – after a vowel and before another vowel in stressed (e.g. 'butter') or unstressed (e.g. 'guitar') position

This study of /t/ collates the latter 3 phonological contexts, i) Word-medial pre-syllabic nasal, ii) Word-medial pre-syllabic /l/ and iii) Word-medial intervocalic. These three phonological contexts can be associated via all having a preceding vowel and a proceeding sonorant. A sonorant is a speech sound that is continuous and predominantly made by the voicing of the vocal folds. That is to say that sonorants are not made by full or partial obstruction in the vocal tract to create a closure and release (plosive) or friction (fricative), respectively. As is mentioned above, the sonorants syllabic /l/ and syllabic nasals (/m/ /n/ and /ŋ/) behave like vowels in that they take up a whole syllable. However, words that may end in a syllabic /l/ /m/ /n/ or /ŋ/ can also be pronounced as the respective consonant with an unstressed vowel, or with a full vowel (Ogden, 2017). To put it another way, just like the phoneme /t/ has the allophones [t], [ʔ], and [ɾ], the syllabic consonant phonemes have the allophones i) syllabic consonant, ii) unstressed vowel plus consonant, and iii) full vowel plus consonant (see table 7.1 for examples).

Table 7.1. Summary of the allophones of the syllabic consonant phonemes in word-final in British English (conventions C̩ = syllabic consonant, v = unstressed vowel, V = full vowel, C = consonant).

Word	Phoneme	Allophones		
		C̩	vC	VC
bottle	Syllabic /l/	[bɒtl̩ʸ]	[bɒtəlʸ]	[bɒtʊlʸ]
bottom	Syllabic /m/	[bɒtⁿm̩]	[bɒtəm]	[bɒtʌm]
button	Syllabic /n/	[bʌtⁿn̩]	[bʌtən]	[bʌtʌn]

Therefore, when the full vowel plus consonant allophone is used, the context for the preceding /t/ becomes word-medial intervocalic. Thus, by collating these three phonological contexts this continuum of pronunciations is accounted for. Further rationale for focusing on the context 'vowel-/t/-sonorant' will be given below.

7.1.2.3 Realisation of word medial /t/

Wells states that:

> "One of the most striking characteristics of American pronunciation to the ears of a non-American is the intervocalic consonant in words such as *atom*, *better*, *waiting*. To English people it sounds like /d/ rather than /t/. Phonetically it is usually a rapid tap rather than a more deliberate plosive; it is also frequently voiced." (Wells, 1982, p. 248)

This realisation of intervocalic /t/ as [ɾ] can be referred to as T-voicing, and T-tapping when the /t/ is not voiced and [ɾ̥] is produced. T-voicing/T-tapping can also occur under other conditions most notably when a vowel precedes and a syllabic /l/ follows, such as in the words 'bottle' or 'little', but not before syllabic nasals (Wells, 1982).

While it is phonologically acceptable for /t/ to be pronounced as [ɾ] in the British Accents RP and SSBE, this is rare (Lindsey, 2019), and other allophones as preferred (Tollfree, 1999) defying Wells's (1982, p. 250) prediction that T-Voicing will be "the first distinctly American phonetic innovation likely to

spread in time to all accents of English". In RP intervocalic /t/ is mostly realised as [t]. Indeed, Eckert (2008, p. 468) describes /t/ release (an emphatic [t] so that a small puff of air is produced upon releasing the hold phase, transcribed as [tʰ]) is a common resource for Americans imitating British English.

Another realisation of intervocalic /t/ is [ʔ], this process being referred to as T-glottaling. The realisation of /t/ as [ʔ] is one of the most well researched variables in UK sociolinguistics. With the apparent growth of T-glottaling among younger speakers across the UK (e.g. Cardiff (Mees and Collins, 1999), Derby (Docherty and Foulkes, 1999), Sheffield (Stoddart, Upton and Widdowson, 1999), West Midlands (Mathisen, 1999), Glasgow (Stuart-Smith, 1999), Edinburgh (Schleef, 2013)) it is thought to be "one of the most dramatic, wide-spread and rapid changes to have occurred in British English in recent times" (Trudgill, 1999, p. 136).

Where research has considered intervocalic /t/ the results are mostly consistent. Carmen Llamas' (2007) study in Middlesbrough, Jennifer Smith and Sophie Holmes-Elliott's (2017) study in Buckie, northeast Scotland, and Hazel Richards' (2008) study in Morley, Leeds, all found that their youngest speakers almost categorically used [ʔ] for intervocalic /t/, as did Williams and Kerswill (1999) among their working class speakers in Hull, Reading and Milton Keynes. Further, an interesting class and gender pattern was found in Hull and Reading (Williams and Kerswill, 1999) and Sandwell, West Midlands (Mathisen, 1999): middle-class women using [ʔ] intervocalically more than middle class men, although its use was still a minority compared to [t]. A similar gender pattern was also found in children in Newcastle and interpreted as boys adhering to local norms and girls aspiring to supra-local norms (Milroy *et al.*, 1994).

Previous descriptions have identified a key difference between RP and "popular speech in the south-east of England", which we will take to mean SSBE, is intervocalic T-glottaling (Wells, 1982, p. 253) with more recent studies of London speakers agreeing with this observation. Laura Tollfree's (1999) study recruited participants from a number of suburbs in south-east London. She categorised each speaker's speech as either: i) "South East London English"

which she describes as "medially to maximally broad varieties", 'broad' indicating a locally distinctive way of speaking, and ii) "South East London Regional Standard", "the local form of near-RP" (ibid, p. 164). She found that the near-RP speakers rarely T-glottaled intervocalically, whereas there was a high incidence in this position in the broader accent group. Also, Altendorf's (1999) study of the speech of public-school and comprehensive school children in London found that in both word-medial intervocalic and word-medial pre-syllabic /l/ contexts /t/ was almost categorically realised as [t] by public school children but comprehensive school children almost categorically used [ʔ] over 50% of the time. This isn't to say that T-glottaling is not present in RP at all. It has been observed in RP in word-final position but the consensus is that it is unlikely intervocalically (Fabricius, 2000; Lindsey, 2019). Finally, the only phonological context where T-glottaling is found in GA English is word-medially before a syllabic nasal (Wells, 1982).

To summarise, in a word-medial intervocalic phonological context [ɾ] dominates in GA, with [t] less likely and [ʔ] being rarely used. [t] is the typical variant for RP, with use of [ɾ] being rare, and while [ʔ] has made some inroads overall, it is unlikely intervocalically. [ʔ] is increasingly used by young speakers in many accents in Britain, with it generally being agreed that one differentiation between RP and SSBE is use of glottals intervocalically in conversation. [t] and [ɾ] are also possible in SSBE with [ɾ] being the least likely. Table 7.2 summarises how likely each of these realisations for intervocalic /t/ are in the three accents in question.

Table 7.2 A summary of the relative use of the allophones [ɾ], [t] and [ʔ] for word-medial intervocalic /t/ across GA, RP, and SSBE.

	[ɾ]	**[t]**	**[ʔ]**
General American (GA)	Most likely	Less likely	Rare
Received Pronunciation (RP)	Rare	Most likely	Less likely
Standard Southern British (SSBE)	Rare	Less likely	Most likely

As is evident from the literature review, use of [ɾ] for /t/ in some word-medial contexts maximally contrasts GA and RP and SSBE. This study is considering whether Zoella's speech is influenced by her audience, particularly as her American audience grows, and so her use of [ɾ] may evidence this influence. Equally, her use of the allophones [t] and [ʔ] may indicate her audience influences her speech in different ways, as will be outlined in section 7.2. To satisfy the principle of accountability (see section 3.3), the phonological contexts for /t/ that are studied herein must permit all three allophones to be used. After surveying Zoella's speech, I estimate her accent to be SSBE, a phonological system that permits [ɾ], [t] and [ʔ] in word-medial pre-syllabic nasal, word-medial pre-syllabic /l/ and word-medial intervocalic phonological contexts. However, it should be acknowledged that if Zoella uses [ɾ] in her speech, and this use is influenced by her growing American audience, then she may use [ɾ] as she imagines an American speaker would. While there is alignment across GA, RP and SSBE in regard to the allophones of /t/ in word-medial intervocalic and word-medial pre-syllabic /l/ contexts, the phonological context of word-medial pre-syllabic nasal adds a complication. While [t], [ʔ], and [ɾ] are all permitted in RP and SSBE, [ɾ] is not used in this context in GA (Wells, 1982). Equally, it would be inappropriate to assume that Zoella has in-depth insight into the phonological system of GA. Rather, I predict that the similarity of the other two word-medial contexts across GA, RP and SSBE will encourage Zoella to apply the same usage permissions to the three allophones of /t/ in word-medial pre-syllabic nasal.

7.1.3 Connotations of /t/

Associations between speech features and social identities are indirect and mediated by social qualities and stances that are activated within interaction and so are underspecified, multiple and mutable (Eckert, 2008). Therefore, the same allophone can carry both positive and negative connotations. Just like what particular social qualities and stances are indexed, whether a speech feature indexes positive or negative social qualities and stances depends on the listener.

Herein, the likely social qualities and stances that each allophone for /t/ can index will be reviewed for both American and British listeners.

7.1.3.1 Americans' view of [t] and [ɾ]

As detailed in section 7.1.2.3, [ɾ] is the dominant realisation of /t/ in GA English, with [t] being less likely and [ʔ] being rare. Regardless, for American listeners [t] is associated with distinct social identities, carries certain social qualities, and can be used in interaction to take specific stances. One notable social identity that Americans associate with [t] is the British. As mentioned above, Penelope Eckert (2008, p. 468) describes [tʰ] as a common resource for Americans imitating British English. She also points to three other social identities that Americans associate with [t]: nerd girls (Bucholtz, 2011), Orthodox Jewish boys (Benor, 2001), and gay men (Podesva, Roberts and Campbell-Kibler, 2001). The social qualities that mediate the association between [t] and these social identities is intelligence and being educated. Mary Bucholtz (Bucholtz, 2011) found that the nerd girls' use of [t] was to give the impression of being independent thinkers rather than conforming to the conventional educational content at school. Sarah Benor (2001) found that the boys that used [t] most at an Orthodox Jewish school were those who had formally studied the Talmud (a religious text) in a Yeshiva (an educational method and system specific to the Jewish religion). Podesva (2007) found that a medical doctor called Heath used [t] more often in clinical work settings than personal, home settings. The social qualities of intelligence and education naturally links [t] to other social qualities also, such as competence and professionalism, and being articulate (Eckert, 2008), demonstrating how a linguistic feature may have multiple related indexical meanings at the same time. [t] can also carry negative social qualities so the speaker may be viewed as artificial, prissy, or effeminate (Wells, 1982, p. 250). Finally, to understand the indexical field of a speech feature one must also understand its alternatives, is aptly illustrated with Eckert's (2008, p. 468) comment that:

> "[t]he contrast between the flapped intervocalic /t/ of the United States and the released /t/ of British English further evokes stereotypes of the

British as cultured, refined, and articulate, and Americans as anti-intellectual and loutish."

7.1.3.2 British view of [ɾ]

In a similar vain to [t] and GA, because the dominant realisation of /t/ in two British English accents is [t] (RP) and [ʔ] (SSBE) and [ɾ] is used rarely in either, British speakers associate [ɾ] with the social identity of 'American'. Just like Americans use [t] to impersonate Brits, it is likely that British speakers would use [ɾ] when imitating Americans (as Stuart-Smith (1999) found one of her speakers did). To date, the social qualities that British speakers associate with Americans, and thus [ɾ] indirectly, has not be systematically investigated. Based upon my own intuitions as a native British speaker, Eckert's (2008, p. 468) observation that [ɾ] indexes anti-intellectualism and loutishness for American speakers holds for British speakers also. Other negative and positive social qualities that I think may mediate [ɾ] and American social identity for British listeners are materialism and egocentricity, power and confidence, respectively.

7.1.3.3 British view of [t]

The positive social qualities that [t] carries from a British listener's perspective also align with those from an American listener's perspective. However, while for American listeners the social qualities of [t] connect to a general 'British' social identity, British listeners would connect them to an upper-class social identity (as evidenced in (Alderton, 2019), because of the dominance of RP amongst these speakers and [t] being characteristic of RP. Use of [t] is viewed as educated (Alderton, 2020) and professional (Kirkham and Moore, 2016) as well as articulate, reliable, and posh (Erik, 2014). Equally, this connection between [t], RP and the upper class can also index negative social qualities also. As stated in 7.1.3.1, RP was once prestigious and aspirational but is now disliked by many (Trudgill, 2000; Lindsey, 2019). There is a lack of systematic research on what negative social qualities [t] may carry for some speakers. However, by combining Lindsey and Trudgill's comments with my intuition as a native British English speaker and the understanding of the positive connotations of [t]

in the literature, a reasonable estimation would be that [t] could also be seen as posh, snobby or arrogant.

7.1.3.4 British view of [ʔ]

Historically, [ʔ] has been highly stigmatised and "widely regarded as ugly and also a lazy sound" (Wells, 1982, p. 35) and negative social associations still exist today, primarily that its speakers are 'uneducated' (Alderton, 2019). The greatest evidence of [ʔ]'s stigmatisation is its lack of use amongst older generations (Tollfree, 1999; Llamas, 2007; Richards, 2008a; Smith and Holmes-Elliott, 2017)) as well as media commentary (see (Kirkham and Moore, 2016; Alderton, 2020) for examples). It is not difficult to argue that the primary reason for these negative connotations is the link between use of [ʔ] and the working class (evidenced in (Erik, 2014; Alderton, 2019) findings), via accents such as SSBE.

However, [ʔ] is increasingly used by young speakers in many accents in Britain with evidence that it can be perceived as indicating solidarity, friendliness, youthfulness, and trendiness (Kirkham and Moore, 2016) as well as being more casual and more down-to-earth (Erik, 2014). Now, as T-glottaling has increased in younger generations and it even begins to be heard in RP, some would argue it has begun to lose its stigma (e.g. (Fabricius, 2000)). Of course, how acceptable T-glottaling is somewhat depends on the phonetic context it is used in, and intervocalic T-glottaling remains the most stigmatised relative to other syllable and word positions (Fabricius, 2000; Lindsey, 2019). To summarise,

> "[t]he intermediate status of /t/-glottaling in RP speech may help to explain the extreme reactions to the perceived use of this form in the speech of individuals in the public domain" (Kirkham and Moore, 2016, p. 90).

7.1.3.5 Summary

To summarise, [t], [ʔ], and [ɾ] can all index positive and negative social qualities for both American and British listeners. Further, the literature indicates considerable similarity in how American and British listeners perceive these

three allophones, namely that [t] indexes being educated, and [ʔ] and [ɾ] index being uneducated, and presumably social qualities that align with these also (e.g. competent and professional, and incompetent and unprofessional, respectively).

7.2 Case Study Research question and hypotheses

The case study research question is:

> Does the direct written feedback received through the commenting function influence a YouTuber's speech?

And, as has already been stated in section 4.3, the commenters describe Zoella as 'cute' less over time, with a noticeable decline in the first three years and a near plateau for the following three years. In addition to this, we now know from the digital ethnography that place is created by continually contrasting Britishness and Americanness. Also, there appears to be an increase over time in the portion of her audience that is American and that Zoella makes adjustments to cater for them. Thus, there are multiple imagined audiences that Zoella could be responding to through her speech, and these potentially change over time.

In light of this, the speech variable word medial /t/ was selected. In comparison to the previous speech variable that I studied, which operationalised the variable as uptalk or not, there are more variants, and each one can carry multiple indexical meanings. Further, these indexical meanings are rooted in two different cultural perspectives (British or American) and even a third in that Zoella's vision of an American's perspective may not align with reality. These multiple indexical meanings and cultural perspectives create a web where it is not possible to untangle a specific hypothesis in regard to which variant will index youthful femininity.

However, in regard to youthful femininity and the possible indexical rationales behind the variants of word medial /t/, I posit two suggestions. The first is that [ɾ] could index youthful femininity. [ɾ] is the typical realisation of word-medial trochaic /t/ in General American English and its use in British English (both RP and SSBE) is rare, therefore one would assume that increasing use of [ɾ] would reflect increasingly engaging with an American audience. Connecting this to youthful femininity, I would argue that an American audience would view a British YouTuber catering to them through her speech as endearing. The second suggestion is that [t] could index youthful femininity. As is detailed in section 7.1.4.1, Americans associate [t] with Britishness, and one cultural stereotype is that the Americans find the British-English endearing, as well prissy. And so, use of [t] may be viewed as cute by her American audience as it emphasises Zoella's British-Englishness.

7.3 Word medial /t/ data collection and analysis

First, potential tokens were identified by reading the orthographic transcripts. Read and dog-directed speech had already been marked in these transcripts and tokens within these types of speech were not considered. The words with the word medial phonological contexts of a preceding vowel and either i) a vowel, ii) syllabic /l/ or iii) syllabic nasal (/m/ /n/ or /ŋ/) after were noted. These potential tokens were then reviewed while listening to the audio and the realisation of the word-medial /t/ coded. The use of auditory analysis follows most other sociolinguistic studies on T-glottaling (e.g. Fabricius, 2002; Llamas, 2007; Kirkham and Moore, 2016; Smith and Holmes-Elliott, 2017). Tokens where the /t/ was omitted were excluded, for example several tokens of "little" were produced as one syllable ("lil" [lɪɫ]) and all productions of "battery" omitted the middle syllable so its pronunciation was akin to "batry" e.g [batɹi] or [baʔɹi]. A small number of tokens that were ambiguous were discarded. The reasons for difficulties in distinguishing the token auditorily included echo, background

noise such as traffic outside, foreground noise such as rustling from opening packages and the showing of items, and music overlaid post filming.

Each token was listened to and its realisation coded as either [t], [ʔ] or [ɾ] based on the criteria below.

- [t] – an audible stop closure and release including aspirated [tʰ], affricated [tˢ], nasally released [tⁿ] and laterally released [tˡ] realisations.
- [ʔ] – total replacement with an auditorily single glottal stop [ʔ] or a period of creaky voice. Tokens where both the preceding and proceeding syllables were fully pronounced with a creaky voice were discarded as these were deemed to be creaky voice rather than glottal realisation.
- [ɾ] - a voiced tapped / flapped realisation

This coding system and the descriptions above give the impression that there are hard delineations between [t], [ʔ] and [ɾ], but actually there are two continua: [t] to [ʔ] and [t] to [ɾ]. The first is a continuum of glottal /t/ realisations. Realisation can range from [t] to its replacement with a single [ʔ] or, effectively, multiple [ʔ] to produce a short period of creaky voice as mentioned in section 7.1.2. I set a clear guideline to delineate tokens where a period of creaky voice was a glottal realisation (partial creak allowed on both the preceding and proceeding syllables or full creak allowed on one of these syllables) from those where the creaky voice is present across the entire preceding and proceeding syllables. Further, it is possible to produce a word with a medial [t] with a creaky voice. Such tokens were discarded. The second is the continuum [t] to [ɾ]. An intermediary between these two realisations is an unvoiced, flapped/tapped alveolar ([ɾ̥]). Such tokens were also discarded. However, in both cases these were very, very rare.

After this process of token identification, inclusion and exclusion, and first round of coding, there were 2,385 tokens in the dataset. Then, after a period of several months, I recoded the entire dataset. During this second round of coding a further 73 tokens (3.1%) were discarded for various reasons, mainly that upon a second analysis I was not confident in the initial coding but uncertain of how to recode the token, and finding that several tokens where the /t/ was in word-

internal foot initial position meaning the preceding vowel was unstressed and the /t/ was stressed (e.g. *tattoo*). At the same time, I removed tokens that were brand names or names of products (86, 3.6%), for reason that will be explained in section 7.5. This resulted in a dataset of 2,226, of which I changed the code for 41 tokens (1.8%) indicating an overall percentage agreement of 98.2% with the first round of coding.

7.4 Results

First, an overview of how word medial /t/ was realised will be given. This will identify that [ɾ] is used rarely and so it is discarded (section 7.4.1). Use of [t] and [ʔ] in relation to the comments is then explored. First, their correlations are examined to test if they are associated and then the potential for a cause-effect relationship to be present is tested using simple linear regression (both are reported in section 7.4.2). Next, section 7.4.3 is dedicated to moderation analysis. First, an overview of this analytical approach is given from both conceptual and statistical standpoints, and time is tested as a moderator (7.4.3.1). Time is then broken down into two other variables: i) Zoella's comment engagement, and ii) Status (whether Zoella is an amateur or professional YouTuber) (7.4.3.2). With evidence that both of these variables moderate the effect of the comments on speech, section 7.4.3.3 explores their combined effect. The statistical analysis finds evidence to support a model of additive multiple moderation. R Studio (RStudio Inc., 2019) was used for the calculations and visualisations in 7.4.1 and 7.4.2. The models reported in 7.4.3 were constructed using the PROCESS Macro (Hayes, 2020) in SPSS (IBM Corp, 2019) and the visualisations were created by transferring data provided by SPSS into R Studio.

7.4.1 Realisation of word-medial /t/

The analysis yielded 2,226 tokens of word-medial /t/, a mean of 39 tokens per video. The number of realisations of /t/ as [ɾ] was negligible (65 tokens, 2.9%).

Therefore, this variant was discarded from the analysis at this point. This left a final dataset of 2,161 tokens, a mean of 39 tokens per video. An overview of the dataset can be found in table 7.3. Removing [ɾ] leaves the word-medial /t/ variable as binary. Therefore, the results of one will be mirrored in the other (e.g. negative values becoming positive). Rather than give a full report of all results, in-text descriptions will focus on [t] although some graphs, tables, and annotated screenshots of results from SPSS for [ʔ] are included.

Table 7.3. Number of and percentage of word medial trochaic /t/

Realisation	N (percentage %)
[t]	1,439 (66.6%)
[ʔ]	722 (33.4%)

Figure 7.1 is a line graph of the percentage of [t] and [ʔ] tokens per video across time. There is great variation in their use from 2011 to 2013. At times the majority of the pronunciations are [t], at other times it's [ʔ], and there are times when they are relatively equal also. However, the overall trend is that [t] increases and [ʔ] decreases (see figure 7.2). Then there is a change in the pattern from early 2013 where [t] becomes the preferred pronunciation.

7.4.2 Correlation and Simple Linear Regression

Figures 7.3 and 7.4 plot the youthful-femininity comment data against the /t/ data. Upon visual inspection there appears to be some co-ordination between the comments and the two variants of /t/. The overall trend is that there is an initial decline in the comments referring to youthful femininity which then plateaus and remains relatively stable, an overall trajectory that is opposed by the [t] data. The stabilising of the comment data is later than in the speech variants; from 2014 onward rather than 2013 (as is clearest in figure 7.4). Although there is a large fluctuation in the comments in the last few videos of the dataset (figure 7.3) this is not noticeable in the line of best fit (figure 7.4).

After establishing that [t] and [ʔ] are not normally distributed (both W = 0.926, *p* <0.01), their correlations with the comments were assessed. A statistically

significant negative correlation was found for [t]. This suggests that there is an association between Zoella's realisation of [t] and the youthful femininity comments. However, while this correlation highlights an association between [t] and the comments this statistical test does not indicate the direction of or what kind of relationship that might be. Thus, linear regression was performed. A statistically significant result was found suggesting that a causal relationship may be present. See table 7.4 for details of the results.

Table 7.4. Results of correlation and linear regression tests

	[t]	**[ʔ]**
Spearman's rank correlation coefficient	S = 443 *p* <0.01 rho = -0.436	S = 174 *p* <0.01 rho = 0.436
Linear regression	coeff = -2.38 *p*<0.01 t = -3.99 Adjusted R^2 = 0.21	coeff = 2.38 *p* <0.01 t = 3.99 Adjusted R^2 = 0.21
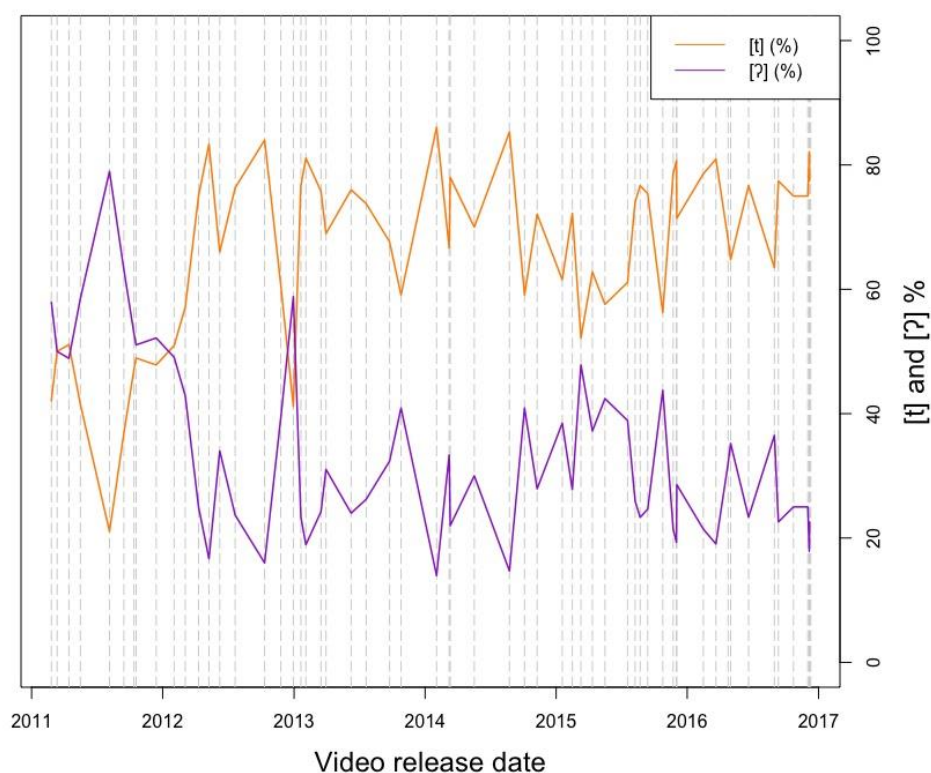
Figure 7.1. Percentage of [t] and [ʔ] in each video
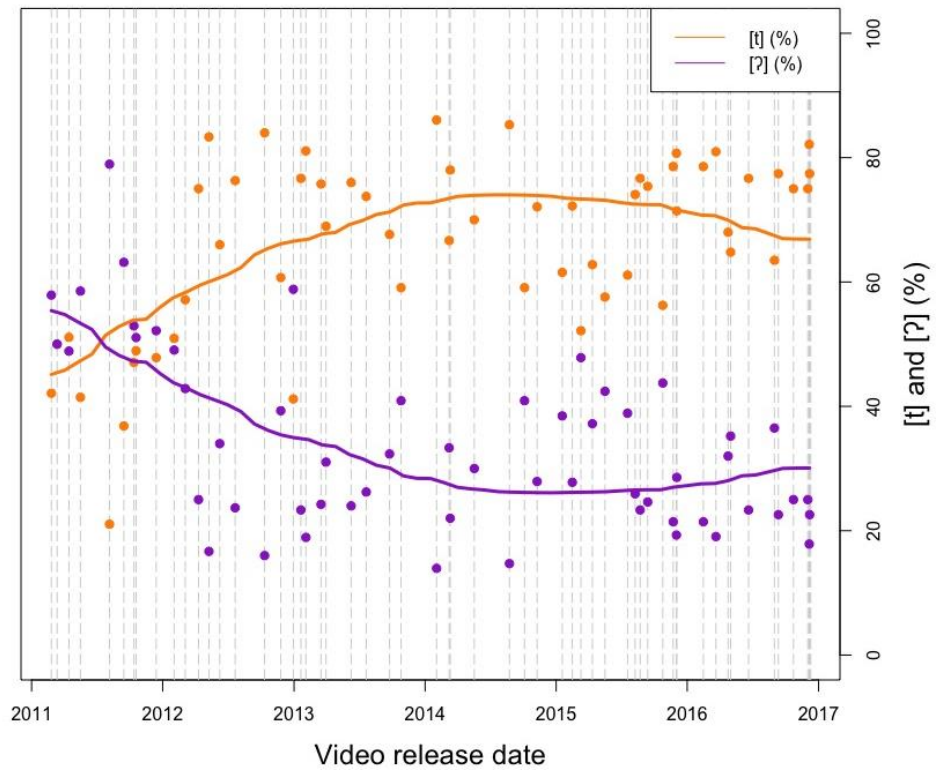
Figure 7.2. [t] and [?] lines of best fit (fitted with 2 polynomials)
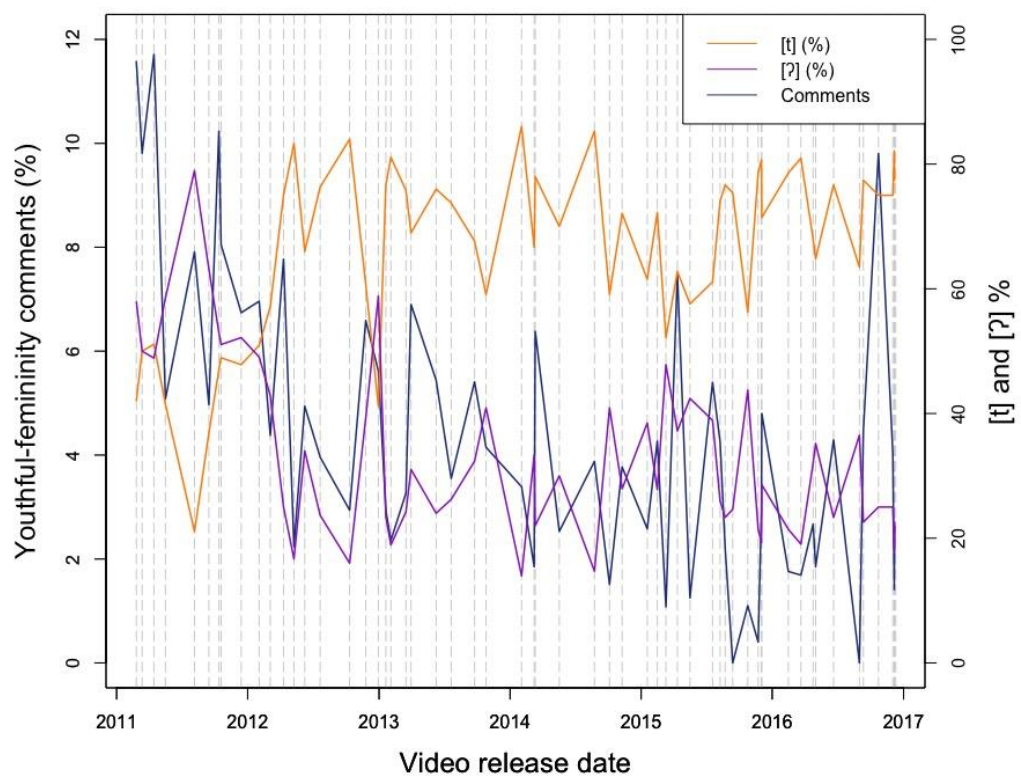
Figure 7.3. Percentage of [t], [?] and youthful-femininity comments per video
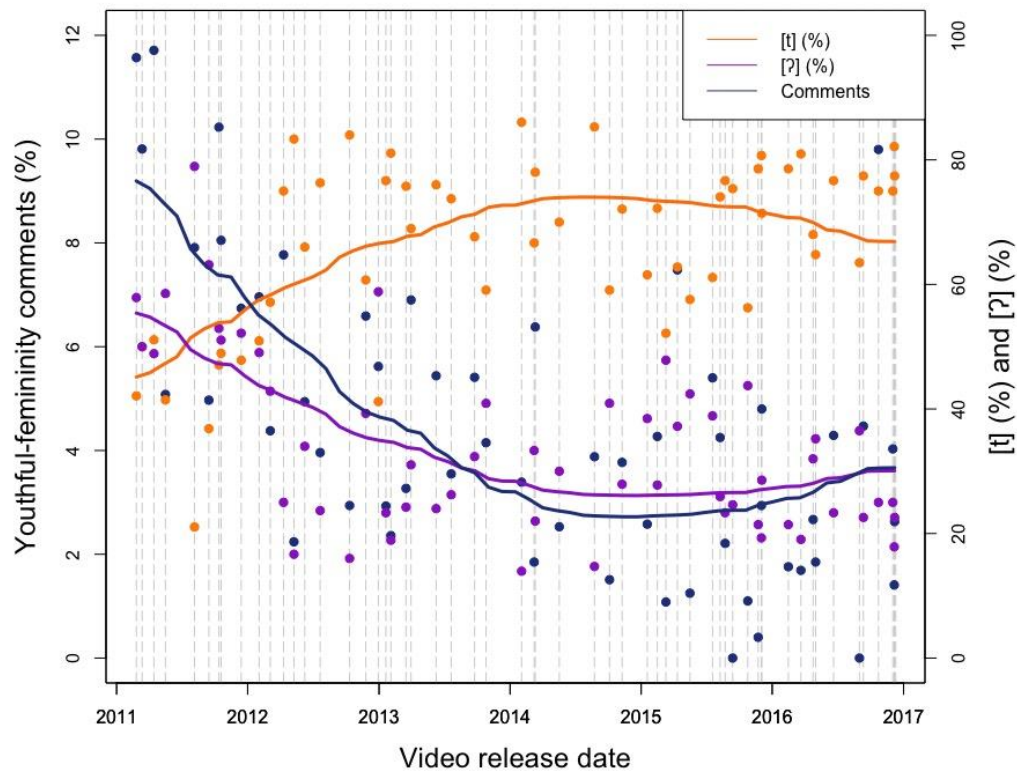
Figure 7.4. [t], [ʔ] and youthful-femininity comments lines of best fit (fitted with 2 polynomials)

7.4.3 Moderation analysis

According to Jaccard (2013), in causal theory three of the fundamental types of relationships that can occur are; i) Direct Causal Relationship, ii) Mediated Causal Relationship and iii) Moderated Causal Relationship. Historically, mediators and moderators have often been confused, or the terms used interchangeably, although they can be clearly distinguished by their differing properties (Baron and Kenny, 1986; Frazier, Tix and Barron, 2004). A *mediator* is an intermediary variable that an independent (antecedent) variable 'works through', fully or partially, to impact the dependent (consequent) variable. In other words, an antecedent variable causes a change in a mediator variable that then goes on to cause a change in the consequent variable. Whereas a *moderator* variable is able to affect both the direction and the strength (either amplifying or weakening the effect) of the relationship between an antecedent and a consequent variable. In other words, an antecedent variable may affect a consequent variable in some instances but not others depending on whether the

moderator variable is present or absent, or there is a certain amount or degree of the moderator present. The term 'interaction', which is synonymous for moderation, is also widely used in the literature. Therefore, rather than just asking if an association (correlation) and a relationship of some kind (linear regression with one dependent and one independent variable) exists, mediation and moderation analysis asks how and when, respectively, two variables may be causally related (Hayes, 2018). Figure 7.5 is a conceptual diagrams that illustrate these different types of causal relationships with the arrows indicating the direction of the effects.

To take an example from the literature, Reinikainen and colleagues (2020) found variables that mediated as well as variables that moderated the effect that a viewers' parasocial relationship with a YouTuber has upon their intention to purchase items featured in a product endorsing video from said YouTuber. A viewer's parasocial relationship with a YouTuber is the connection the viewer imagines that they have with them, such as feeling that they know the YouTuber well. The mediating variables were influencer credibility and brand trust. In other words, how well a viewer feels they know a YouTuber influences how credible they feel the YouTuber is, and how credible they feel the YouTuber is influences how much they trust the brand that the YouTuber is promoting, and how much they trust the brand influences how strong their intention to purchase a product promoted by said YouTuber is. Thus, the parasocial relationship between a viewer and YouTuber indirectly effects their intention to buy an endorsed product, and this indirect effect happens through influencer credibility and brand trust. Further, the moderating variable was the viewer seeing other viewers comments. In the experiment, in the condition where a viewer was asked to read other viewers' comments about the YouTuber (which were all positive) the effect the parasocial relationship had on influencer credibility was amplified in comparison to the condition where the viewer did not read any comments.
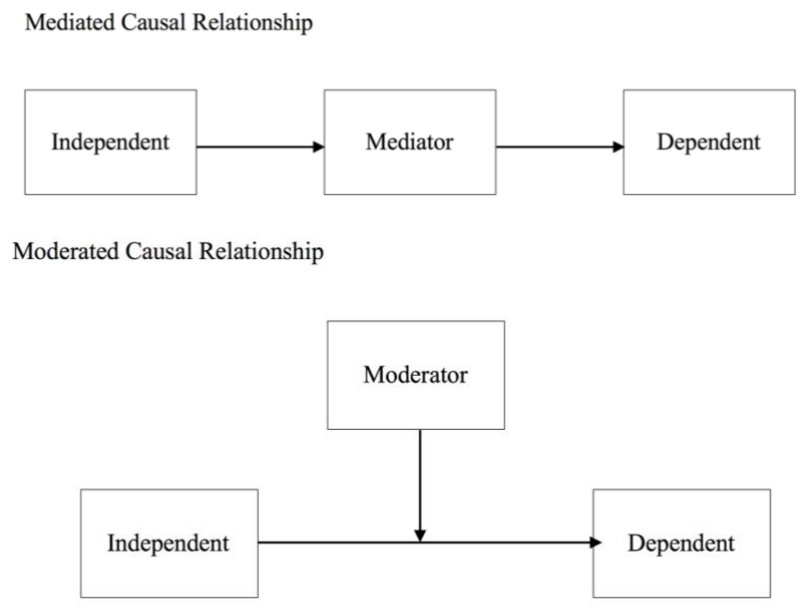
Figure 7.5. Conceptual diagrams of a simple mediated causal relationship and a simple moderated causal relationship. Recreated based on figures 1.1 (p.7) and 1.2 (p.8) in (Hayes, 2018).

7.4.3.1 Time

In this study, a variable that may act as a moderator is time. Time may be a moderator in that the strength of the effect that the comments have on Zoella's speech may depend on when they were posted. Time cannot be a mediator, however. First, this is because temporal order is vital in arguing that a variable is a mediator. The antecedent must be thought to precede the mediator, and the mediator must be thought to precede the consequent. Second, when the variables are in the required order it must be logical to argue that the preceding variable affects the proceeding variable (Hayes, 2018). The percentage of comments describing Zoella as cute cannot happen before time happens nor can they change time, of course.

Moderation is examined using a regression-based approach. Figure 7.6 is a statistical diagram for a simple moderation model. A statistical diagram is a representation of the conceptual diagram through the set of equations that are required to test it. Each line represents a linear regression, with the variable at the arrowhead being regressed onto the variable at the arrow's base (e.g. Y is regressed on X). The values of the resulting regression coefficient, R^2 , bootstrap

confidence intervals, and *p* provide insight into the relationship between the two regressed variables. In a simple moderation model the three calculations are: i) the effect of X on Y while holding W as 0 (b_1), ii) the effect of W on Y while holding X as 0 (b_2), and iii) the effect of the product of X and W (referred to as the interaction term) on Y (b_3). The coefficients derived from these three calculations are conditional effects – the effect of an antecedent variable (X, W or XW) on a consequent variable. (Y). Figure 7.7 is a visual representation of these three linear regressions to illustrate how they relate. Therefore, most simply, the null hypothesis that W is not a moderator can be rejected if the interaction term (b_3) is statistically significantly different from 0 because this indicates that the effect of X on Y is dependent on W.

It is evident that there is great debate amongst statisticians in regard to best practices in mediation and moderation analysis. Therefore, to conduct the statistical analysis herein (Hayes, 2018) alone was used for guidance. In addition to being one of the most recommended introductory texts, its content aligns with the use of this statistical approach in the social sciences, in comparison to other texts that focus on its application in, for example, the medical sciences (e.g. Vanderweele, 2015).
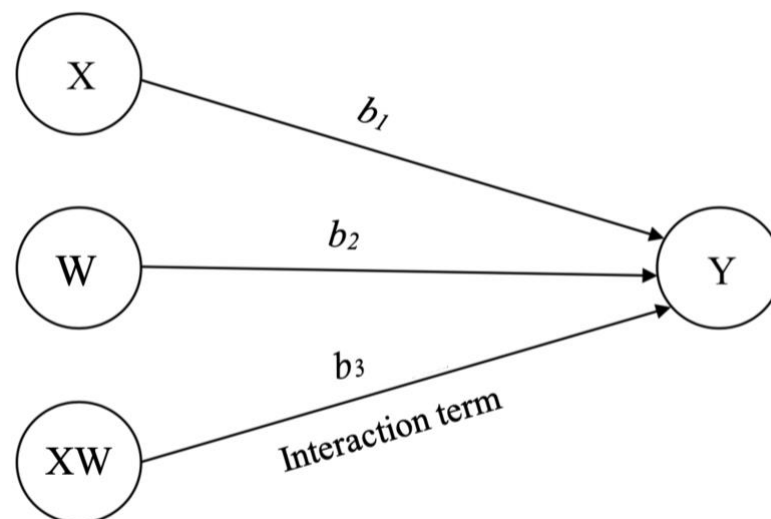
Figure 7.6. Statistical diagram of simple moderation model. Recreated based on figure 7.5B (p.235) in (Hayes, 2018).
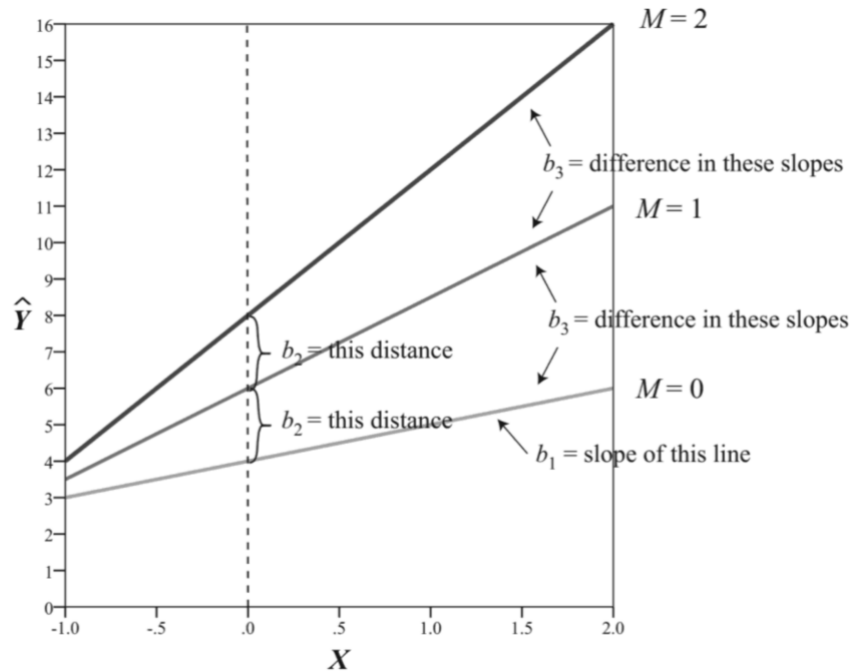
Figure 7.7. Visual representation of the linear regressions performed in a simple moderation model from (Hayes, 2018, p. 230 figure 7.4).

To assess if time is a moderator simple moderation models were tested, one with Y as [t] and the other with Y as [ʔ] (named *AlvT* and *Glottal* in the SPSS read out, respectively). X was *Comments* and W was *Time*. Note, time has been linearised as a percentage of the dataset in order to accommodate for the fact that the videos under study are not posted at regular intervals. Figure 7.8 is a display of the results from testing these models with annotations to assist in their interpretation. In regard to [t], the comment-time interaction term was found to be statistically significant with a positive coefficient ($b_3 = 0.042$, $p < 0.05$). Further, including the interaction term accounted for an additional 7.7% of variance in the data ($R^2 = 0.077$) compared to a model that included comments and time only.

The Johnson-Neyman (JN) test can give further insight into the moderator's effect by identifying the boundaries where it transitions between statistically significant and not, should there be such boundaries. The easiest way to explain this test is to describe the manual approach that was performed prior to its invention. This was the 'pick-a-point' approach. When the variables are continuous, a value of the moderator is selected and the standard error of the conditional effect of X on Y is calculated. Then the ratio of the conditional effect

231

to the standard error is calculated. Under the null hypothesis, this value equals 0 when at the selected moderator value. Finally, a p-value for this ratio is obtained by consulting a *t* distribution table. In comparison, the JN test is the pick-a-point approach in reverse. The ratio of the conditional effect of X on Y given W to its standard error is calculated. A *p*-value for this ratio is then derived by using the *t* distribution. In other words, the coefficient for each antecedent variable is taken from the linear model that includes the interaction term, and their ratio calculated. Then the probability of that ratio value is calculated based on a *t* distribution being formed from the data. This indicates if W is a 'Threshold' moderator; that there is a critical value which when reached by W changes its effect on X's influence on Y. For [t] the interaction term was deemed statistically significant until the value of time was 42%.

```
******************* **************************************************************
Model  : 1          OUTCOME VARIABLE:
  Y  : AlvT           AlvT
  X  : Comments
  W  : Time          Model Summary
                            R       R-sq        MSE          F        df1        df2          p
Sample                  .6344      .4025   126.7832    11.8985     3.0000    53.0000      .0000
Size:  57

                    Model
                              coeff         se          t          p       LLCI       ULCI
                    constant  73.8486     7.1263    10.3628      .0000    59.5549    88.1422
         b₁ ->  Comments     -3.0817      .9749    -3.1609      .0026    -5.0372    -1.1262
         b₂ ->  Time          -.0295      .0992     -.2977      .7671     -.2284      .1694
         b₃ ->  Int_1          .0418      .0161     2.6050      .0119      .0096      .0740


                    Test(s) of highest order unconditional interaction(s):
                            R2-chng        F        df1        df2          p
                    X*W      .0765     6.7860     1.0000    53.0000      .0119

                    Moderator value(s) defining Johnson-Neyman significance region(s):
                         Value    % below    % above
                       41.9547    42.1053    57.8947
******************* **************************************************************
Model  : 1          OUTCOME VARIABLE:
  Y  : Glottal        Glottal
  X  : Comments
  W  : Time          Model Summary
                            R       R-sq        MSE          F        df1        df2          p
Sample                  .6344      .4025   126.7832    11.8985     3.0000    53.0000      .0000
Size:  57

                    Model
                              coeff         se          t          p       LLCI       ULCI
                    constant  26.1514     7.1263     3.6697      .0006    11.8578    40.4451
         b₁ ->  Comments      3.0817      .9749     3.1609      .0026     1.1262     5.0372
         b₂ ->  Time           .0295      .0992      .2977      .7671     -.1694      .2284
         b₃ ->  Int_1         -.0418      .0161    -2.6050      .0119     -.0740     -.0096


                    Test(s) of highest order unconditional interaction(s):
                            R2-chng        F        df1        df2          p
                    X*W      .0765     6.7860     1.0000    53.0000      .0119

                    Moderator value(s) defining Johnson-Neyman significance region(s):
                         Value    % below    % above
                       41.9547    42.1053    57.8947
```

Figure 7.8. Results from testing time (W) as a moderator

These results allow for the null hypothesis that time does not moderate the relationship between the comments and speech to be rejected. Time moderates the effect of the comments on [t] in that as the value of time increases the strength of the negative effect that the comments have on [t] decreases. This is illustrated in figure 7.9.

The JN test further informs this conclusion. The interaction term was deemed statistically significant until the value of time was 42%. Time had been linearised for the purposes of statistical analysis, and so 42% equated to the date 01/08/2013, in between video 24 and 25. Therefore, it is evident that the comments had a greater effect on Zoella's speech in her early videos.

While this analysis has given a statistically significant result, it is nevertheless difficult to interpret. Based on the data gathered in this study so far, the date 01/08/2013 does not align with a particular event. In an attempt to gain greater insight, the JN result was plotted onto two graphs, placing this value within the context of the rest of the data. The first graph is the lines of best for the antecedent variable (comments) and the two consequent variables ([t] and [?]) (figure 7.10). The JN result appears to mark when the speech data begins to stabilise, which is not long before the comment data does. While this visualisation reconfirms this is a notable period of change, it does little to explain the data further. The second graph (figure 7.11) displays Zoella's comment engagement (see section 4.7.2) as well as marks out the time period before Zoella signed with a talent management agency (see section 6.4.2). Here, the JN result is soon after the first video where Zoella does not post a comment and approximately 8 months after she signs with Gleam. The variable time is, of course, not just the passing of time itself but the events that occurred as time passed. And the position of the JN result in relation to Zoella's comment engagement and her transitioning from a micro-celebrity to an A-List YouTuber suggests that these two events may have had some influence on the comments effecting her speech, and so will be investigated further.
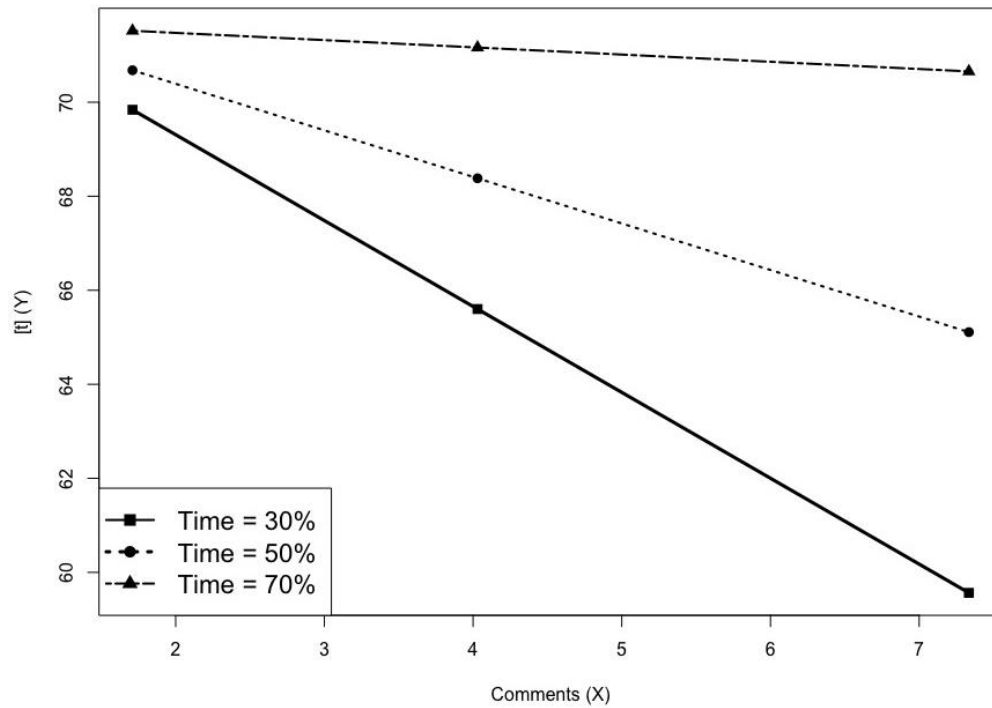
Figure 7.9. Visual representation of the moderation of the effect of comments (X) on [t] (Y) by time (W).
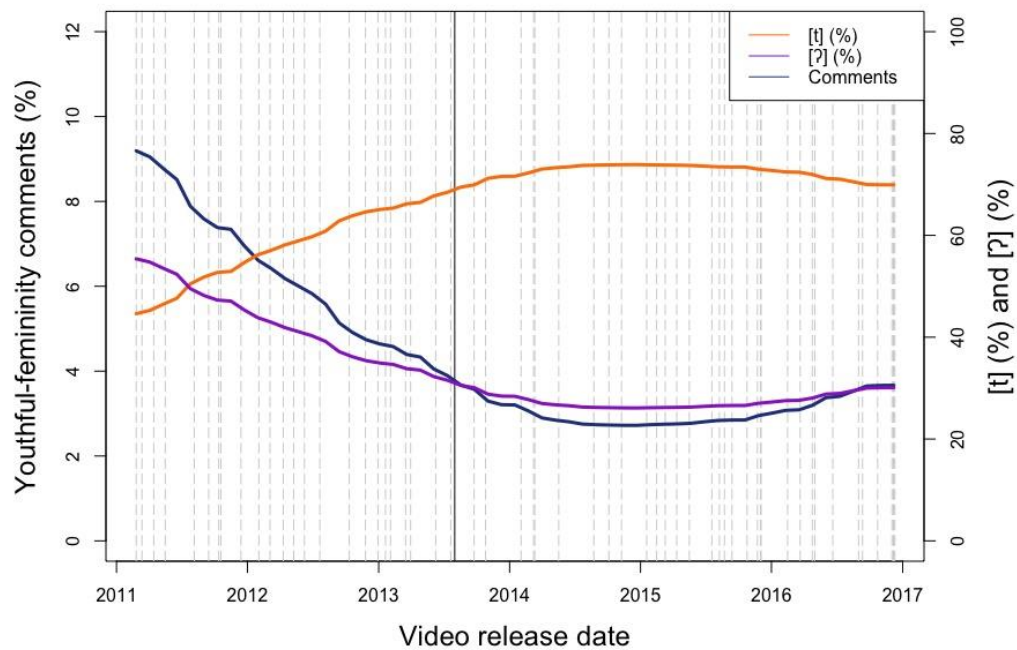
Figure 7.10. [t], [ʔ] and comments lines of best fit (fitted with 2 polynomials) with vertical black line indicating the boundary condition of time's moderation (left of line *p*<0.05)
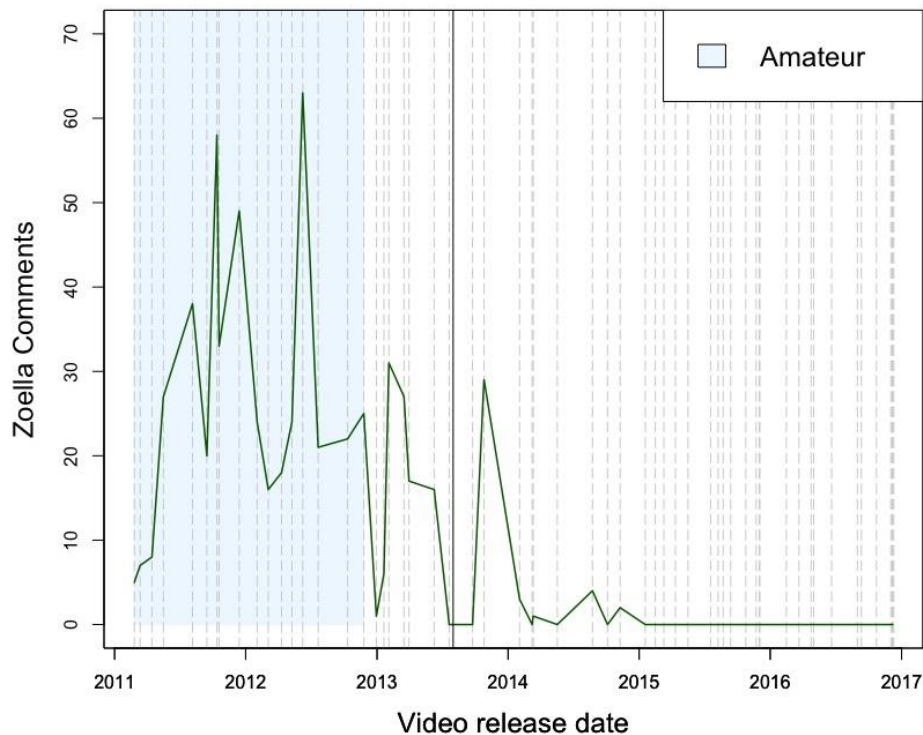
Figure 7.11. Zoella comment engagement (raw values) with amateur period of time marked, and vertical black line indicating the boundary condition of time's moderation (left of line *p*<0.05)

7.4.3.2 Zoella's comment engagement and Status

First, whether Zoella's Comment Engagement and Status will be tested as either mediator or moderators must be decided. Logically, there could be a direct causal relationship between the comments and speech, if Zoella is aware of the comment content. As is described in section 4.7.2, data of when and how much Zoella reads the comments on her videos is not available to us. Therefore, how much she posts comments on her own videos is being used as a proxy because it is reasonable to assume that before she replies she reads other comments in addition to the one that she is replying to. Thus, one hypothesis is that Zoella's comment engagement moderates the effect they have upon her speech. Further, several data sources indicate that there are many changes that occur along with Zoella transitioning from an amateur to professional YouTuber. These include the observations made through the digital ethnography as well as the youthful femininity comment data, namely that comments plateau along with [t] and [ʔ]. Thus, a second hypothesis is that whether Zoella is a A-List YouTuber or not moderates the effect the comments have upon her speech.

235

Neither of these variables are hypothesised to be mediators. As was stated in the previous section, this is because temporal order is vital in arguing that a variable is a mediator. The antecedent must be thought to precede the mediator, and the mediator must be thought to precede the consequent. Also, when the variables are in the required order it must be logical to argue that the preceding variable affects the proceeding variable (Hayes, 2018). Here, there is little argument that either the comments precede Zoella's status or that the comments alone affect Zoella's status (although they may make a contribution in conjunction with other factors that are not analysed in this study). Equally, while we know that Zoella commenting on her own video occurs after some (probably not all) of the comments analysed as a part of this study are posted, based on the data gathered at this stage in the project it is not reasonable to assume that the comments influence how much she engages. Therefore, neither Zoella's comment engagement nor Status will be tested as mediators.

First, whether Zoella's comment engagement moderates the effect of the comments on speech was tested. Figure 7.12 displays the results. For [t] the interaction term (b_3) is statistically significant ($p<0.05$) and the model including the interaction term accounts for an additional 6.1% ($R^2 = 0.061$) of the variance than a model without the interaction term. The interaction term was deemed statistically significant after the value 1.0552 and remains significant for the rest of the dataset. These results suggest that for the comments to have an effect on speech Zoella must engage with them. Also, because Zoella does not post any comments on 28 of the 57 videos (making the 'Zoella's comment engagement' value 0 for said videos) 54% of the data has a value below 1.0552 and 45% of the data has a value above 1.0552. This relatively even distribution of data across the boundary value 1.0552 strengthens this conclusion.

These results allow for the null hypothesis that Zoella's comment engagement does not moderate the relationship between the comments and speech to be rejected. Zoella's comment engagement moderates the effect of the comments on [t] in that as her engagement increases the strength of the negative effect that the comments have on [t] increases also. This is illustrated in figure 7.13.

```
*******************   ***********************************************************
Model  : 1            OUTCOME VARIABLE:
  Y   : AlvT          AlvT           Product terms key:
  X   : Comments                       Int_1    :       Comments x        ZEngage
  W   : ZEngage       Model Summary
Sample                       R       R-sq       MSE        F       df1       df2         p
Size: 57                  .5698     .3247    143.2749   8.4954   3.0000   53.0000     .0001

                      Model
                                   coeff        se         t         p       LLCI      ULCI
                      constant    73.9391    3.2780    22.5559    .0000    67.3642   80.5141
            b₁ ->     Comments    -1.2620     .6934    -1.8201    .0744    -2.6528    .1288
            b₂ ->     ZEngage       .3142     .2583     1.2164    .2292     -.2039    .8324
            b₃ ->     Int_1        -.0866     .0397    -2.1811    .0336     -.1662    -.0070

                      Test(s) of highest order unconditional interaction(s):
                               R2-chng       F        df1       df2         p
                      X*W        .0606    4.7572    1.0000   53.0000     .0336

                      Moderator value(s) defining Johnson-Neyman significance region(s):
                           Value    % below    % above
                          1.0552    54.3860    45.6140
*******************   ***********************************************************
Model  : 1            OUTCOME VARIABLE:
  Y   : Glottal       Glottal        Product terms key:
  X   : Comments                       Int_1    :       Comments x        ZEngage
  W   : ZEngage       Model Summary
Sample                       R       R-sq       MSE        F       df1       df2         p
Size: 57                  .5698     .3247    143.2749   8.4954   3.0000   53.0000     .0001

                      Model
                                   coeff        se         t         p       LLCI      ULCI
                      constant    26.0609    3.2780     7.9502    .0000    19.4859   32.6358
            b₁ ->     Comments     1.2620     .6934     1.8201    .0744     -.1288    2.6528
            b₂ ->     ZEngage      -.3142     .2583    -1.2164    .2292     -.8324    .2039
            b₃ ->     Int_1         .0866     .0397     2.1811    .0336      .0070    .1662

                      Test(s) of highest order unconditional interaction(s):
                               R2-chng       F        df1       df2         p
                      X*W        .0606    4.7572    1.0000   53.0000     .0336

                      Moderator value(s) defining Johnson-Neyman significance region(s):
                           Value    % below    % above
                          1.0552    54.3860    45.6140
```

Figure 7.12. Results from testing Zoella's comment engagement (W) as a moderator of [t] and [ʔ] (Y).

Figure 7.13. Visual representation of the moderation of the effect of comments (X) on [t] (Y) by Zoella's comment engagement (W).

Next, whether being a micro-celebrity or A-List YouTuber moderates the effect of the comments on speech was tested, a variable that will be referred to as 'status'. While it has been argued that social media creates a context of 'celebrification', a continuum of 'celebrifying' one's self through a variety of behaviours, as was detailed in chapter 6, it is unclear how one would quantify such a continuum. However, a notable change that reflects Zoella's YouTube activities transitioning from a hobby to one where she earns an income is signing with a talent management company. It is not clear the exact date that Zoella signed with talent management company Gleam Futures but a video posted on the 9th of December 2012 stated the Gleam office as Zoella's postal address in the information box (Sugg, 2012), and so the first haul video posted after this (video 18, 30/12/2012) will be defined as her transition to an A List YouTuber. To examine the impact of this transition through statistical analysis I created a new variable that I named "Status", with videos posted prior to video 18 on the 30/12/2012 deemed to be 'amateur' and video 18 and those posted after being deemed 'professional'. To examine whether status moderates the effect the comments have on Zoella's speech, simple moderation models were tested. Rereferring to figure 7.6, X is the comments, Y is the speech variable ([t] or [?]) and W is Status (amateur or professional). Therefore, four simple moderation models were tested, detailed in table 7.5. b_1 is the conditional effect of X on Y when W is 0 and therefore, can provide insight when testing a dichotomous variable if the two conditions are coded as 0 and 1, with the condition that is being tested in a model being coded 0. To clarify, in models 1 and 3 the amateur videos were coded as 0 and the professional videos coded as 1, and this was reversed for models 2 and 4.

Table 7.5. The four simple moderation models tested.

	X	**Y**	**W**
Model 1	Comments	[t]	Amateur
Model 2	Comments	[t]	Professional
Model 3	Comments	[?]	Amateur
Model 4	Comments	[?]	Professional

The results for these models are displayed in figure 7.14. The comment-status interaction term was found to be statistically significant for all four models (all *p* <0.05). This result allows for the null hypothesis that Zoella's status does not moderate the effect of the comments on speech to be rejected. Further, all four models accounted for an additional 6.3% of variance in the data ($R^2 = 0.063$) compared to models that did not include their respective interaction terms. This indicates that the effect of the comments on Zoella's speech depends on whether she is an amateur or professional.

Because status is a dichotomous variable it is not possible to perform a JN test. So, the 'pick-a-point' approach was used, with the points in question being '0' and '1'. As stated above, b_1 is the conditional effect of X on Y when W is 0, therefore the regression coefficient of b_1 estimates the conditional effect of the status designated as '0' in the model. For [t] b_1 was statistically significant for the amateur condition (*p* <0.01) but not statistically significant for the professional condition (*p* = 0.8542). These results indicate that when she is an amateur the comments effect Zoella's speech but when she is a professional YouTuber, they do not. In the amateur condition b_1 was negative (coeff = -3.21). So, the effect the comments have on [t] is negative when she is an amateur. See figures 7.15 for a visualisation.

However, from visualising these moderator variables together earlier it is evident that they are associated. Figure 7.11 suggests that when Zoella is amateur she engages with the comments more than when she is professional. Therefore, the two apparent moderators Zoella comment engagement and status will be tested together in the same model.

```
******************** ***********************************************************
Model  : 1           OUTCOME VARIABLE:        ┌─────────────────────────────────────────┐
  Y  : AlvT            AlvT                    │ Product terms key:                      │
  X  : Comments                               │   Int_1     :       Comments x     StatP0 │
  W  : StatP0                                 └─────────────────────────────────────────┘
                      Model Summary
                          R        R-sq      MSE        F        df1       df2        p
Sample                  .6112    .3736    132.8997    10.5378   3.0000    53.0000    .0000
Size:  57

                      Model
                                 coeff       se        t         p        LLCI      ULCI
                      constant  71.2270    3.4520   20.6336    .0000    64.3031   78.1508
     b₁ -> Comments    -.1572     .8511    -.1847    .8542    -1.8642    1.5498
     b₂ -> StatP0      5.6276    8.2217     .6845    .4967   -10.8631   22.1183
     b₃ -> Int_1      -3.0511    1.3192   -2.3129    .0246    -5.6971    -.4052

                      Test(s) of highest order unconditional interaction(s):
                              R2-chng      F        df1       df2         p
                      X*W       .0632    5.3496    1.0000   53.0000     .0246

                      Conditional effects of the focal predictor at values of the moderator(s):

                          StatP0    Effect      se        t         p        LLCI      ULCI
     Professional -> .0000      -.1572     .8511    -.1847    .8542    -1.8642    1.5498
        Amateur -> 1.0000      -3.2083    1.0079   -3.1831    .0024    -5.2300   -1.1867
******************** ***********************************************************
Model  : 1           OUTCOME VARIABLE:        ┌─────────────────────────────────────────┐
  Y  : AlvT            AlvT                    │ Product terms key:                      │
  X  : Comments                               │   Int_1     :       Comments x     StatA0 │
  W  : StatA0                                 └─────────────────────────────────────────┘
                      Model Summary
                          R        R-sq      MSE        F        df1       df2        p
Sample                  .6112    .3736    132.8997    10.5378   3.0000    53.0000    .0000
Size:  57

                      Model
                                 coeff       se        t         p        LLCI      ULCI
                      constant  76.8546    7.4619   10.2996    .0000    61.8879   91.8213
     b₁ ->  Comments   -3.2083    1.0079   -3.1831    .0024    -5.2300   -1.1867
     b₂ ->  StatA0     -5.6276    8.2217    -.6845    .4967   -22.1183   10.8631
     b₃ ->  Int_1       3.0511    1.3192    2.3129    .0246      .4052    5.6971

                      Test(s) of highest order unconditional interaction(s):
                              R2-chng      F        df1       df2         p
                      X*W       .0632    5.3496    1.0000   53.0000     .0246

                      Conditional effects of the focal predictor at values of the moderator(s):

                          StatA0    Effect      se        t         p        LLCI      ULCI
        Amateur -> .0000      -3.2083    1.0079   -3.1831    .0024    -5.2300   -1.1867
     Professional -> 1.0000      -.1572     .8511    -.1847    .8542    -1.8642    1.5498
******************** ***********************************************************
Model  : 1           OUTCOME VARIABLE:        ┌─────────────────────────────────────────┐
  Y  : Glottal         Glottal                 │ Product terms key:                      │
  X  : Comments                               │   Int_1     :       Comments x     StatP0 │
  W  : StatP0                                 └─────────────────────────────────────────┘
                      Model Summary
                          R        R-sq      MSE        F        df1       df2        p
Sample                  .6112    .3736    132.8997    10.5378   3.0000    53.0000    .0000
Size:  57

                      Model
                                 coeff       se        t         p        LLCI      ULCI
                      constant  28.7730    3.4520    8.3352    .0000    21.8492   35.6969
     b₁ -> Comments     .1572     .8511     .1847    .8542    -1.5498    1.8642
     b₂ -> StatP0      -5.6276    8.2217    -.6845    .4967   -22.1183   10.8631
     b₃ -> Int_1        3.0511    1.3192    2.3129    .0246      .4052    5.6971

                      Test(s) of highest order unconditional interaction(s):
                              R2-chng      F        df1       df2         p
                      X*W       .0632    5.3496    1.0000   53.0000     .0246

                      Conditional effects of the focal predictor at values of the moderator(s):

                          StatP0    Effect      se        t         p        LLCI      ULCI
     Professional -> .0000       .1572     .8511     .1847    .8542    -1.5498    1.8642
        Amateur -> 1.0000       3.2083    1.0079    3.1831    .0024     1.1867    5.2300
******************** ***********************************************************
Model  : 1           OUTCOME VARIABLE:        ┌─────────────────────────────────────────┐
  Y  : Glottal  Glottal                        │ Product terms key:                      │
  X  : Comments                               │   Int_1     :       Comments x     StatA0 │
  W  : StatA0  Model Summary                  └─────────────────────────────────────────┘
                          R        R-sq      MSE        F        df1       df2        p
Sample                  .6112    .3736    132.8997    10.5378   3.0000    53.0000    .0000
Size:  57

                      Model
                                 coeff       se        t         p        LLCI      ULCI
                      constant  23.1454    7.4619    3.1018    .0031     8.1787   38.1121
     b₁ -> Comments    3.2083    1.0079    3.1831    .0024     1.1867    5.2300
     b₂ -> StatA0      5.6276    8.2217     .6845    .4967   -10.8631   22.1183
     b₃ -> Int_1      -3.0511    1.3192   -2.3129    .0246    -5.6971    -.4052

                      Test(s) of highest order unconditional interaction(s):
                              R2-chng      F        df1       df2         p
                      X*W       .0632    5.3496    1.0000   53.0000     .0246

                      Conditional effects of the focal predictor at values of the moderator(s):

                          StatA0    Effect      se        t         p        LLCI      ULCI
        Amateur -> .0000       3.2083    1.0079    3.1831    .0024     1.1867    5.2300
     Professional -> 1.0000       .1572     .8511     .1847    .8542    -1.5498    1.8642
```

Figure 7.14. Results from testing Status (W) as a moderator of [t] and [ʔ]
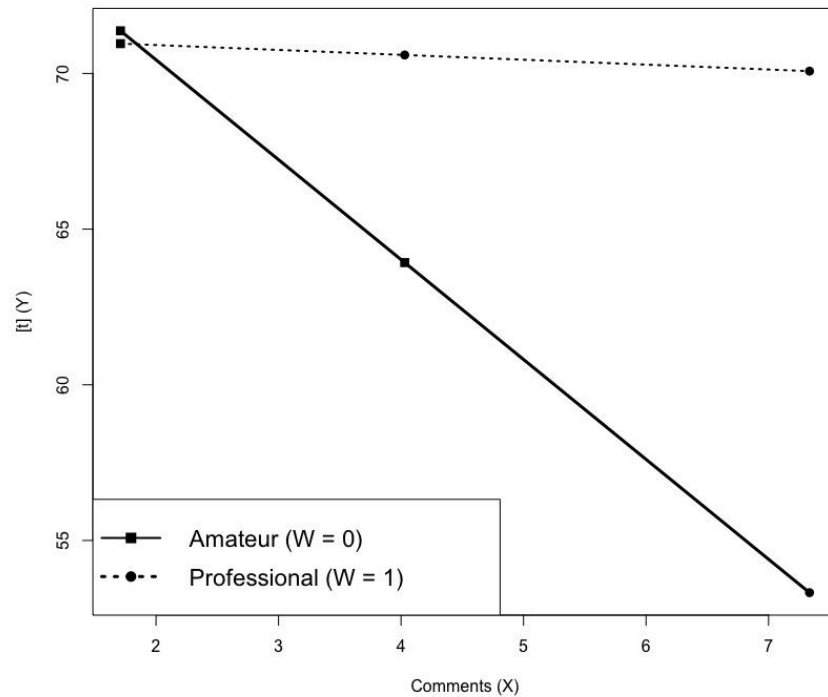
Figure 7.15. Visual representation of the moderation of the effect of comments (X) on [t] (Y) by status (W).

7.4.3.3 Covariation, moderated moderation models and additive multiple moderation

This section reports the three approaches that were taken to ascertain if and how the two apparent moderator variables both moderate the effects of comments on Zoella's speech at the same time. First, each apparent moderator variable underwent retesting as a moderator while controlling for the other. This was to establish that the apparent moderation caused by one is actually not due to the effect of the other. Second, two moderated moderation models were tested to see if Zoella's comment engagement was moderated by status and vice versa. In other words, these models tested whether one of the apparent moderators influences the effect of comments on Zoella's speech by moderating the other apparent moderator. Third, an additive multiple moderation model was tested to see if the comments' effect on Zoella's speech is moderated by both Zoella's comment engagement and status *simultaneously*.

Hayes (2018) states that when performing regression analysis it should be assumed that the antecedent variables are correlated with each other. So far, the analysis herein has not included covariates in the models. Including a covariate (C) in a simple moderation model adjusts the model so that it tests for W's moderation of the effect of X on Y while controlling for C. In doing so, the influence of W without the influence of C can be ascertained. In other words, whether the moderating effects that the variables i) Zoella's comment engagement and ii) status have are actually because they are so closely related can be investigated. Figure 7.16 provides a conceptual diagram and figure 7.17 provides a statistical diagram of such a model. First, simple moderation models testing Zoella's comment engagement and controlling for status were tested. The results are displayed in figure 7.18. To be concise, only the results where Amateur = 0 for status is used. When status is controlled for all the values of the model are exactly the same (apart from the constant) for a model where Professional = 0 with the only difference being the positivity/negativity of the coefficient, *t*, lower and upper 95% confidence interval values (labelled LLCI and ULCI, respectively) are reversed, and swapped too if they do not cross 0.
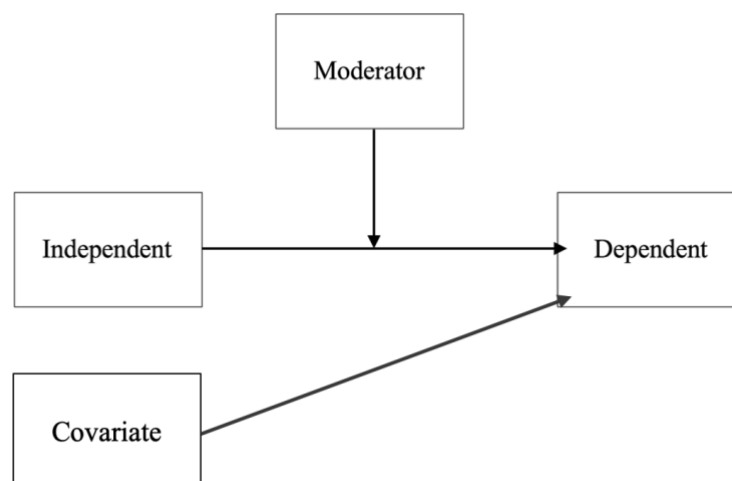
Figure 7.16. Conceptual diagram of simple moderation model with covariate.
Recreated based on figure 8.4A (p.279) (Hayes, 2018)
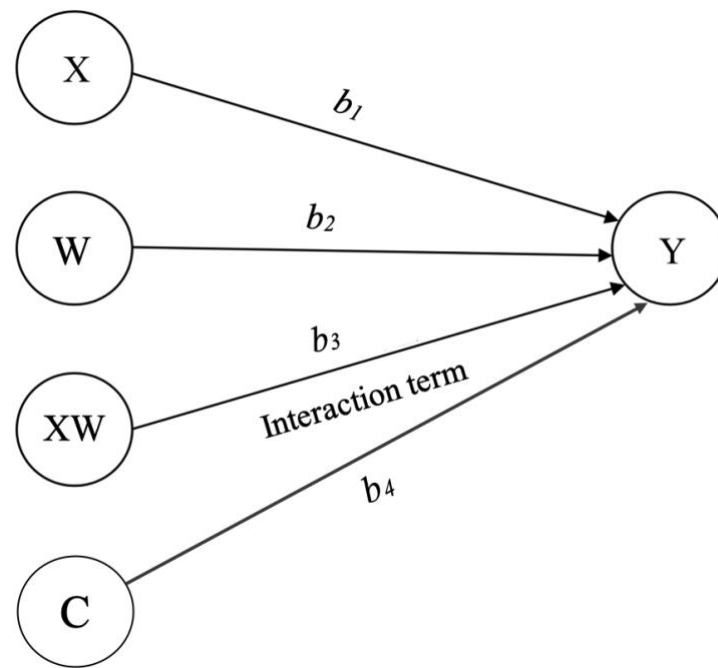
Figure 7.17. Statistical diagram of simple moderation model with covariate. Recreated based on figure 8.4B (p.279) (Hayes, 2018).

The interaction term of comments and Zoella comment engagement remained statistically significant and negative for [t] (b_3 = -0.092, p <0.05). Further, the interaction term accounted for an additional 6.8% of variance in the data (R^2 = 0.068) compared to a model that included comments and Zoella comment engagement only. The result of the JN test is that the interaction term is statistically significant when the moderator value is above 7.4661.

Next, the models where *status* is and is not controlled will be compared (see figures 7.12 and 7.13 for a reminder of the results for the models where *status* is not controlled). For [t] the strength of the moderator's effect increases (from - 0.087 to - 0.092) and the amount of variance accounted for increases (from R^2 = 0.061 to 0.068) in the model where *status* is controlled. Most interesting is the increase in the JN test result from 1.0552 to 7.4661. This indicates that a minimal level of engagement with the comments by Zoella is required before the comment's affect her speech, rather than the impression gained previously that any degree of engagement moderates the comments' effect on her speech.

```
******************** *******************************************************************
Model  : 1        OUTCOME VARIABLE:
  Y  : AlvT         AlvT                        ┌─────────────────────────────────────┐
  X  : Comments                                 │ Product terms key:                  │
  W  : ZEngage    Model Summary                 │   Int_1    :      Comments x   ZEngage│
                          R      R-sq     MSE        F       df1       df2        p
Covariates:             .6157   .3791  134.2648   7.9383   4.0000   52.0000    .0000
 StatA0
                  Model
Sample                          coeff      se        t         p       LLCI      ULCI
Size:  57         constant    61.5914   6.5977   9.3353    .0000    48.3520   74.8308
          b₁ ->   Comments     -.6300    .7336   -.8587    .3945    -2.1021     .8422
          b₂ ->   ZEngage       .5144    .2671   1.9261    .0596     -.0215    1.0504
          b₃ ->   Int_1        -.0918    .0385  -2.3826    .0209     -.1690     -.0145
          b₄ ->   StatA0      10.8362   5.0764   2.1346    .0375      .6496   21.0228


                  Test(s) of highest order unconditional interaction(s):
                          R2-chng       F       df1       df2        p
                  X*W      .0678     5.6768   1.0000   52.0000     .0209


                  Moderator value(s) defining Johnson-Neyman significance region(s):
                        Value     % below    % above
                       7.4661    64.9123    35.0877


******************** *******************************************************************
Model  : 1        OUTCOME VARIABLE:
  Y  : Glottal      Glottal                     ┌─────────────────────────────────────┐
  X  : Comments                                 │ Product terms key:                  │
  W  : ZEngage    Model Summary                 │   Int_1    :      Comments x   ZEngage│
                          R      R-sq     MSE        F       df1       df2        p
Covariates:             .6157   .3791  134.2648   7.9383   4.0000   52.0000    .0000
 StatA0
                  Model
Sample                          coeff      se        t         p       LLCI      ULCI
Size:  57         constant    38.4086   6.5977   5.8215    .0000    25.1692   51.6480
          b₁ ->   Comments      .6300    .7336    .8587    .3945     -.8422    2.1021
          b₂ ->   ZEngage      -.5144    .2671  -1.9261    .0596    -1.0504     .0215
          b₃ ->   Int_1         .0918    .0385   2.3826    .0209      .0145     .1690
          b₄ ->   StatA0     -10.8362   5.0764  -2.1346    .0375   -21.0228     -.6496


                  Test(s) of highest order unconditional interaction(s):
                          R2-chng       F       df1       df2        p
                  X*W      .0678     5.6768   1.0000   52.0000     .0209


                  Moderator value(s) defining Johnson-Neyman significance region(s):
                        Value     % below    % above
                       7.4661    64.9123    35.0877
```

Figure 7.18. Results from simple moderation models testing the moderating effect of Zoella's comment engagement (W) with Status (C) as a covariate.

Next, four simple moderation models testing status and controlling for Zoella comment engagement were tested (see table 7.5 for variables X, Y and W). The results are displayed in figure 7.19, and those for [t] are reported here. The interaction term of comments and status remained statistically significant (p <0.05) for the amateur condition and remained not statistically significant for the professional condition. All models including the interaction term accounted for an additional 6.8% of variance in the data ($R^2 = 0.068$) compared to a model that included comments and status only. As stated above, b_1 is the conditional effect of X on Y when W is 0, therefore the regression coefficient of b_1 estimates the conditional effect of the status designated as '0' in the model. b_1 was statistically significant for the amateur condition (p <0.01) but not statistically significant for the professional conditions ($p = 0.9278$). The coefficient was negative ($b_3 = -3.295$). Comparing the models, the strength of the moderator's effect increases (from - 3.21 to - 3.295) and the amount of variance (R^2) accounted for increases (from 0.063 to 0.068) in the model where Zoella's comment engagement is controlled.

These results suggest that even when one moderator variable is controlled for, the other still moderates X's effect on Y. This indicates that both of the variables apply their own moderation to the effect that the comments have on Zoella's speech. The results reported here mirror those where covariates are not controlled, most notably in regard to statistical significance and the direction of the coefficients. The one key difference is the value of the JN test increasing for *Zoella's comment engagement*. This suggests that this variable's moderating effect is not triggered by its mere presence (the JN in the previous model was 1.0552) but by a certain amount of engagement with the comments by Zoella. In other words, the moderation that *Zoella comment engagement* and *status* perform have been untangled from one another.

Now that it has been identified that *Zoella comment engagement* and *status* impose their own individual moderation of the effect of the comments on Zoella's speech, how this occurs will be investigated.

```
******************** ********************************************************************
Model : 1           OUTCOME VARIABLE:
  Y  : AlvT           AlvT                       ┌─────────────────────────────────────┐
  X  : Comments                                  │Product terms key:                   │
  W  : StatA0                                    │ Int_1   :        Comments x    StatA0│
                    Model Summary                └─────────────────────────────────────┘
Covariates:                 R       R-sq       MSE         F        df1        df2          p
  ZEngage             .6160    .3794    134.1981     7.9487    4.0000    52.0000      .0000

Sample              Model
Size: 57                        coeff         se          t          p        LLCI        ULCI
                    constant   79.9975     8.7463     9.1465      .0000     62.4466     97.5483
          b₁ ->     Comments   -3.2951     1.0204    -3.2291      .0022     -5.3428     -1.2475
          b₂ ->     StatA0     -8.6702     9.3411     -.9282      .3576    -27.4147     10.0743
          b₃ ->     Int_1       3.2165     1.3466     2.3886      .0206       .5143      5.9187
          b₄ ->     ZEngage     -.0942      .1350     -.6980      .4883      -.3652       .1767

                    Test(s) of highest order unconditional interaction(s):
                              R2-chng         F        df1        df2          p
                    X*W         .0681     5.7055     1.0000    52.0000      .0206
******************** ********************************************************************
Model : 1           OUTCOME VARIABLE:
  Y  : AlvT           AlvT                       ┌─────────────────────────────────────┐
  X  : Comments                                  │Product terms key:                   │
  W  : StatP0                                    │ Int_1   :        Comments x    StatP0│
                    Model Summary                └─────────────────────────────────────┘
Covariates:                 R       R-sq       MSE         F        df1        df2          p
  ZEngage             .6160    .3794    134.1981     7.9487    4.0000    52.0000      .0000

Sample              Model
Size: 57                        coeff         se          t          p        LLCI        ULCI
                    constant   71.3273     3.4718    20.5448      .0000     64.3606     78.2940
          b₁ ->     Comments    -.0786      .8626     -.0911      .9278     -1.8095      1.6523
          b₂ ->     StatP0      8.6702     9.3411      .9282      .3576    -10.0743     27.4147
          b₃ ->     Int_1      -3.2165     1.3466    -2.3886      .0206     -5.9187      -.5143
          b₄ ->     ZEngage     -.0942      .1350     -.6980      .4883      -.3652       .1767

                    Test(s) of highest order unconditional interaction(s):
                              R2-chng         F        df1        df2          p
                    X*W         .0681     5.7055     1.0000    52.0000      .0206
******************** ********************************************************************
Model : 1           OUTCOME VARIABLE:
  Y  : Glottal        Glottal                    ┌─────────────────────────────────────┐
  X  : Comments                                  │Product terms key:                   │
  W  : StatA0                                    │ Int_1   :        Comments x    StatA0│
                    Model Summary                └─────────────────────────────────────┘
Covariates:                 R       R-sq       MSE         F        df1        df2          p
  ZEngage             .6160    .3794    134.1981     7.9487    4.0000    52.0000      .0000

Sample              Model
Size: 57                        coeff         se          t          p        LLCI        ULCI
                    constant   20.0025     8.7463     2.2870      .0263      2.4517     37.5534
          b₁ ->     Comments    3.2951     1.0204     3.2291      .0022      1.2475      5.3428
          b₂ ->     StatA0      8.6702     9.3411      .9282      .3576    -10.0743     27.4147
          b₃ ->     Int_1      -3.2165     1.3466    -2.3886      .0206     -5.9187      -.5143
          b₄ ->     ZEngage      .0942      .1350      .6980      .4883      -.1767       .3652

                    Test(s) of highest order unconditional interaction(s):
                              R2-chng         F        df1        df2          p
                    X*W         .0681     5.7055     1.0000    52.0000      .0206
******************** ********************************************************************
Model : 1           OUTCOME VARIABLE:
  Y  : Glottal        Glottal                    ┌─────────────────────────────────────┐
  X  : Comments                                  │Product terms key:                   │
  W  : StatP0                                    │ Int_1   :        Comments x    StatP0│
                    Model Summary                └─────────────────────────────────────┘
Covariates:                 R       R-sq       MSE         F        df1        df2          p
  ZEngage             .6160    .3794    134.1981     7.9487    4.0000    52.0000      .0000

Sample              Model
Size: 57                        coeff         se          t          p        LLCI        ULCI
                    constant   28.6727     3.4718     8.2588      .0000     21.7060     35.6394
          b₁ ->     Comments     .0786      .8626      .0911      .9278     -1.6523      1.8095
          b₂ ->     StatP0     -8.6702     9.3411     -.9282      .3576    -27.4147     10.0743
          b₃ ->     Int_1       3.2165     1.3466     2.3886      .0206       .5143      5.9187
          b₄ ->     ZEngage      .0942      .1350      .6980      .4883      -.1767       .3652

                    Test(s) of highest order unconditional interaction(s):
                              R2-chng         F        df1        df2          p
                    X*W         .0681     5.7055     1.0000    52.0000      .0206
```

Figure 7.19. Results from simple moderation models testing Status (W) with Zoella's comment engagement (C) as a covariate.

246

Next, moderated moderation models were tested (see figure 7.20 for a conceptual diagram). Simply, the moderating effect of one variable is itself moderated by another variable, so there is a *primary moderator* (W) and a *secondary moderator* (Z). For example, it could be that Zoella's comment engagement (*primary moderator*) is moderated by status (*secondary moderator*). In other words, whether Zoella's status is amateur or professional moderates how much she cares about the comments she engages with. Figure 7.21 provides a statistical diagram of such a model. In such a model Zoella's comment engagement is W and status is Z, and comments is X and speech is Y as always.

As can be seen from figure 7.22 and 7.23, none of the linear regressions performed as part of testing a moderated moderation model are statistically significant for either [t] or [ʔ]. Therefore, the null hypothesis that Status does not moderate Zoella's comment engagement which does not moderate the effect of Comments on speech cannot be rejected.
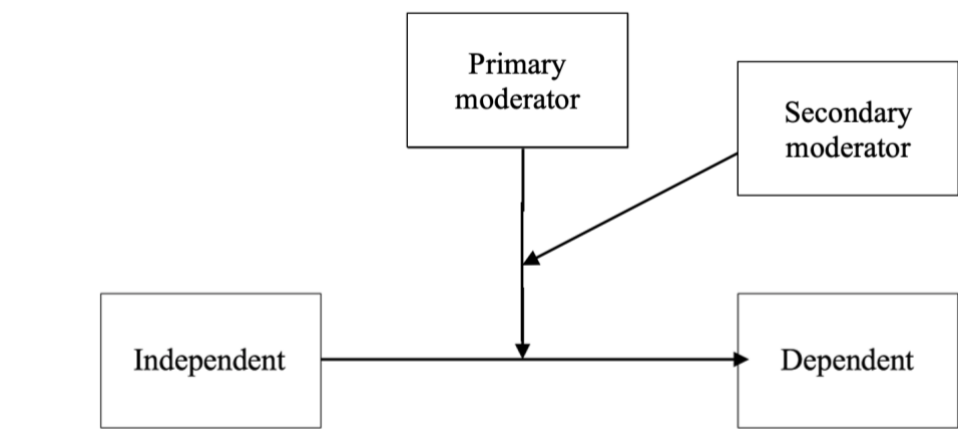
Figure 7.20. Conceptual diagram of a moderated moderation model. Recreated based on figure 9.4A (p.331) in (Hayes, 2018)
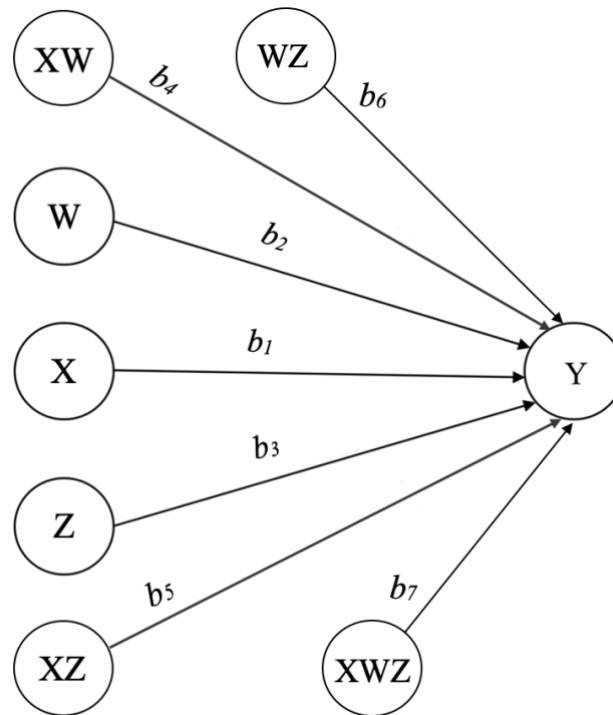
Figure 7.21. Statistical diagram of a moderated moderation model. Recreated based on figure 9.4B (p.331) in (Hayes, 2018)

```
******************* *******************************************************************
Model  : 3          OUTCOME VARIABLE:
  Y  : AlvT           AlvT
  X  : Comments
  W  : ZEngage       Model Summary
  Z  : StatA0               R        R-sq        MSE          F        df1        df2          p
                         .6381       .4072    136.0447     4.8082     7.0000    49.0000      .0004
Sample
Size:  57            Model
                                    coeff         se          t          p        LLCI        ULCI
                     constant     75.7275    15.8183     4.7873      .0000     43.9392    107.5158
          b₁ -> Comments       -2.5564     1.7912    -1.4272      .1599     -6.1560      1.0432
          b₂ -> ZEngage          .1297      .5775      .2245      .8233     -1.0309      1.2902
          b₄ -> Int_1           -.0385      .0693     -.5563      .5805      -.1778       .1007
          b₃ -> StatA0         -5.6945    16.2297     -.3509      .7272    -38.3096     26.9206
          b₅ -> Int_2           2.7779     2.0226     1.3734      .1759     -1.2867      6.8426
          b₆ -> Int_3            .5033      .7955      .6327      .5299     -1.0953      2.1018
          b₇-> Int_4            -.1118      .1426     -.7841      .4367      -.3984       .1748

                     Product terms key:
                      Int_1    :        Comments x        ZEngage
                      Int_2    :        Comments x        StatA0
                      Int_3    :        ZEngage  x        StatA0
                      Int_4    :        Comments x        ZEngage  x        StatA0

                     Test(s) of highest order unconditional interaction(s):
                            R2-chng         F         df1        df2          p
                     X*W*Z     .0074      .6149     1.0000    49.0000      .4367
```

Figure 7.22. Results from moderated moderation models predicting [t] with Zoella's comment engagement as primary moderator (W) and Status as secondary moderator (Z).

```
****************************  ************************************************************
Model  : 3          OUTCOME VARIABLE:
   Y  : Glottal       Glottal
   X  : Comments
   W  : ZEngage     Model Summary
   Z  : StatA0               R       R-sq       MSE         F       df1       df2         p
                          .6381      .4072   136.0447    4.8082    7.0000   49.0000     .0004
Sample
Size:  57          Model
                                 coeff        se         t         p       LLCI       ULCI
                   constant    24.2725   15.8183    1.5345     .1313    -7.5158    56.0608
            b₁ -> Comments      2.5564    1.7912    1.4272     .1599    -1.0432     6.1560
            b₂ -> ZEngage       -.1297     .5775    -.2245     .8233    -1.2902     1.0309
            b₄ -> Int_1          .0385     .0693     .5563     .5805     -.1007      .1778
            b₃ -> StatA0        5.6945   16.2297     .3509     .7272   -26.9206    38.3096
            b₅ -> Int_2        -2.7779    2.0226   -1.3734     .1759    -6.8426     1.2867
            b₆ -> Int_3         -.5033     .7955    -.6327     .5299    -2.1018     1.0953
            b₇-> Int_4           .1118     .1426     .7841     .4367     -.1748      .3984

                   Product terms key:
                    Int_1   :        Comments x       ZEngage
                    Int_2   :        Comments x       StatA0
                    Int_3   :        ZEngage  x       StatA0
                    Int_4   :        Comments x       ZEngage  x        StatA0

                   Test(s) of highest order unconditional interaction(s):
                            R2-chng         F        df1       df2         p
                   X*W*Z      .0074      .6149    1.0000   49.0000     .4367
```

Figure 7.23. Results from moderated moderation models predicting [?] with Zoella's comment engagement as primary moderator (W) and Status as secondary moderator (Z).

Finally, to confirm that the two apparent moderator variables impose their own individual moderation simultaneously, an additive multiple moderation model was tested. In this model X's effect is conditional on both W and Z. Also, as this model was deemed to give the greatest insight into the causal mechanism of comments on speech, Time was included as a covariate in this model. This was to statistically take into account the inherent relationship between Zoella's speech in each of the videos. There is the potential for the comment influence to accumulate over time. For example, Zoella's awareness of comments on video 2 and 3 may influence her speech in video 5 along with the comments on video 4, rather than the comment influence being limited to those posted on video 4 alone. In other words, the same percentage of comments may have different influential values depending on when they were posted, regardless of the two moderators. So, the final model is where all three antecedent variables (X,W,Z), their interaction terms (XW, XZ) and the covariate (C) are included. Figure 7.24 provides a conceptual diagram, figure 7.25 provides a statistical diagram, and figure 7.26 provides the results for [t].
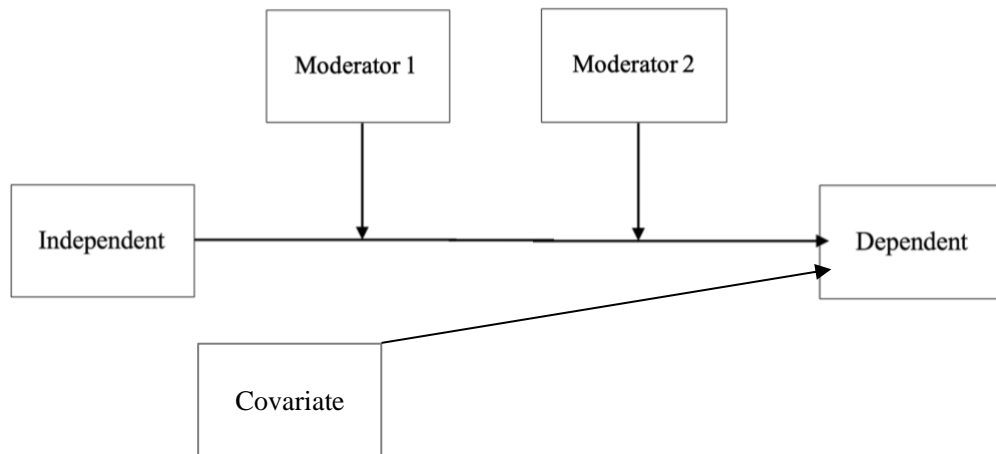
Figure 7.24. Conceptual diagram of an additive multiple moderator model with covariate. Recreated based on figure 9.1A (p.322) in (Hayes, 2018)
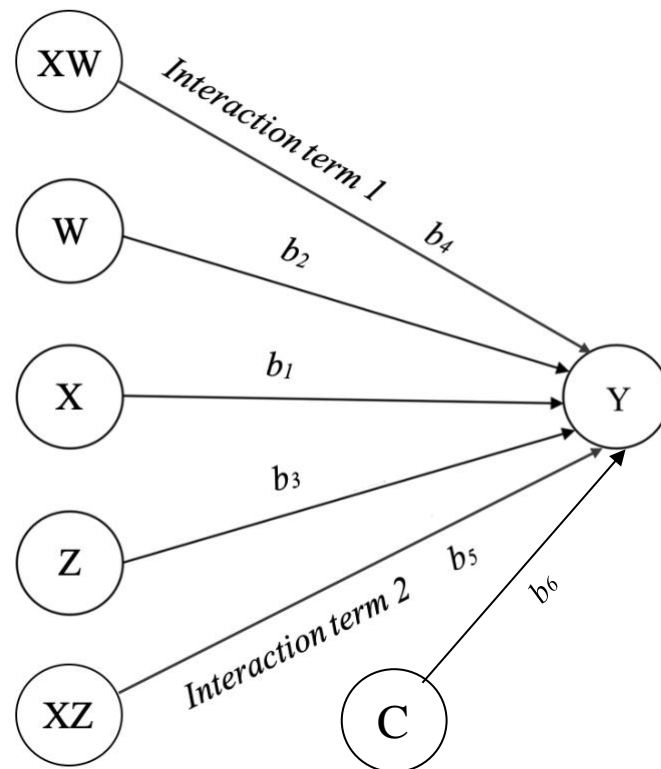
Figure 7.25. Statistical diagram of an additive multiple moderator model. Recreated based on figure 9.1B (p.322) in (Hayes, 2018)

In this model the antecedents (*b1*, *b2*, *b3*), the interaction terms (*b4* and *b5*) and the covariate (*b6*) are not statistically significant ($p > 0.05$). But the model where all the antecedents are included as well as both the interaction terms and the covariate is statistically significant ($p < 0.05$) (labelled "BOTH" in the "Test(s) of highest order interaction(s)" section in figure 7.26). See figure 7.27 for a

250

visual representation of this model. This model accounts for 10.6% ($R^2 =$ 0.1062) more variance in the data than the model including the antecedent variables alone. Thus, the null hypothesis that neither Zoella's comment engagement nor Status moderates the effect of Comments on [t] can be rejected. Also, although not statistically significant, the coefficients of *b4* and *b5* can provide further insight. Both *b4* and *b5* are negative. Here, it should be noted that *b5* reflects the Amateur condition even though Z is assigned as "StatP0". In an additive multiple moderator model *b5* determines the effect of X on Y as Z increases by 1 and W is held constant. As in the previous models, coding of Z's conditions is manipulated to maximise its interrogation. Here, coding Professional = 0 and Amateur = 1 allows *b5* to provide the coefficient for the Amateur condition, the condition that is of greater interest as it has been found to be statistically significant in the previous models. In summary, as Zoella's comment engagement increases, its negative moderating effect on [t] also increases, with this increase being greater in the Amateur condition.

In the next section, the results of the statistical analysis will be considered in further detail in conjunction with the literature review in section 7.1.4 to answer the case study research question.

```
*******************  *****************************************************************
Model  : 2          OUTCOME VARIABLE:
   Y  : AlvT           AlvT
   X  : Comments
   W  : ZEngage
   Z  : StatP0       Model Summary
                             R        R-sq        MSE           F        df1         df2           p
                          .6654      .4428    125.3216      6.6217     6.0000     50.0000       .0000
Covariates:
 Time
                    Model
Sample                             coeff          se            t           p        LLCI        ULCI
Size:  57           constant     56.2332      8.0948      6.9468       .0000     39.9741     72.4922
             b₁ ->   Comments       .3975       .8626       .4608       .6469     -1.3351      2.1301
             b₂ ->   ZEngage        .5731       .3430      1.6708       .1010      -.1158      1.2620
             b₄ ->   Int_1         -.0894       .0468     -1.9104       .0618      -.1834       .0046
             b₃ ->   StatP0        3.7772     10.4977       .3598       .7205    -17.3084     24.8627
             b₅ ->   Int_2        -1.4374      1.5828      -.9082       .3682     -4.6166      1.7417
             b₆ ->   Time           .1819       .0923      1.9703       .0544      -.0035       .3673

                    Product terms key:
                     Int_1     :         Comments  x         ZEngage
                     Int_2     :         Comments  x         StatP0

                    Test(s) of highest order unconditional interaction(s):
                             R2-chng         F         df1         df2           p
                     X*W       .0407      3.6497      1.0000     50.0000       .0618
                     X*Z       .0092       .8248      1.0000     50.0000       .3682
                     BOTH      .1062      4.7644      2.0000     50.0000       .0128
```

Figure 7.26. Results from an additive multiple moderator model predicting [t] with Zoella's Comment Engagement (W) and Status (Z) as moderators and Time as a covariate (C).
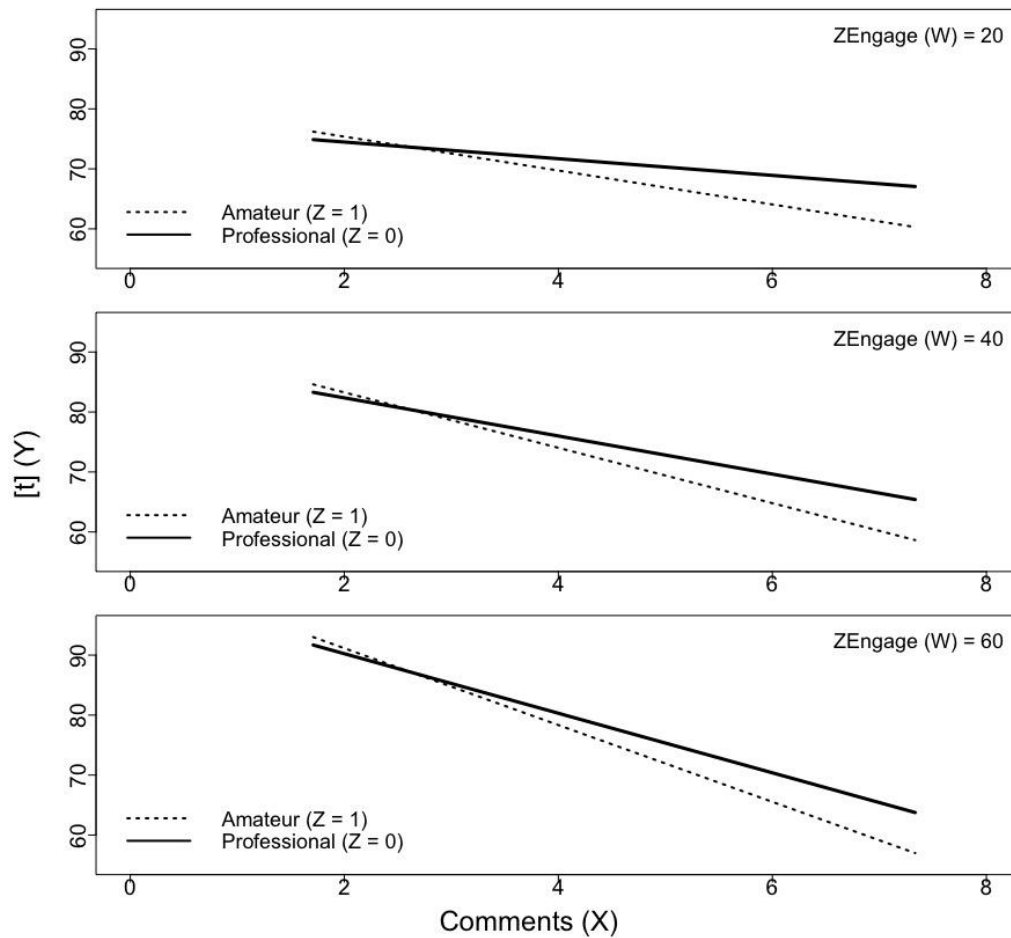
Figure 7.27. Visual representation of simultaneous moderation of the effect of comments (X) on speech [t] (Y) by Zoella's comment engagement (W) and Status (Z) while controlling for the covariate Time (C).

7.5 Case Study Discussion and Conclusion

To briefly summarise, through testing a series of moderation models it became evident that both Zoella's comment engagement and her status as an amateur YouTuber moderate the effect the comments have upon her speech. The more Zoella sees what the comments say, the more influence they have over how she speaks, but what the comments say has influence on her speech when she is an amateur and not when she is a professional. Here, it should be reiterated that we do not know when or how much Zoella reads the comments, therefore evidence of her responding to comments on her own videos is being used as a proxy. It is also evident that these moderation effects are separate from one another, as the

252

results of the models detailed in section 7.4.3.3 evidence. Therefore, the moderating effects of these variables overlap and accumulate but are not interdependent.

The next element to consider is the direction of the effects. Focusing on the patterning of the data, rather than the effects as displayed in the statistical graphs, as the comments decrease, and Zoella's comment engagement decreases, [ʔ] decreases and [t] increases. In regard to status, when an amateur [t] decreases and [ʔ] increases as the comments increase. Because the comments are those which describe Zoella as 'cute', from this result it could be assumed that [ʔ] indexes youthful femininity and that [t] does not. Why this may be so requires more careful thought.

In regard to youthful femininity and the possible indexical rationales behind the variants of word medial /t/, I posited two suggestions in section 7.2. The first was that [ɾ] could index youthful femininity. As detailed in section 7.1.2.3, [ɾ] is the typical realisation of word-medial trochaic /t/ in General American English and its use in British English (both RP and SSBE) is rare. Therefore, one would assume that increasing use of [ɾ] would reflect increasingly engaging with an American audience. Connecting this to youthful femininity, I would argue that an American audience could view a British YouTuber catering to them through her speech as endearing. However, as stated in section 7.4.1, [ɾ] was negligible (65 tokens, 2.9%) and so was discarded prior to statistical analysis. At first, this lack of [ɾ] may seem counter to the observations made in the digital ethnography (detailed in chapter 6). Then again, with so many other ways of catering for an American audience and with this catering possibly being more noticeable or explicit, maybe there was little motivation or necessity to cater to an American audience through speech also. The second suggestion was that [t] could index youthful femininity. As is detailed in section 7.1.3.1, Americans associate [t] with Britishness, and one cultural stereotype is that the Americans find the British endearing, as well prissy. Therefore, use of [t] may be viewed as cute by her American audience as it emphasises Zoella's Britishness. However, it transpires that [ʔ] patterns with the comments. One argument for the way [ʔ] could index youthful femininity is via approachability and friendliness although

this is the variant that's predicted to be stigmatised by the British (section 7.1.3.4) and has little social meaning to Americans.

But so far, an important element to this dataset has been overlooked: the percentage of youthful femininity comments decrease. In other words, her commenters describe her as cute less over time. This patterning of the data also raises the question as to whether the youthful femininity comments were welcomed by Zoella because there is little evidence in the data that she resists the declination in her pronunciation of word medial /t/ as [ʔ]. Further, the analysis above focuses on how the comment and speech data pattern with one another but not how the moderators pattern with them. In this regard, little further insight can be gained through examining Zoella's comment engagement. It is evident that this is the mechanism through which the comments affect speech, with greater comment engagement leading to the youthful femininity comments having a greater effect, but this does not inform our analysis of why [ʔ] and the youthful femininity comments pattern. In comparison, that the amateur condition moderates the effect that the comments have on speech (but the professional condition does not) warrants greater attention.

The moderation of the comments effect on speech by status encourages youthful femininity to be considered within the context of being an amateur or a professional on YouTube. First, it is reasonable to presume that an amateur aspires to become a professional or at least be seen as competent. Here, a YouTuber being referred to as 'cute' may have the positive connotations of youth and femininity, of course, and physical attractiveness, qualities that are expected of a beauty and fashion YouTuber. Also, describing someone as 'cute' may be seen as suggesting they are approachable or friendly, as [ʔ] has been found to be. But equally, this could be viewed as patronising or condescending, and suggest this YouTuber should not be took seriously, or is incompetent. It has become evident that success on YouTube, especially within this genre, is heavily dependent upon apparent expertise and so youthful femininity may be a less desirable trait. Further, it is difficult to envisage how describing a professional YouTuber as 'cute' could be positive, an intuition that is supported in this study through there being fewer youthful femininity comments in the professional

condition relative to the amateur condition and that the percentage of these comments plateau suggesting youthful femininity is incongruent with this career stage.

Reframing the analysis to focus on how [t] may index qualities that diverge from youthful femininity, rather than focusing on how [ʔ] indexes youthful femininity, provides greater insight. Previous literature consistently evidences that [t] indexes qualities such as intelligence and education for both American and British listeners (section 7.1.3). This naturally links also [t] to other social qualities, such as competence and professionalism, and being articulate. These are social qualities that someone aspiring to be a professional YouTuber is likely to desire. That [t] is the dominant pronunciation and the percentage of comments referring to youthful femininity stabilises after Zoella becomes a professional YouTuber, and that the professional condition does not moderate the effect of the comments on speech, aligns with this hypothesis. This pattern suggests that once Zoella has become a professional, her competence and expertise being affirmed by the industry, she is less concerned with the feedback in the comments or views their content as less important in comparison to the amateur condition where she was developing her skills, knowledge and reputation, and presumably highly motivated to please her audience. Therefore, there may not actually be an indexical relationship between [ʔ] and youthful femininity but simply [ʔ] is one of the pronunciation options available and [t] is tied to other indexical meanings that Zoella wants to portray.

This explanation does not incorporate observations of the geography of the two dominant imagined audiences or how these change over time, however. On the one hand, who her imagined audiences are becomes somewhat redundant as it just so happens that American and British perceptions of use of [t] in regard to professionalism and competency are in agreement. On the other, Zoella's imagined audience changing may play a subtle role. [ʔ] is not used by Americans and if Zoella had a sense of this it may be that her audience becoming apparently more American was an additional catalyst for increasing use of [t]. Equally, [t] is the more standard pronunciation and prior research has found that as an audience becomes more geographically diverse more standard features are used with the

aim of being inclusive (e.g. Pavalanathan and Eisenstein, 2015; Shoemark *et al.*, 2017, although this was found in written data). Regardless as to whether any of these motivations are present or not, the apparent causal relationships between the comments, Zoella's speech, Zoella's engagement with the comments and Zoella's status still hold.

Finally, the apparent role that YouTube's auto-generated closed captioning facility plays in YouTuber visibility, and the resulting influence this may have on pronunciation, should be acknowledged. As first described in section 2.3.3, Sophie Bishop (2019) argues that YouTube's use of auto-generated caption data in their search algorithm encourages several YouTuber behaviours that are used in the hope of creating caption data that results in their video receiving greater visibility. One of these behaviours is "carefully and crisply pronouncing keywords" (ibid, p.27) to encourage an accurate auto-generated caption (however, it should be noted that no linguistically based analysis was reported). Therefore, if a YouTuber believed that careful pronunciation was required to maximise their exposure it is highly likely that they would favour [t] over [ʔ]. However, there are a number of reasons as to why this potential explanation has little relevance to this data. First, tokens that were brand names or branded product names were removed from the dataset, although it is possible that the influence that the desire for accurate captions has on speech pervades beyond keywords. Second, while videos began to have auto-generated captions back in 2009 (Harrenstien, 2009), that auto-generated captions are input into YouTube's various algorithms was not made public knowledge until much later, possibly around 2017 (e.g. Kaver, 2017). We cannot assume that YouTube's most successful partners, such as Zoella, were not privy to this information earlier than the more typical user, but if they were it is unlikely that this was several years earlier. Actually, the altering of pronunciation to generate accurate captions would go some way to explaining why the professional condition does not statistically significantly moderate the effect of the comments on speech, because in this scenario Zoella's speech would not be motivated by the content of the comments. Still, the conclusions above in regard to the amateur condition do not need altering.

To summarise, to answer the case study question, Zoella responds to the feedback in the comments by diverging from them through her use of [t]. The degree of divergence depends on how much Zoella engages with the comments, and how important the comment content is, as a result of her on career status.

7.6 Case Study Limitations and Future work

Now that the case study is complete there are several limitations that should be acknowledged in addition to those outlined in section 4.4. The most notable limitation to this study is the use of Zoella posting comments on her videos as a proxy for her actually reading the youthful femininity comments. While including a variable that details the number of comments Zoella read on each video would probably result in a more exact understanding of this moderator's moderation of the causal relationship between the comments and her speech, it would not be possible to collect such a data set retrospectively. However, it is reasonable to presume that Zoella would not post a comment herself without looking at what her viewers were saying in their comments. So, we can have confidence that this variable is not counting a behaviour that doesn't exist, and actually, one can argue that using Zoella's comment engagements as a proxy for her reading the comments is a more conservative estimate of how her reading the comments would moderate the effect of the comments on speech.

Second, in sociolinguistic studies it is typical for the data points of the consequent variable to be individual speech tokens. In doing so, linguistic factors, such as following place of articulation, grammatical category, and lexical frequency, can be tested as antecedent variables and their influence over the consequent variable considered. Although I would argue that the great consistency in the language content of the videos across the dataset reduces the variability of the words used, and thus constrains these linguistic factors somewhat, some sociolinguistic researchers may want confirmation of the role

that these linguistic factors play in this case study. This could be considered as a piece of future work. However, mixed-effects logistic regression (the more typical statistical approach used in sociolinguistics) would have to be used because Hayes (2018) discourages the use of mediation and moderation analysis on a dichotomous consequent variable.

Finally, a criticism that can be made of many linguistic studies can be levied at this study. This is a case study of an individual and it is questionable how transferable the conclusions in this study are to other YouTubers, types of video content and audiences, or other public video sharing sites. In response I would argue that very few YouTubers across the globe have been as successful as Zoella and so it would be unwise to generalise these conclusions beyond this particular instance anyway. And, of course, this is the first study of its kind and we will have to see if future work uncovers the same or similar causal relationships. Again, future work could examine YouTubers who have had different career journeys, are from different genres and so produce different types of content, whose viewers appear to contribute different types of feedback, as well as speakers of different accents or languages, in comparison to Zoella to interrogate these essential elements that made up this case study. For an example, see section 8.1.1.4.

7.7 Reflecting on Thesis Research Questions

The work reported in this chapter contributes to answering the final thesis research question: "What statistical approaches could be used in studies of speech in online public video considering that time can be operationalised with greater granularity?" To recap section 3.4, real-time sociolinguistic studies of single speakers have generally only had a few data points for time and in such studies the statistical analysis has centred on comparing phonetic realisations at different points in time using tests such as ANOVAs and t-tests, but time itself is rarely included as a variable in the statistical analysis. Where time has been included as a variable in statistical analysis it has predominantly been a

categorical variable, e.g. year, decade, or before and after particular life events. Also, these different categories of time are typically tested using an exploratory, trial and error approach, effectively carving out the data in different ways to see which operationalisation of time is most predictive of speech along with other variables in mixed-effects regression modelling. Further, when time has been included in such models there is little evidence that its interaction with other variables has been considered. Therefore, to date the role of the variable time has generally been limited to being one of the many antecedent variables that is tested to find the model that best fits the speech data and thus predict it.

As is evidenced in this case study, YouTube has the potential to provide many more data points for the variable time. One result of this is that it may be rationalised as a continuous, rather than categorical, variable. However, testing time as a continuous moderator was not fruitful in itself in regard to being able to explain the data. But the result of the Johnson-Neyman test did encourage two other variables to be tested (Zoella's comment engagement and status) which lead to an interesting set of findings. Therefore, in the end, time was operationalised as a categorical variable (status as amateur or professional) and so this study did not deviate from typical sociolinguistic practices to date.

Before reflecting on the relevant thesis research question, some similarities and differences between the typical use of mixed-effects regression analysis and moderation analysis in PROCESS should be highlighted. First, both are regression based approaches and both include the testing of interaction between variables (because the terms moderation and interaction are synonymous). However, they essentially reflect different study designs and approaches. Mixed-effects regression analysis is exploratory, where multiple predictor/antecedent variables are each included and removed in turn to identify the best fitting model. Whereas moderation analysis only allows one X variable, reflecting a more hypothesis driven approach. Further, moderation analysis assumes that variables have fixed-effects whereas variables can be designated as random-effects in a mixed-effects regression analysis.

This thesis research question looks to consider the wider impact of a specific variable of interest (time) being operationalised in a specific way (continuous). First, the statistical role of time in this study contrasts with that in most other sociolinguistics research. Namely, herein time (and subsequently identified aspects of time) is tested to see if it may be a moderator (a mechanism of change) of a focal antecedent variable rather than being one of many antecedent variables. There is evidence to suggest that moderation (or interaction as it is also known) is used across linguistics research (Baayen, 2013) but it appears that this statistical approach is novel for sociolinguistic studies that focus on speech change across time particularly those which are case studies of an individual. Further, although it should be acknowledged that the literature review is not exhaustive, more complex models have not been reported either, such as the additive moderator model described herein.

Second, with more data points along the dimension of time, other time-based variables can be considered. In this case study this was the youthful femininity comments which would not have been a viable variable if only a few videos were analysed. Further, the youthful femininity comments taking on the role of the focal antecedent evidences how other time related variables becoming available prompts novel kinds of research questions.

Third, more data points along the dimension of time allows a speech consequent variable that is typically dichotomous (a phonological analysis where there is a choice of one of two variants) to become continuous (percentage of use) also. Similarly to time (see section 3.4.3), this variable's operationalisation does not pose the need for statistical innovation in itself because different variations of mixed effects regression modelling can be used for continuous (mixed-effects linear regression) and categorical (mixed-effects logistic regression) consequent variables. However, Hayes (2018) discourages the use of mediation and moderation analysis on a dichotomous consequent variable, and so time could not be tested as a moderator if the speech variable was not continuous. Of course, using the percentage of [t] and [ʔ] as the consequent variable means this is technically a different variable and thus asking a different research question in contrast to considering whether each individual instance of word medial /t/

results in a [t] or [ʔ]. In other words, rather than asking whether [t] or [ʔ] is more likely to be used in a particular instance, this case study is asking when [t] / [ʔ] is more likely to be used more and when it is more likely to be used less.

Finally, it should be noted that the use of moderation in this thesis is reflective of a small aspect of this statistical approach and so merely scratches the surface of this topic. Whether the focal antecedent and potential moderator are dichotomous, categorical or continuous, and how many potential moderators there are all impose different requirements onto performing and interpreting the results of the statistical analysis, let alone then also considering the presence of mediators as well. In other words, the description of moderation analysis given herein is limited to what is relevant to this case study. Therefore, the contribution made in this thesis is to give a basic argument for the use of this statistical approach, provide an initial illustrative example, and point to the resources that researchers can use to find out more.

7.8 Summary

To summarise, the core of this thesis (chapters 4 to 7) has reported an exploratory case study that aimed to answer the question 'Does the direct written feedback received through the commenting function influence a YouTuber's speech?'. Through comment analysis (chapter 4) and speech analysis (chapter 7) that was informed by a digital ethnography (chapter 6) and an examination of the causal relationships between these and other variables using moderation analysis, it can be concluded that the feedback in the comments appears to influence Zoella's speech at certain times in her career and the degree of that influence depends on how engaged she is with the comments.

In the next chapter, all four thesis research questions will be reflected upon and addressed by collating the experience and learnings gained through performing the case study. As a result, a set of working guidelines for future sociolinguistics research practice in online public video are proposed.

Chapter 8.

Discussion and Conclusions

> *"Most basically, a method is nothing more nor less than a*
> *means of getting something done"*
> (Markham, 2006, p. 50)
> *"the capabilities and limitations of methods are revealed in*
> *practice."*
> (Hine, 2005, p. 21)

The research presented in this thesis explored how speech can be studied within the context of YouTube. Earlier in the thesis I established that examining speech in CMC research is a rarity because typical research practices do not easily transfer from studies of offline speech and language or language-online to studies of speech-online. In this regard, four key methodological issues that need to be addressed were identified: i) Formulating Research Questions, ii) Research Ethics, iii) Selecting Linguistic Variables, and iv) Statistical Analysis. Then, I conducted a case study which acted as a vehicle through which these four key methodological issues were investigated. By recording and reflecting upon the decision-making processes and research practices that were required as a part of the case study, I generated insights into how studies of speech in online public video taking a sociolinguistic approach can be conducted.

This thesis set out to make a methodological contribution in the form of guidelines for research practice that can benefit those who aim to analyse speech in online public video. This chapter collates and reflects upon all the learnings that I have gained from performing the Zoella case study in regard to research decision making and practices for YouTube data. In this chapter a response to each of the four **thesis questions** is given, the insights that these responses are based on are reviewed, and relevant guidelines are described in turn (sections 8.1, 8.2, 8.3 and 8.4). I also report on research studies that were attempted before the Zoella case study and that are planned for after this thesis is complete (section 8.1.1). The insights gained from this new information further supports some of the guidance that I make in response to thesis question 1 as a result of conducting the Zoella case study. Section 8.5 will outline future work in regard to monitoring and reviewing the guidelines (8.5.1), and the development of resources and tools that would assist in conducting sociolinguistics research on YouTube (8.5.2) as well as the implications for the thesis learnings for research into other kinds of YouTube data and Broadcast Media when a sociolinguistic approach is took (8.5.3). Finally, the thesis will be summarised with some concluding remarks in section 8.6. In addition to relaying the guidance in this chapter, I have also drafted a number of resources for researchers to use and these will be referred to throughout (see appendices 3, 4 and 5).

8.1 Reflections on Thesis Question 1

The first thesis question is: 'What strategies or approaches could a researcher use to i) find sociolinguistically interesting YouTube data and ii) formulate appropriate research questions for that data?' My response to thesis question 1 is that researchers i) should not go searching for data but utilise their own and others experience and knowledge of YouTube, and then ii) systematically establish the data's qualities so that these naturally impose boundaries upon what research questions can and cannot be answered.

This thesis question was identified as a result of reviewing different types of sociolinguistic data (tailored, found, and raw) and their typical associated practices for data collection and formulating research questions, as well as comparing these to said practices when researching online written data. There is currently little documentation of these practices for YouTube data across the CMC literature and so four key descriptors of YouTube (massive, heterogeneous, accidental and disordered) were used to structure the theorising of challenges that a researcher may experience in trying to find data that is appropriate to research and formulate research questions to ask of it. It became evident that formulating a research question to ask of found or raw data requires a negotiation of prior literature and the data's qualities, with the research question that is eventually formulated representing a middle ground between the two. However, it is difficult for a researcher to have a clear grasp on what data qualities the video data is likely to possess because the search facilities do not aligned with sociolinguistic interests. Without these data qualities naturally imposing limitations and boundaries, a researcher may lack direction or inspiration.

To answer the first half of this thesis research question, through performing the Zoella case study it was evident that I was reliant upon my own prior experience and knowledge of YouTube. By considering the data of specific YouTubers and specific types of videos that I was already familiar with the pool of potential data was reduced to what felt like a manageable amount. In regard to the second half of this thesis research question, mapping out the data qualities of the videos and then seeing if they mapped on to those that would be required to answer a pre-formulated research question was more efficient as I was familiar with the data somewhat. My thoughts moved between evaluating the literature and evaluating the YouTube data, iteratively refining my definition of each to design a case study that was essentially a compromise of the two. Further confirmation of these insights has been gained through planning future projects as well as projects that were abandoned prior to the Zoella case study, as is detailed below.

8.1.1 Plans and Abandoned Projects

So far, the content of this thesis suggests that its main contribution (guidance for researchers who wish to conduct sociolinguistic research using YouTube data) has only been informed by the experience and learnings gained from performing the Zoella case study. However, it is important to highlight that I have also generated a lot of insight through i) formulating research questions that I then chose not to research because I could not find appropriate data, and ii) planning other case studies where identifying the data was straightforward. These insights have informed the guidelines just as much as the Zoella case study, and so I will report on them now.

First, two abandoned case studies (Scottish Referendum 8.1.1.1 and Response Videos 8.1.1.2) are described. The key learning here was that a researcher cannot solely be guided by the literature and must take advantage of their own knowledge of YouTube when designing a research study. Second, two research studies that are planned for the future (Doug DeMuro 8.1.1.3 and Co-optional 8.1.1.4) are described. The YouTube data for both of these studies was happened upon by chance as a result of finding out about the YouTube content that someone close to me knew well.

8.1.1.1 Scottish Referendum

As a part of applying for a PhD position I began to design the 'Scottish Referendum' project. I was inspired by Lauren Hall-Lew's findings that politicians' pronunciations of key words indicated political alignment (e.g. Hall-Lew, Coppock and Starr, 2010; Hall-Lew, Friskney and Scobbie, 2017). I wanted to see if these studies' conclusions held in the speech of the general public, essentially using YouTube as a repository of speech samples. The premise was to collect YouTube videos that discussed the Scottish referendum from multiple YouTubers, comparing pronunciation from Remainers and Leavers, and also referendum and non-referendum content. I planned the project as though I was collecting tailored data and presumed there would be an

abundance of data on YouTube for me to select from. While sound in principle, in practice this proved impossible. I could hardly find any relevant data via the YouTube search box. The results of my many searches were awash with clips from news broadcasts, comedy and political panel shows, and daytime chat shows. Very few results were videos from YouTubers, and so abandoned the project.

To summarise, the planning and abandoning of this research clearly illustrates how a researcher that approaches YouTube data, or other online public video, with a preformulated research question based on the literature alone may not be able to scale the hurdle that is finding appropriate research data.

8.1.1.2 Response Videos

After abandoning the Scottish Referendum project I rethought my PhD research plan. My next plan for this thesis was that it would contain more than one case study and that these would be far smaller than the Zoella case study that is detailed herein. Instead, I planned three case studies, i) Zoella (the analysis of uptalk only), ii) Co-optional (described in section 8.1.1.4) and iii) Response videos. The premise was to examine speech change in different interaction contexts (which create different interaction dynamics as a result) that can be found on YouTube: i) within a video (Co-optional), ii) within a video interface (Zoella), iii) across videos (Response videos). The idea behind the response video project, and why it was not successful, is detailed here.

Prior to 2013, it was possible to make connections between videos. A user could upload a 'response video' and attach it to another, already existing YouTube video (referred to here as the 'starter video' for clarity). Multiple response videos could be attached to a starter video, and small thumbnails for these response videos were displayed underneath the starter video. The intention was to create video threads, so that a dialogue could be created between videos. Indeed, Adami (2009) and Pihlaja (2011) both performed qualitative analyses of interaction across response videos, focusing on sign-making and metaphor, respectively. However, the video response functionality was removed in 2013

because YouTube found that the Click-Through Rate (the percentage of users that clicked on a response video after watching a starter video) was four out of a million users (Panzarino, 2013; YouTube Team, 2013).

Yet, the concept of the 'response video' persists. First, response videos that were posted before the response video function was removed still exist on YouTube, of course. But there is also a notable body of videos posted after the function was removed that contain "Response Video" in their titles, often also quoting the title of the video that is being responded to. Therefore, I imagined a study that considered convergence (if speakers' speech became more like the person that they were directing their speech to in order to demonstrate affinity and encourage relationship building) or divergence (the opposite to convergence) (Giles, Coupland and Coupland, 1991) would be interesting. Based on what I had learnt from attempting to find data for the Scottish Referendum project, I believed that finding appropriate data would be more straightforward because I had chosen a general research topic and my inspiration had come from YouTube, rather than having a preformulated research question that was inspired by the literature alone. However, there were two barriers to finding data for this project. First, without the infrastructure connecting start videos with their response videos, I had to begin by searching for response videos, which was time consuming. Second, very few response videos actually reply to their starter videos in a direct dialogue. Rather, most response videos are actually mocking the YouTuber in the starter video, often with clips from the starter video that the response video is retorting. Therefore, speakers in response videos are actually directing their speech to their viewers and so it would not be possible to study convergence or divergence.

To summarise, finding appropriate data to research requires more than merely an awareness that a type of video exists. Rather, a researcher must be certain that they know this genre or type of video well in terms of its style and typical content across multiple YouTubers. Also, attempting to research an interface feature, function or design that actually no longer exists is futile.

8.1.1.3 Co-Optional

After these abandoned plans I adjusted my approach so that I used YouTube data that I was aware of as the starting point for designing a research project. As was stated in section 8.1.1.2, my initial plan for this thesis was that it would contain more than one case study. However, as the results of the uptalk analysis began to be established and I reflected upon my research decision-making and practices thus far, it became clear to me that I had only just begun to gain enough insight to be able to make recommendations to others in regard to one type of study design (longitudinal panel study of a single speaker). I decided that the Zoella data needed more examining and as my work on this case study grew and grew, the Co-Optional case study shrunk and shrunk until I decided that it would be more appropriate for the thesis to consider the Zoella data alone. However, this other case study is now planned for the near future[11].

This study was designed to take advantage of 'The Co-Optional podcast'[12]. This is a weekly video where four YouTubers, three regulars and a guest, stream from different locations to discuss computer/video games. The interaction is computer-mediated through video streaming technologies, creating a split-screen video. It can be argued that the three regular YouTubers are representatives of a Community of Practice (CoP). A CoP is a group of people who share a passion or concern and learn in regard to that topic as they interact with each other, a term coined by Lave and Wenger (1991) within a learning/workplace context and took up in sociolinguistics to become a fundamental concept. Speech style can also be studied in a CoP as its members may speak similarly and so speech can be an identifying feature of being a member of that group. Examples of work where the concept of CoP has been used include Eckert's (1989) 'Jocks' and 'Burnouts', and Mendoza-Denton's (2008) Latina girl gangs. Coupland (2007) theorises that CoPs develop their speech styles through members regularly interacting and their speech converging during the interaction, so that over time these convergences become stable. Converging is when a speaker adjusts their

---

[11] To be explicit, to date I have been granted ethical clearance by the Institutional Review Board and have downloaded the relevant video data. However, I have not begun orthographic transcription.

[12] Example episode: https://www.youtube.com/watch?v=D1x5Lw6eloU

speech to be more similar to the speech of their interaction partner and may be used to demonstrate affinity and encourage relationship building, according to Communication Accomodation theory (Giles, Coupland and Coupland 1991). However, to date there have been no studies investigating this theory, probably because of the practicalities of collecting an appropriate amount of data across a suitable time period. With each podcast lasting at least 2 hours 30 minutes, and a podcast being posted once a week consistently from January 2012 to date, this archive of interactions presents the opportunity to examine Coupland's (2007) theory. The research question would be something akin to: "If the same speakers interact on a regular basis and accommodate during this interaction, do these accommodations become stable and thus define the speech style of the community of practice?"

I became aware of The Co-Optional podcast via my younger brother who is an avid fan of the podcast, the three regular YouTubers and many of their guests. When I was looking for inspiration for my PhD project proposal (May 2017) by asking family and friends what they watched on YouTube, I quickly began to formulate a research question in regard to speech convergence in response to my brother reminding me of the Co-Optional podcast. In addition to him bringing this YouTube series to my attention, my brother also provided important insights and information as I planned this case study, most notably he immediately knew which guest has been the most reoccurring on the podcast. It was this information that helped me identify the dataset: 8 videos (24:05:14 footage, across a period of 3 years, 4 months and 14 days) where the same four YouTubers are present.

Here, the conclusion is similar to that of the subsection above and thus further supports the following guidance for working with YouTube data in sociolinguistic research: Utilising your own and others knowledge and experience of YouTube can reduce the pool of potential data so the researcher is more likely to find sociolinguistically interesting data and be able to assess its data qualities more efficiently.

8.1.1.4 Doug DeMuro

I stumbled upon the data for my next study[13] serendipitously and I formulated the research question almost instantly as a result of recently being immersed in the relevant literature.

I was at home with my husband and daughter in early January 2020. He was in the kitchen cooking dinner with YouTube videos playing (loudly) on his ipad, a typical practice in our house. My daughter and I were in our living room, next door to the kitchen. I heard (represented orthographically and then using the International Phonetic Alphabet):

"Thththththththththththiiissssssssssssssssssssssss is a (car name I can't remember)"[14]
" [ð:::ɪs:::] is a (car name I can't remember)"
I jumped up, went into the kitchen, and asked my husband what he was watching. He told me:

"Thththththththththththiiissssssssssssssssssssssss is Doug DeMuro."
" [ð:::ɪs:::] is Doug DeMuro."

But not only did he mimic the exaggerated, elongated pronunciation of "this", I now know that when he was waving his hands around (placing them on his hips at the start of "this" and then sweeping them up and out so his forearms are parallel to the floor, palms up, by the end of the word) he was also mimicking Doug's typical gestures. At my dumbfounded look, he went to Doug DeMuro's YouTube channel, and almost every one of the most recent videos (posted in 2019) that he clicked on had this same introduction.

---

[13] To be explicit, to date I have not applied to my Institutional Review Board for ethical clearance, nor have I downloaded any video or comment data. However, I have spent some time surveying Doug's YouTube channel in order to assess the data's qualities, identify specific videos of interest and thus establish that a study would be possible.
[14] See https://www.youtube.com/watch?v=pTb_cOYQctc for an example.

I asked myself "how did this introduction come about? And "Could his commenters have played a role in its formation?". I decided to investigate further: Doug DeMuro is a YouTuber who makes videos about reviewing cars, assessing their "quirks and features" and giving them a "Doug score"[15]. The first video that he began with "This" was posted in April 2016 but Doug's use of this introduction appears to be sporadic until mid 2017 where it becomes the standard introduction to his videos. This is reflected in the thumbnail images which all have him stood behind the bonnet of a car, elbows bent at the waist, palms up and parallel to the floor, as well as the "this" compilation at the beginning of his video celebrating 1 million subscribers[16]. Further, upon an initial assessment of his videos, the gesture of having his hands on his hips and then moving his hands up and out to a more open posture is used inconsistently and is not necessarily in time with the speech. He appears to experiment with the gesture, for example starting the gesture with his hands clasped together in front of his torso or stuffed in his pockets. Also, from a brief scan of the comments it is evident that this is a source of discussion amongst his viewers. The reading of one comment (here rephrased in an attempt to anonymise) confirmed to me that this would be an interesting and sociolinguistically relevant data set: "Watching this video years after its posted and the "this" is longer than yesterdays!".

From initially surveying the data, I plan to formulate a research question akin to that used in the Zoella case study that is reported herein ("Does the direct written feedback received through the commenting function influence a YouTuber's speech?") and conduct the research through a similar method to that taken in the Zoella case study. However, in contrast to the Zoella case study, the comment analysis would not need to aim to identify social qualities to then guide the selection of the linguistic variable to study. This is because I have already identified the speech feature/s of interest because they drew my interest to the data in the first place. Also, length of fricatives does not currently index a social identity (based on my intuitive knowledge of English only) but are highly noticeable and lay people would be able to discuss them explicitly, as seems to

---

[15] The Doug DeMuro YouTube channel:
https://www.youtube.com/channel/UCsqjHFMB_JYTaEnf_vmTNqg
[16] https://www.youtube.com/watch?v=usmTkNC5Jh4

be evidenced in the comments. The speech features of interest are the two fricatives in the word "this" (the first and last sounds in the word, [ð] and [s] respectively) and their duration.

To summarise, the planning of this research further supports a suggestion that I would make as a result of performing the Zoella case study: That an appropriate strategy for finding sociolinguistically interesting YouTube data is to take advantage of your own and others knowledge and experience of YouTube. This, effectively, massively reduces the pool of potential data, namely to specific YouTubers and specific types of videos, and an amount of data that would be far more manageable in regard to assessing data qualities.

8.1.2 Guidance for Formulating Research Questions

The knowledge and insights gained from asking thesis research question 1 has resulted in the creation of the resource in appendix 3. This resource is designed to assist a researcher in formulating a research question to ask of online public video data, whether this is through interrogating their own YouTube experience and knowledge or having been struck with inspiration after stumbling across interesting data. From the experience I gained conducting this thesis' case study, I believe that the pragmatic way to formulate a research question for such data is through efficiently ascertaining the data's qualities, for these constrain what research questions can be asked. Therefore, these resources guide a researcher in summarising the content of a specific channel. Summarising the content may help a researcher to narrow down the data to that which will be most pertinent to answering a research question that they already have in mind. Or, if the researcher does not already have a research question in mind, help them to sift through the content of a specific channel so that a dataset that both inspires a research question and would be needed to answer it rises to the surface. The intention here, to paraphrase Van Herk's (2013, p. 165) comments on working with raw data, is to turn this "bunch of words and stuff" into something we can analyse. Of course, when attempting to formulate a research question the researcher should be in dialogue with the literature as well to find a middle

ground between it and the data's qualities. So, in opposition to tailored data where "much of our data collection is blind to eventual purpose" (Van Herk, 2013, p. 165), these resources assist researchers in simultaneously defining the specific videos to collect data from and the questions that will be asked of it.

The first page of appendix 3 is designed to record essential information about a YouTuber and their channel. How much of the form is completed is up to the researcher and some questions may require more investment in regard to time watching a channel compared to others that may be easy to answer with a quick internet search. The same social and geographical factors that are important in more traditional / offline sociolinguistic research are included along with factors that are important from the perspective of analysing social media such as whether the YouTuber could be considered A-List. Of course, as described in section 6.1, celebrification is a continuum not a binary distinction of 'celebrity' and 'non-celebrity', but establishing if these indicators of being an A-List are present gives a researcher further insight into the YouTuber's online history. Further, these pages can be continually added to as the researcher gains more information through becoming familiar with the data. For clarity, it should be highlighted that this resource is not intended to be a substitute for an in-depth ethnography, although may assist a researcher in deciding whether applying this method may be fruitful and also in regard to what time period or how to limit their field (for a reminder of these issues see section 6.3.2). Also, because the YouTubers history is the primary focus it may be beneficial for a researcher to map out all the events that are noted along a timeline. This layering of data may reveal interesting periods to contrast or moments of change where before and after can be compared. Finally, it may be that a researcher examines several YouTubers' channels because they are believed to align or contrast in interesting ways.

The next step (page 2) summarises the types of videos the YouTuber posts on their channel. Many YouTubers post different types of videos on their channel. For example, in addition to hauls Zoella's main channel also contains the following video types: vlogs, question and answers, tours (e.g. room tour), monthly favourites, tutorials (e.g. makeup, hair, baking), outfit diaries and

lookbooks, challenges (e.g. my brother does my makeup, my boyfriend does my clothes shop), storytime (recounting personal experiences), and reactions (reacting to other videos, home videos and photos from childhood). Within each video type one can expect some consistency across factors such as overall structure, the items used or shown, the activities performed, the people present or referred to, the physical location, amount and type of editing, camera angles, and potentially linguistic content. Establishing what video types a YouTuber typically posts may help a researcher to identify a video type to assess further using page 3. Or finding out how different video type contrast may reveal factors that potentially influence speech and so may be fruitful to research in addition to categorising the channel's content. Further, each YouTuber will bring their own unique style to these videos which may introduce other elements that may be interesting to research.

Finally, the third page is designed to assist the researcher in efficiently establishing the data qualities that individual videos possess. Each row should refer to a factor such as those assessed on page 2 (e.g. the people present, the physical location, whether audio is live or overlayed) or others the researcher is interested in, and each column should refer to a video from the channel, with each row-column crossover providing a space for notes about that specific factor in that specific video. Here, the aim is to select videos that are consistent across certain factors and differ across other factors in order to establish the video dataset to be analysed. It may be a video type has been identified as interesting as a result of completing the form on page 2. For example, a researcher may want to see if a YouTuber's speech differs between make-up tutorials where sponsored products are being used to those where the product choices are not sponsored. In order to do so a researcher would want to ensure as many other potentially influential factors to be as consistent as possible. Comparatively, a researcher may wish to compare a YouTuber's speech in videos that are different video types but are consistent in regard to some other factor. For example, a researcher could identify that they are interested in comparing a YouTuber's speech in vlogs to question and answer videos when both include their partner. Essentially, the researcher is encouraged to establish a draft set of inclusion/exclusion criteria and then assess the videos on the channel of interest.

If it transpires that the data available does not match this inclusion/exclusion criteria then the researcher has the opportunity to revise it. Further, the strategy of identifying video types and then examining one or two of these closely is to reduce the amount of data that needs to be considered. With some channels possessing hundreds or thousands of videos it would be inappropriate to expect a researcher to assess every video that has been posted. Multiple copies of page 3 may need to be used, depending on how many videos are being assessed.

This resource is not prescriptive and can be edited and used however a researcher sees fit, both with YouTube data and data from other online public video sharing platforms. Although to give a word of advice, it would be advantageous to fill it in digitally so that videos or other sites that are being referred to can be hyperlinked, and colour coding or symbols could help ensure the notes are concise. In summary, this resource supports the researcher in performing an efficient, structured assessment of potential YouTube data in order to formulate a research question and identify a specific data set for analysis.

8.2 Reflections on Thesis Question 2

The second thesis question is: "What are the ethical issues in taking a sociolinguistic approach to researching speech in YouTube data and how could they be addressed?". My response to thesis question 2, if I was advising a researcher of a project design that minimises ethical complexities, is that:

    i)   the video data should have been produced by a YouTuber who can be described as A-List at the time of researching,

    ii)   the YouTuber should be informed your intention to research so they have the opportunity to object,

    iii)   identifying the YouTuber can be justified to assist in interpreting the findings but also to credit the YouTuber's production of the data,

iv)   a researcher may argue that disregarding Terms of Service is justifiable as these contradict both UK and USA Copyright and Fair Use law.

This thesis question was identified as a result of discovering that, although ethical guidance for conducting research using online data already exists, there is a paucity of guidance in regard to online audio and video, and no guidance for researching speech in particular. From reviewing the literature, it was evident that the main ethical considerations around using YouTube data are the tensions between i) anonymity and credit, ii) public data and informed consent, and iii) terms of service and data collection requirements. While this is also true for online data in general, these considerations are amplified as the research focus moves to speech. In comparison, the ethical considerations for researching YouTube comment data is relatively uncomplex and there are many sources of advice and considerable discussion about using online public written data. Therefore, the discussion herein centres on video/speech data.

During this case study, engaging with the literature guided me in regard to what ethical issues to consider and how to reflect upon them and their particular nuances in YouTube video data. I independently came to some conclusions in regard to i) anonymity and credit, and ii) public data and informed consent in this project as a result. However, after realising the significant impracticalities and risks associated with attempting to conduct sociolinguistic research on YouTube video data by streaming alone, I sought expert help from Hugh Rhodes, Enterprise Manager and Lawyer at Northumbria University. His knowledge and expertise assisted me in addressing the conundrum of iii) terms of service and data collection requirements.

8.2.1 Guidance for navigating ethical issues

By performing the case study detailed in this thesis I have gained some insight into navigating the ethical issues that may arise when conducting sociolinguistic research on YouTube data and I have collated these insights into an ethics

decision-making map (appendix 4). A map has been used, rather than a decision-making tree, because these ethical issues are interwoven and cannot be separated or arranged in a sequence. Also, I have written an email template which a researcher can use to inform a YouTuber that they intend to download their data and analyse their speech (appendix 5) the rationalisation for which is described in section 5.7.1 and so will not be repeated here. Therefore, this section focuses on describing appendix 4 and its intended use.

I spent a notable amount of time researching what the ethical issues could be when it comes to studying online data, and then translating these into the context of YouTube and analysing speech from YouTube. In producing the resource in appendix 4 my intention is to streamline the ethical-analysis and decision-making process for other researchers. Therefore, the ethics decision-making map is designed to guide researchers in navigating the ethical implications of their intended research, as well as help identify data and research designs that pose minimal ethical complexities, and encourage continual reflection upon the repercussions of making certain decisions upon the research subjects and the researcher. Of course, the ethical considerations that this resource encourages a researcher to reflect upon are not exhaustive and others are likely to arise as a result of the data that is being assessed and its social context. Further, this resource does not dictate a definitive criterion of what is or is not an ethical study of speech on YouTube. As previously stated, for research that incorporates online data there cannot be a 'one size fits all' approach and the ethical considerations for each research project needs to be assessed individually (Brake *et al.*, 2020).

The intention of the ethics decision-making map is for a researcher to identify the place where their intended project sits amongst the web of ethical considerations. For example, whether they will credit the data to its producer or anonymise the data. Equally, whether they intend to use data that can be described as 'public' or 'private'. It is very important that the terms 'private' and 'public' are not used in the technical sense but in the same vein as Nissenbaum's (2004) theory of 'contextual integrity' and Lange (2007) findings of publicly-private and privately-public as discussed in section 3.2.3. The map primarily

focuses on the tensions between i) anonymity and credit, and ii) public data and informed consent as it is apparent that the third issue (terms of service and data collection) was resolved by engaging with UK copyright law (see section 5.7.1). Pervading across both these continua is the decision as to whether informed consent from the data producer is required, or informing them of the intention to conduct research on their data and provide the opportunity to object is sufficient. Of course, the ability to use either approach is dependent on knowing who should be contacted to give permission or to object.

Identifying the place where their intended project sits will allow a researcher to consider whether adjustments to their intended method (e.g. chosen data, research question) will result in a more defensible project design in regard to ethical implications. The map is annotated with questions to prompt this reflection. Again, this set of questions is not exhaustive, but it is hoped these provide an initial prompt for researchers to then consider the particular nuances of their intended project. The positioning of the questions is intentional; they are designed to critique whether it is appropriate to use public-credited, public-anonymous and private-anonymous data. There are no questions in regard to credited-private data because it is difficult to imagine a scenario where using such data would be an ethically defensible decision, and thus the default decision should be to anonymise private data, hence why there are no questions here either.

In summary, this resource intends to synthesise the discussions of ethical practice in regard to online data in the literature that are relevant to researching speech. Actually, from surveying appendix 4 it is evident that this could be a useful resource for online data more generally. Therefore, its value to researching speech is most evident when approached from a sociolinguistic perspective and with an understanding of the particular technical and social nuances of online public video data, as are described throughout this thesis.

8.3 Reflections on Thesis Question 3

The third thesis question is: "What strategies could be used to guide the selection of linguistic variables in online data where place is ambiguous?" My response to thesis question 3 is that online ethnography can be used to conceptualise place and that global linguistic variants may be a fruitful subset of speech features.

This thesis question was identified as a result of realising that the backdrop against which linguistic variables are selected in both offline studies and the vast majority of online studies is place. Thus, selecting linguistic variables to study is complicated in YouTube data because place is ambiguous. In this thesis two strategies were tested.

The first strategy was for the variable to be a global speech feature. The rationale was that a YouTuber such as Zoella may imagine their audience as geographically broad and so may respond to their comments through a speech feature that unites multiple, geographically dispersed audiences. In other words, using the lowest common denominator speech feature (to adopt the term used by (Androutsopoulos, 2014, p. 66) and (Gil-Lopez *et al.*, 2018, p. 127)). Therefore, the selection of uptalk was primarily motivated by the physical *space* that the potential actual audience inhabits. While this strategy did not lead to a clear result on this occasion, there is still a strong rationale behind it and so it may transpire to be useful in future studies. Rather than the strategy being inappropriate it may be that speech feature uptalk was not the right one to select.

The second strategy was to employ online ethnography to gain insight into how Zoella and her commenters co-create place through the resources available, and ask whether this place changes over time. The rationale was that it may be that the definition of the imagined audience used to motivate selecting uptalk as the linguistic variable did not align with Zoella's definition. It has been found that social media users take cues from the social media environment to imagine their audience, and equally that the imagined audience becomes visible when it influences the information that users choose to broadcast, and so insight can be

gained into Zoella's definition of the imagined audience by examining the data contained in the YouTube interface. Therefore, the selection of word medial /t/ was primarily motivated by the conceptual *place* of the imagined audience. While in the end the defining of place played little role in the interpretation of Zoella's speech because the two primary imagined audiences agreed in regard to the indexical connotations of the two main variants, the online ethnography allowed me to gain insight into Zoella's development into an A-List YouTuber over time as well, without which the use of word medial /t/ could not be interpreted and the statistical examination of Zoella's status as amateur or professional would not have been prompted.

To summarise, in regard to guidance for other researchers in selecting and interpreting linguistic variables when studying speech in online data where place is ambiguous, online ethnography may provide clear direction and rationale for this decision making. In addition to confirming that ethnography can be used as a strategy for identifying speech variables, some knowledge of how ethnography can be adapted in order to respond to an online context, where spatiality and temporality are unhelpfully complicated, has also been given. Therefore, recommending that online ethnography is used in this way in future studies is the primary contribution in regard to thesis question three.

8.4 Reflections on Thesis Question 4

The final thesis research question is: 'What statistical approaches could be used in studies of speech in online public video considering that time can be operationalised with greater granularity?" My response to thesis question 4 is that there appears to be value in applying moderation analysis.

This thesis question was defined as a result of identifying YouTube's potential to provide data at many more points in time in comparison to current offline studies. The initial prediction was that this finer granularity would allow time to be operationalised as a continuous variable. However, as Zoella's YouTube data

was reviewed it became apparent that other variables also become available for analysis as a result of an expansion in the number of time datapoints. Namely, the comments posted on videos. Further, this prompted me to reconsider time's role beyond being one of many antecedent variables that may influence Zoella's speech. And so, considering novel kinds of data lead to novel kinds of research questions, and alternative statistical approaches to mixed effects regression analysis were explored.

The statistical approach used in this case study was moderation analysis, with the possibility of mediation analysis also considered but discarded. Specifically, the approach and tools recommended by Hayes (2018) were used, which I believe is novel for sociolinguistics research based on my review of the literature. As was stated in section 7.7 whether the focal antecedent and potential moderator are dichotomous, categorical or continuous, and how many potential moderators there are, all impose different requirements onto performing and interpreting a moderation analysis. Therefore, the description of moderation in this thesis is limited to what is useful to the case study, and so it would be inappropriate to provide guidance in regard to applying this method. Therefore, the contribution in this thesis does not go beyond pointing out the potential value of this statistical approach to sociolinguistic research in online public video where time is of a greater granularity or other variables are analysable as a result of an increase in the number of time data points by providing an initial illustrative example, and pointing to the resources that researchers can use to find out more.

8.5 Future work

Ultimately, this thesis has asked 'how can we conduct sociolinguistic research using online public video?' but cannot give an exhaustive answer, of course. While this case study evidences the possibilities and value of researching speech in online public video from a sociolinguistic perspective, many of its insights merely scratch the surface because the work herein is i) focused on a particular topic (audience design) and ii) the potential causal relationship between two

specific variables (comments and speech), iii) and is limited to a particular platform (YouTube). Therefore, future, long term work would be to i) ask research questions about different topics, ii) analysing different kinds of variables, and iii) conduct work on online public video sharing platforms other than YouTube. Through these activities, over time a research community specifically interested in speech in online public video could evolve, with research methods being continually refined until a core set of practices becomes established.

This grander vision builds upon the initial methodological guidance defined in this thesis and its appendices. Therefore, it would be beneficial to implement a strategy by which methodological insights that researchers gain, that are not necessarily reported in research publications, can be collated. In this vain, how others' use the guidelines and may give feedback, and the editing of the resources to reflect this generation of further knowledge, will be described in section 8.5.1.

Another area of future work is the making of tools and resources that would make the process of conducting research, particularly in regard to data preparation, more efficient. While such tools would be beneficial to all types of research this can be argued to be a pertinent issue when researching online data because of the pace in which the technical infrastructure and its social use changes. Some initial ideas for other tools and resources that are inspired by my experience as detailed in this thesis are suggested in section 8.5.2.

Finally, in this section the potential for the content of this thesis to influence research into speech contained in other types of YouTube video content and in Broadcast Media content will be discussed.

8.5.1 Monitoring use of the guidelines

A key outcome of this thesis is a set of resources to support sociolinguists in conducting research using online publicly shared video data. However, the

documents in the appendices are a first draft of these resources and only reflect the knowledge and experience of one researcher (myself) conducting one case study. Therefore, appendices 3 to 5 will be posted on a Project page on my ResearchGate profile (Sutton, 2020) in an attempt to encourage other researchers to use the documents and give feedback so that future versions may reflect the collective generation of experience and insight. ResearchGate is a website where academics can set up a profile and post information about their research, such as describing ongoing projects and posting draft publications, and network with other researchers. Thus, ResearchGate can provide visibility for and access to research prior to peer-reviewed publication, and so can be harnessed by researchers who are seeking feedback or collaboration. Also, ResearchGate tracks a number of analytics, such as document reads, which can be helpful in regard to monitoring the use of the documents prior to citation in peer-reviewed publication. The introduction on the project page will invite researchers to use the documents and give feedback on how helpful they were, and what they learned in regard to methods when using such data.

8.5.2 Other resources and tools

In performing the work described in this thesis I identified a number of tasks that were cumbersome or time consuming. As a result, I have several suggestions for other resources and tools that could make preparing YouTube data for conducting sociolinguistics research more efficient, primarily by taking advantage of its auto-generated, time stamped captions.

First, a script where the linguistic content of a video could be searched and summarised would assist researchers in establishing what videos to include and exclude from their study in regard to the potential linguistic variable of interest. As was mentioned in section 3.1.7, in contrast to online written data, one issue with YouTube data is the inability to search transcripts for linguistic content. However, a script could be written that performs this task by taking each word in the auto-generated captions, finding it in a pronunciation dictionary (e.g. The Carnegie Mellon Dictionary (Carnegie Mellon University, 2020)) and then

identifying if it contains the phoneme of interest by consulting the dictionary's transcription. The search term used would probably be a phoneme and the phonological context of interest could also be a search criterion too. Ideally, the script would then return the number of words found in the captions. As was mentioned in section 3.3, it is advised that a linguistic variable be frequent so there is plenty of data for analysis. So, having this script count the number of potential tokens, and what phonological contexts they are in, would be highly beneficial when designing a research study using YouTube data. Of course, the success of such a script would depend on the accuracy of YouTube's auto-generated captions.

Second, a script that takes YouTube's auto-generated time stamped captions and inserts them into a praat (Boersma and Weeink, 2018) textgrid would speed up data preparation. Of course, this would need to be carefully checked for not just transcription accuracy but also time stamping. However, this would take a lot less time than a researcher transcribing the data into praat themselves.

8.5.3 YouTube User Copied and Edited Content, and Broadcast Media

As stated in section 1.11, from the outset this thesis delimited its interests to 'user generated content' (a video the user has recorded with the intention of uploading it to YouTube) because this type of content evidences a unique interaction context where video creators communicate to their viewers via video and viewers communicate back via various means (e.g. views, likes, comments). But this thesis has implications for another type of YouTube video which I will refer to as 'user copied and edited content'. These types of videos, rather than being a mere copy of video that was recorded for other purposes and originally distributed outside of YouTube (e.g. television, movies, music videos, live streams), are edited and even spliced with other copied or user generated content, such as some music videos identified by Liikkanen and Salovaara (2015). While the speech produced in these videos was not in response to a YouTube audience, its selection and editing was, and thus, in theory, the content

creator may take into account viewer feedback in the comments when making these decisions. Further, these decisions may change over time. Therefore, the methodological insights documented in this thesis could have implications for such as study.

The learnings from this thesis could also be considered when conducting sociolinguistic research on speech in broadcast media. Engaging with social media while also watching broadcast media (a phenomenon called 'second-screening' (e.g. Feltwell et al 2017) has become the norm. Public figures receive feedback about their appearances on broadcast media via various social media platforms, such as Twitter. So, one could consider researching the impact of this social media feedback upon the speech of a speaker in broadcast media (such as a talk show host) over time. This sort of research would come with the additional challenge that the connection between the social media feedback and the speech is less tangible and direct compared to on YouTube, however. Broadcast media speech and social media feedback are found on different types of media streams (broadcast versus online), let alone different websites, whereas with YouTube the function to give feedback and the content being feedback on is within the same webpage. Because of this structure, there is less ambiguity in what speech is being commented on in YouTube compared to a tweet about a recent TV appearance. This, of course, has implications for the conclusions that can be made when researching the impact of social media feedback on speech in broadcast media over time.

8.6 Final Words

Beyond the many contributions to knowledge and practice detailed in this thesis, I would like to highlight one more. In performing the case study reported herein I am one of the first researchers to conduct a quantitative analysis of video data that was specifically created to be shared on the world's largest and longest running online, public video sharing platform; YouTube. In doing so, I challenge

this data's apparent reputation of being exotic, novel or un-pin-down-able, and am optimistic that other researchers may now see it as more approachable.

Of course, the conclusions in this thesis and its main contribution (guidelines for research practice) only scratch the surface of this new field of enquiry and will need to be revisited, extended, and fine-tuned. Therefore, I am highly aware that

> "[w]ith the pleasure of being the first goes the certainty of being wrong" (Labov, 1972, p. 98).

Appendix 1: List of videos

1. Sugg, Z. *Primark Haul*. Zoella.

https://www.youtube.com/watch?v=JUxDKZLOKYQ Published 25/02/2011

2. Sugg, Z. *Haul: Primark, H&M & Lush*. Zoella.

https://www.youtube.com/watch?v=dSLPyyYtvIs Published 13/03/2011

3. Sugg, Z. *Haul: Topshop, New Look, H&M & Superdrug*. Zoella.

https://www.youtube.com/watch?v=VUkj9waw864 Published 15/04/2011

4. Sugg, Z. *Mahusive Collective Haul*. Zoella.

https://www.youtube.com/watch?v=6uCY8EO5PDY Published 17/05/2011

5. Sugg, Z. *Primark Haul*. Zoella.

https://www.youtube.com/watch?v=iMwxRg9XZOI Published 06/08/2011

6. Sugg, Z. *Haul: Makeup & Car Booting*. Zoella.

https://www.youtube.com/watch?v=lnz3rx_MTPM Published 15/09/2011

7. Sugg, Z. *Primark Haul*. Zoella.

https://www.youtube.com/watch?v=PZx9DRQvj9M Published 13/10/2011

8. Sugg, Z. *Haul: New Look,Topshop, H&M & Bootsale*. Zoella.

https://www.youtube.com/watch?v=eEcp6jFZGZU Published 19/10/2011

9. Sugg, Z. *Primark Haul*. Zoella.

https://www.youtube.com/watch?v=wlJEvAytN1E Published 14/12/2011

10. Sugg, Z. *Collective Haul: Topshop, New Look, Soap & Glory, Style Compare, Orange Circle & Vintage*. Zoella.

https://www.youtube.com/watch?v=n-bb2LY6-GQ Published 02/02/2012

11. Sugg, Z. *Haul: Primark & New Look*. Zoella.

https://www.youtube.com/watch?v=GYgTyPwYGto Published 04/03/2012

12. Sugg, Z. *Primark Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=_kOPu-SXQDU Published 10/04/2012

13. Sugg, Z. *Little Haul: Mac, Revlon, Car Booty, Disney etc. | Zoella*. Zoella.

https://www.youtube.com/watch?v=Wyn0J3Rg80o Published 09/05/2012

14. Sugg, Z. *Little Haul : FeelUnique, Boots & Ebay | Zoella*. Zoella.

https://www.youtube.com/watch?v=Qv_DRdaF3IE Published 08/06/2012

15. Sugg, Z. *Collective Haul : Mac, Zara, Primark, American Apparel etc... | Zoella*. Zoella. https://www.youtube.com/watch?v=RCZNbABwBqM Published 21/07/2012

16. Sugg, Z. *Collective Haul : Topshop, Lush, H&M, FeelUnique & AA | Zoella.* Zoella. https://www.youtube.com/watch?v=7ZIPQqBzNuo Published 11/10/2012

17. Sugg, Z. *Makeup & Beauty Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=q4YmLkcFL_8 Published 25/11/2012

18. Sugg, Z. *Winter Primark Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=jPNcldD6Dmc Published 30/12/2012

19. Sugg, Z. *Big Drugstore Beauty Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=6jstRStk_cM Published 20/01/2013

20. Sugg, Z. *Topshop Haul & £500 Giveaway | Zoella*. Zoella. https://www.youtube.com/watch?v=IH8rWvom_oc Published 03/02/2013

21. Sugg, Z. *Huge Collective Haul & Giveaway | Zoella*. Zoella. https://www.youtube.com/watch?v=TCeyzwtIwtU Published 17/03/2013

22. Sugg, Z. *Huge Florida Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=TrfuLmHcD_A Published 31/03/2013

23. Sugg, Z. *Very Haul & Giveaway | Zoella*. Zoella. https://www.youtube.com/watch?v=hZcqemFa19w Published 09/06/2013

24. Sugg, Z. *Drugstore Makeup & Beauty Haul | Zoella.* Zoella. https://www.youtube.com/watch?v=Mej9EejFQyI Published 21/07/2013

25. Sugg, Z *Autumn & Winter Fashion Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=gMRnqfL5o5g Published 24/09/2013

26. Sugg, Z. *Huge Lush Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=lJDOtzCHXKo Published 26/10/2013

27. Sugg, Z. *Home "Stuff" Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=ciouSXGeg10 Published 02/02/2014

28. Sugg, Z. *HUGE Beauty & Cosmetics Haul | Zoella.* Zoella. https://www.youtube.com/watch?v=bBJ5vBvDEPE Published 09/03/2014

29. Sugg, Z. *Boohoo Haul & £500 Giveaway | Zoella*. Zoella. https://www.youtube.com/watch?v=L4uqN9BRTVQ Published 12/03/2014

30. Sugg, Z. *Huge Spring Clothing Haul | Zoella*. Zoella. https://www.youtube.com/watch?v=LkBolzwo9eI Published 18/05/2014

31. Sugg, Z. *Home Bits & Clothing Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=1BlZ5yDibV4 Published 24/08/2014

32. Sugg, Z. *Huge Lush Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=ClPu5aB5RPs Published 05/10/2014

33. Sugg, Z. *BooHoo Haul & Giveaway | Zoella*. Zoella.

https://www.youtube.com/watch?v=Un2HHllfzHI Published 09/11/2014

34. Sugg, Z. *Clothing, Homeware & Beauty Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=ALxSVK2j1i8 Published 18/01/2015

35. Sugg, Z. *America Haul | Sephora, Bath & Body Works & Sweets | Zoella*.

Zoella. https://www.youtube.com/watch?v=VoMqWgqIiCo Published

15/02/2015

36. Sugg, Z. *Huge Spring Primark Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=mCpx2M3GRH8 Published 11/03/2015

37. Sugg, Z. *Huge Summer Clothing Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=rZp8eFnJd5g Published 12/04/2015

38. Sugg, Z. *Beauty & Homeware Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=MoIijC6obd8 Published 17/05/2015

39. Sugg, Z. *Topshop & ASOS Clothing Haul | Zoella.* Zoella.

https://www.youtube.com/watch?v=zPdEzRb8Gzs Published 19/07/2015

40. Sugg, Z. *Lush Haul & First Impressions | Zoella*. Zoella.

https://www.youtube.com/watch?v=zaRIxFu-HFE Published 09/08/2015

41. Sugg, Z. *Stationery Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=Ecq7Yd1_vLU Published 23/08/2015

42. Sugg, Z. *Bath & Body Works Autumn Candle Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=rjOSKvkIYUE Published 13/09/2015

43. Sugg, Z. *Halloween & Christmas LUSH Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=YNhMbJCIkrc Published 25/10/2015

44. Sugg, Z. *Baking Haul | Zoella.* Zoella.

https://www.youtube.com/watch?v=oUyu59dNgsQ Published 23/11/2015

45. Sugg, Z. *Christmas Jumper Haul | Zoella.* Zoella.

https://www.youtube.com/watch?v=UEHQ96TF3S4 Published 02/12/2015

46. Sugg, Z. *Christmas Homeware, Clothing & Accessories Haul | Zoella*

Zoella. https://www.youtube.com/watch?v=CN1_RvKmj18 Published

03/12/2015

47. Sugg, Z. *Huge Boots Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=Ty0PScqt1w8 Published 15/02/2016

48. Sugg, Z. *Easter LUSH Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=-xyNw4UxLAI Published 21/03/2016

49. Sugg, Z. *Homeware Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=1PspMpI4B4g Published 24/04/2016

50. Sugg, Z. *Stationery Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=fDpUGd7fUOM Published 01/05/2016

51. Sugg, Z. *Huge Holiday ASOS Haul & Try On | Zoella*. Zoella.

https://www.youtube.com/watch?v=eyhMtyE1x1w Published 20/06/2016

52. Sugg, Z. *Autumn Bath & Body Works Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=nxs6DCR2SDM Published 31/08/2016

53. Sugg, Z. *Huge Disastrous Primark Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=tOafTz1RgXA Published 11/09/2016

54. Sugg, Z. *Autumn & Halloween Home Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=T-5fOe-CLxo Published 23/10/2016

55. Sugg, Z. *Christmas Jumper Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=chTVxX-uIMQ Published 02/12/2016

56. Sugg, Z. *Christmas Bath & Bodyworks Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=C6Mhi-UJBCs Published 06/12/2016

57. Sugg, Z. *Huge Winter ASOS Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=CLLnP4Vrf8A Published 07/12/2016

58. Sugg, Z. *Christmas Home Haul | Zoella*. Zoella.

https://www.youtube.com/watch?v=wiWDDaVbXfw Published 10/12/2016

Appendix 2: Inclusion / Exclusion criteria for uptalk

Transcript conventions:
All breaths: /
Pauses > 0.5 sec approx: .
Cuts/edits: []
Speech while speaker is not visible: ~~strikethrough~~
Speech directed to a specific audience: *italics.*

<u>Token identification process.</u>

(Order below does not reflect order of assessment).

Criteria: An independent clause that is an Intonation phrase (ending in Break Index 4).

Inspect the text prior to each breath and pause:

1. Does the breath/pause coincide with the end of a syntactic structure?
If no, reject.
If the breath/pause is followed by a cut/edit then will need to exercise discretion.

2. Is it a main/independent clause?
If no, reject.

3. Does the breath/pause coordinate with a Break Index 4 and a boundary tone at least in the initial instance?
If no, reject.

Criteria: Declarative

4. Is it a question? (Interrogative, yes-no, or tag question?)
If yes, reject.

5. Is it a command?
If yes, reject.

6. Is it a greeting or farewell?
If yes, reject.

7. Is it **inside** a list? In other words, not the final item.
If yes, reject.

Criteria: Fluent and of analysable quality

8. Does it contain any breaths or pauses?
If yes, reject.

9. Does it contain any hesitations, repetitions, or dysfluencies?
If yes, reject.

10. Is the majority of the clause produced with a neutral vocal setting and modal voicing?
If no, reject.

Criteria: Minimise other discourse factors

11. Does it appear to be a quote or impersonation of self or others? (e.g. 'I was like')
If yes, reject.

12. Does she appear to be reading?
If yes, reject.

13. Does it appear to be a declarative question? (e.g. frowning or squinting, stating uncertainty and addressing the audience "don't know whether you're gonna be able to see"). Reservation, implication or uncertainty, tentativeness.
If yes, reject.

Appendix 3: Guide for summarising online public video data

Online public video data: Essential information   1

Creator Name:		Channel name:	
# subscribers:		# videos:	
Start date:		Genre/topic:	
Video types:			
Other channels:			
Other platforms:			

Social and Geographical history

Gender		Age (approx):	
Language		Accent:	
Current location		Previous location:	
Key people:			
Key events:			
Other notes (e.g. ethnicity, sexuality):			

Celebrification: A-List indicators

Talent management:	
Paid advert/ PR products:	
Mainstream media:	
Commercial ventures:	

Online public video data: Video types   **2**

Creator Name:		Channel Name:	
Video type:		# on channel:	

Activities / content and structure:	
Items shown / used:	

Social and Geographical

Physical location:	
People and their roles:	
Notes on language:	

Media and Technical

Audio (e.g. quality, live or overlayed, equipment visible?)	
Camera (e.g. static or moving, angles, single or split screen):	
Degree and types of editing (e.g. visual manipulations, insertion of clips from other media):	

Celebrification: A-List indicators

Commerical activity (e.g. advert or PR related):	

Online public video data: Data qualities of video types 3

Creator Name:		Channel Name:	
Video type:		# on channel:	

	Vid1	Vid2	Vid3	Vid4	Vid5	Vid6	Vid7	Vid8
Factor 1								
Factor 2								
Factor 3								
Factor 4								
	Vid9	**Vid10**	**Vid112**	**Vid13**	**Vid14**	**Vid15**	**Vid16**	**Vid17**
Factor 1								
Factor 2								
Factor 3								
Factor 4								
	Vid18	**Vid19**	**Vid20**	**Vid21**	**Vid22**	**Vid23**	**Vid24**	**Vid25**
Factor 1								
Factor 2								
Factor 3								
Factor 4								

Appendix 4: Ethics Decision-Making Map

Public

Is it possible to anonymise?

Could the findings be detrimental to the subject?

What could be the repercussions of others finding this data via your work?

Will knowing who the subject is assist in understanding the findings?

Informing of intention to research

Can I identify who should be informed?

Informed Consent

Can I identify who should give consent?

Anonymous

Credited

Private

Appendix 5: Intention to research - email template

Email subject: The use of YouTube videos for Academic Research under Fair Use conditions

To whom it may concern,

My name is [Firstname Surname] and I am a [position] at [institution, Country]. I am contacting you in regard to a matter of copyright for academic research purposes. My research is investigating [e.g. speech phenomena (that is how people speak, why they may speak like this and why this may change based on different social factors) in video data that has been shared publicly online]. To this end, I will be examining some of your YouTube videos. Specifically, the videos I will analyse are:

1. [Video title on Youtube], [date of publication on YouTube], [Hyperlink]
2. [Video title on Youtube], [date of publication on Youtube], [Hyperlink]
3. [Video title on Youtube], [date of publication on Youtube], [Hyperlink]

[If appropriate: These videos also include X other speakers and so I am sending this correspondence to them/their representatives also as indicated in their YouTube "About" page.]

To conduct my research, some of my analysis will require the use of specialist software for which I need to download a copy of the video. The data downloaded will be used solely for my research and will not be distributed to anyone else via any means, and the original source of data will be clearly referenced by using web links.

I have been advised that my reasons for accessing and using this data is clearly within the Fair Use (USA) and Fair Dealing (UK) conditions within copyright law, and within YouTube's own Fair Use guidelines.

If this is not the case, please can you contact me by [DATE] and let me know of your concerns. If I do not hear from you, I will assume that you are comfortable with this use of this YouTube material. In addition, I have sent similar correspondence to YouTube's copyright team.

Finally, thank you for providing such entertaining and interesting videos.

Best wishes,

[NAME]

References

Abidin, C. (2015) 'Internet (in)famous: the mystification and folklore of microcelebrification', in *AoIR Selected Papers of Internet Research 16: The 16th Annual Meeting of the Association of Internet Researchers*. Chicago: Association of Internet Researchers.

Adami, E. (2009) '"We/YouTube": Exploring sign-making in video-interaction', *Visual Communication*, 8(4), pp. 379–399. doi: 10.1177/1470357209343357.

Agha, A. (2003) *The social life of cultural value*, *Language and Communication*. doi: 10.1016/S0271-5309(03)00012-0.

Agnew, J. A. (2002) *Place and politics in modern Italy*. Chicago, Illinois: The University of Chicago Press.

Alam, F. and Stuart-Smith, J. (2015) 'Identity, ethnicity and fine phonetic detail. An acoustic phonetic analysis of syllable-initial /t/ in Glaswegian girls of Pakistani heritage', in Hundt, M. and Sharma, D. (eds) *English in the Indian Diaspora*. Amsterdam, The Netherlands: John Benjamins Publishing Company, pp. 29–53.

Alderton, R. (2019) *Salience and social meaning in speech production and perception*. Lancaster University.

Alderton, R. (2020) 'Perceptions of T-glottalling among adolescents in South East England: A sign of "chavviness", or a key to "coolness"?', *English Today*, 36(3), pp. 40–47.

Alexander, J. (2020) 'YouTube officially rolls out changes to children's content following FTC settlement', *The Verge*, 6 January.

Allan, K. (1986) *Linguistic Meaning, Vol. 2*. London: Routledge & Kegan Paul.

Altendorf, U. (1999) 'Eshtuary English: Is English going Cokney?', *Moderna Språk*, XCIII(1.1-11). Available at: http://phon.ucl.ac.uk/home/estuary/home.htm.

Androutsopoulos, J. (2006a) 'Introduction : Sociolinguistics and computer-mediated communication', *Journal of Sociolinguistics*, 10(4), pp. 419–438. doi: 10.1111/j.1467-9841.2006.00286.x.

Androutsopoulos, J. (ed.) (2006b) *Special Issue: Computer-Mediated Communication*. Journal of Sociolinguistics.

Androutsopoulos, J. (2013) 'Online Data Collection', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 236–250.

Androutsopoulos, J. (2014) 'Languaging when contexts collapse: Audience design in social networking', *Discourse, Context and Media*. Elsevier, 4–5, pp. 62–73. doi: 10.1016/j.dcm.2014.08.006.

Androutsopoulos, J. (2015) 'Networked multilingualism: Some language practices on Facebook and their implications', *International Journal of Bilingualism*, 19(2), pp. 185–205. doi: 10.1177/1367006913489198.

Androutsopoulos, J. and Beißwenger, M. (2008) 'Introduction: Data and Methods in Computer-Mediated Discourse Analysis', *Language@Internet*, 5, pp. 1–7. doi: 10.1017/CBO9781107415324.004.

Androutsopoulos, J. and Ziegler, E. (2004) 'Exploring language variation on the Internet: Regional speech in a chat community', *Language Variation in Europe: Papers from the Second International Conference on Language Variation in Europe, ICLaVE*, 2(155), pp. 99–111. Available at: http://scholar.google.com/scholar?q=related:I37GngAAcGwJ:scholar.google.com/&hl=en&num=20&as_sdt=0,5%5Cnpapers3://publication/uuid/CD7AC788-3CB4-420A-90F0-369969FAD652.

Apple (no date) 'Siri'. Available at: https://www.apple.com/siri/ (Accessed: 13 September 2017).

Aslan, E. and Vásquez, C. (2018) '"Cash me ousside": A citizen sociolinguistic analysis of online metalinguistic commentary', *Journal of Sociolinguistics*, 22(4), pp. 406–431. doi: 10.1111/josl.12303.

Baayen, R. H. (2013) 'Multivariate statistics', in Podesva, R. J. and Sharma, D. (eds) *Research Methods in Linguistics*. Cambridge University Press.

Baird, S. (2001) 'How "to be like" a Kiwi: Verbs of quotation in New Zealand English', *New Zealand English Journal*, 15, p. 6. Available at: http://search.informit.com.au/documentSummary;dn=596472396940529;res=IE LNZC.

Baker, R. and Hazan, V. (2011) 'DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs', *Behavior Research Methods*, 43(3), pp. 761–770. doi: 10.3758/s13428-011-0075-y.

Bamman, D., Eisenstein, J. and Schnoebelen, T. (2014) 'Gender identity and lexical variation in social media.', *Journal of Sociolinguistics*, 18(2), pp. 135–160. doi: 10.1111/josl.12080.

Barbieri, F. (2007) 'Older men and younger women: A corpus-based study of quotative use in American English', *English World-Wide*, 28, pp. 23–45.

Barbieri, F. (2009) 'Quotative be like in American English: Ephemeral or here to stay?', *English World-Wide*, 30, pp. 68–90.

Baron, R. M. and Kenny, D. A. (1986) 'The Moderator-Mediator Variable Distinction in Social Psychological Research. Conceptual, Strategic, and Statistical Considerations', *Journal of Personality and Social Psychology*, 51(6), pp. 1173–1182. doi: 10.1037/0022-3514.51.6.1173.

Bathurst, B. (1996) 'A cute accent?', *The Observer Review*, 24 March.

Beachcomber (2012) '95 years old and still training dolphins', *Express*, 17 January. Available at: https://www.express.co.uk/comment/beachcombers/373352/95-years-old-and-still-training-dolphins.

Becker, K. (2013) 'The Sociolinguistic Interview', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 91–100.

Bell, A. (1984) 'Language Style as Audience Design', *Language in Society*, 13(2), pp. 145–204. doi: 10.1017/S004740450001037X.

Bell, A. (2001) 'Back in style: Reworking audience design', in Eckert, P. and Rickford, J. R. (eds) *Style and Sociolinguistic Variation*. Cambridge: Cambridge University Press, pp. 139–169.

Bell, A. and Johnson, G. (1997) 'Towards a sociolinguistics of style', *University of Pennsylvania Working Papers in Linguistics*, 4.1(1), pp. 2–21.

Benor, S. (2001) 'The learned /t/: Phonological variation in Orthodox Jewish English', *Penn Working Papers in Linguistics: Selected Papers from NWAV 2000.*, pp. 1–16.

Besnier, N. (2013) 'Vignette 3a. Responsibility to Research Participants in Representation', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. Oxon, UK: Routledge, pp. 46–49.

Biel, J.-I. and Gatica-Perez, D. (2011) 'VlogSense', *ACM Transactions on Multimedia Computing, Communications, and Applications*, 7S(1), pp. 1–21. doi: 10.1145/2037676.2037690.

Biel, J.-I., Tsiminaki, V., Dines, J. and Gatica-Perez, D. (2013) 'Hi YouTube! Personality impressions and verbal content in social video', *Proceedings of the 15th ACM on International conference on multimodal interaction - ICMI '13*, pp. 119–126. doi: 10.1145/2522848.2522877.

Biel, J. I. and Gatica-Perez, D. (2010) 'Voices of vlogging', *ICWSM 2010 - Proceedings of the 4th International AAAI Conference on Weblogs and Social Media*, pp. 211–214.

Biel, J. I. and Gatica-Perez, D. (2013) 'The YouTube Lens: Crowdsourced Personality Impressions and Audiovisual Analysis of Vlogs', *IEEE Transactions on Multimedia*. IEEE, 15(1), pp. 41–55. doi: 10.1109/TMM.2012.2225032.

Bishop, S. (2018) *Beauty Vlogging: Practices, Labours, Inequality*. University of East London.

Bishop, S. (2019) 'Vlogging Parlance: Strategic Talking in Beauty Vlogs', in Abidin, C. and Brown, M. L. (eds) *Microcelebrity Around the Globe: Approaches to Cultures of Internet Fame*. Bingley, UK: Emerald Publishing Limited, pp. 21–32.

Blaikie, N. (2007) *Approaches to Social Enquiry*. 2nd edn. Cambridge, UK: Polity Press.

Blyth, C., Recktenwald, S., Wang, J., Blyth, C. and Recktenwald, J. R. S. (1990)

'I'm like, " Say What ?!": A New Quotative in American Oral Narrative', *American Speech*, 65(3), pp. 215–227.

Boellstorff, T., Mardi, B., Pearce, C. and Taylor, T. L. (2012) *Ethnography and Virtual Worlds: A Handbook of Method*. Princeton, New Jersey: Princeton University Press.

Boersma, P. and Weeink, D. (2018) 'Praat: doing phonetics by computer'. Available at: http://www.praat.org/.

Bohmann, A. (2016) 'Language change because Twitter? Factors motivating innovative uses of because across the English-speaking Twittersphere', in Squires, L. (ed.) *English in Computer-Mediated Communication: Variation, Representation, and Change*. De Gruyter Mouton, pp. 149–178.

Bolander, B. and Locher, M. A. (2014) 'Doing sociolinguistic research on computer-mediated data: A review of four methodological issues', *Discourse, Context and Media*. Elsevier, 3(1), pp. 14–26. doi: 10.1016/j.dcm.2013.10.004.

Bolinger, D. (1978) 'Intonation across languages', in Greenberg, J. H. (ed.) *Universals of Human Language II: Phonology*. Stanford: Stanford University Press, pp. 471–524.

boyd, danah (2007) 'Why Youth (Heart) Social Network Sites: The Role of Networked Publics in Teenage Social Life', *MacArthur Foundation Series on Digital Learning – Youth, Identity, and Digital Media*, 7641(41), pp. 1–26. doi: 10.1162/dmal.9780262524834.119.

boyd, danah (2010) 'Social Network Sites as Networked Publics:Affordances, Dynamics, and Implications', in Papacharissi, Z. (ed.) *Networked Self: Identity, Community, and Culture on Social Network Sites*. London: Routledge, pp. 39–58.

boyd, danah (2014) *It's Complicated: The social lives of networked teens*. Yale: Yale University Press.

boyd, danah and Crawford, K. (2012) 'CRITICAL QUESTIONS FOR BIG DATA Provocations for a cultural, technological, and scholarly phenomenon', *Information, Communication & Society*, 15(5), pp. 662–679. doi: 10.1080/1369118X.2012.678878.

boyd, danah michele (2008) *Taken Out of Context*, *PhD Thesis*. doi: 10.1056/NEJMcps0801308.

Boyd, Z., Elliott, Z., Fruehwald, J., Hall-Lew, L. and Lawrence, D. (2015) 'An Evaluation of Sociolinguistic Elicitation Methods', *Proceedings of the 18th International Congress of Phonetic Sciences*, Paper #800.

Boyden, S. (2016) 'Zoella Has Launched Her Homeware, Lifestyle And Stationery Collection And It's All So Instagram Worthy', *MTV.co.uk*, 6 September. Available at: http://www.mtv.co.uk/zoella/news/zoella-has-launched-her-homeware-lifestyle-and-stationery-collection-and-its-all-so-instagram-worthy.

Bradford, B. (1996) 'Upspeak', *Speakout! The Newsletter of the IATEFL Pronunciation Special Interest Group*, 18, pp. 22–24.

Bradford, B. (1997) 'Upspeak in British English', *English Today*, 13(3), pp. 29–36.

Brake, D. J., Kathrine, A., Heise, N., Henriksen, A. H., Hongladarom, S., Jobin, A., Lim, S. S., Locatelli, E., Markham, A., Reilly, P. J. and Wilhelm, C. (2020) 'Internet Research : Ethical Guidelines 3.0 Association of Internet Researchers Unanimously approved by the AoIR membership October 6 , 2019'.

Brandschain, L., Graff, D., Cieri, C., Walker, K., Caruso, C. and Neely, A. (2010) *Greybeard - Voice and aging*, *Proceedings of LREC 2010, the Seventh International Conference on Language Resources and Evaluation*. Edited by N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, M. Rosner, and D. Tapias. Valletta, Malta. Available at: http://www.lrec-conf.org/proceedings/lrec2010/summaries/789.html (Accessed: 21 August 2020).

Brandt, A. and Jenks, C. (2013) 'Computer-Mediated Spoken Interaction: Aspects of Trouble in Multi-Party Chat Rooms', *Language@Internet*, 10(5).

Britain, D. (1992) 'Linguistic change in intonation: the use of high rising terminals in New Zealand English', *Language Variation and Change*, 4, pp. 77–104.

Britain, D. (2013) 'Space, Diffusion and Mobility', in Chambers, J. K. and

Schilling, N. (eds) *The Handbook of Language Variation and Change*. 2nd edn. Chichester, UK: Wiley-Blackwell, pp. 471–500.

Bruckman, A. (2002) 'Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet', *Ethics and Information Technology*, 4(3), pp. 217–231. doi: 10.1023/A:1021316409277.

Bruckman, A. (2014) 'Research ethics and HCI', in Kellogg, W. A. and Olson, J. S. (eds) *Ways of knowing in HCI*. New York, NY: Springer, pp. 449–468.

Bryant, P. (1980) *Australian questioning intonation: An addition to Speakers' response-seeking repertoire*. Australian National University.

Bucholtz, M. (2011) *White Kids: Language, Race, and Styles of Youth Identity*. Cambridge, UK: Cambridge University Press.

Buchstaller, I. (2008) 'The localization of global linguistic variants', *English World-Wide*, 29(1), pp. 15–44.

Buchstaller, I. and D'Arcy, A. (2009) 'Localized globalization: A multi-local, multivariate investigation of quotative be like', *Journal of Sociolinguistics*, 13(3), pp. 291–331. doi: 10.1111/j.1467-9841.2009.00412.x.

Bueno, S. (2019) *Youtubers BRITISH CREW*, *WeHeartIt.com*. Available at: https://weheartit.com/_SaraBueno/collections/8282917-youtubers-british-crew?page=3&before=61885844 (Accessed: 18 December 2019).

Burgess, J. and Green, J. (2009) *YouTube: Online Video and Participatory Culture*. Cambridge, UK: Polity Press.

Burgess, J. and Green, J. (2018) *YouTube: Online Video and Participatory Culture*. 2nd edn. Cambridge, UK: Polity Press.

Burland, K. (2017) 'Where the Black Country Meets "Black Barnsley": Dialect Variation and Identity in an Ex-Mining Community of Barnsley', in Montgomery, C. and Moore, E. (eds) *Language and a Sense of Place: Studies in Language and Region*. Cambridge, UK: Cambridge University Press, pp. 234–257.

Burr, T. (2020) 'Tanya Burr'. YouTube. Available at: https://www.youtube.com/tanyaburr.

Bussmann, H. (1998a) 'declarative sentence', in *Routledge Dictionary of Language and Linguistics*. London: Routledge, p. 227.

Bussmann, H. (1998b) 'main clause', in *Routledge Dictionary of Language and Linguistics*. London: Routledge, p. 716.

Bussmann, H. (1998c) 'phonetics', in *Routledge Dictionary of Language and Linguistics*. London: Routledge, p. 892.

Butler, S. (2019) 'Bras for £2, pyjamas for £5 … how low, low prices keep Primark tills ringing', *The Observer*. Available at: https://www.theguardian.com/business/2019/nov/09/primark-fast-fashion-retail-low-prices-sustainable.

Callier, P. (2016) 'Exploring stylistic co-variation on Twitter: The case of DH', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter, pp. 241–260.

Carnegie Mellon University (2020) 'The CMU Pronouncing Dictionary'. Available at: http://www.speech.cs.cmu.edu/cgi-bin/cmudict#about.

Caron, C., Raby, R., Mitchell, C., Théwissen-LeBlanc, S. and Prioletta, J. (2017) 'From concept to data: sleuthing social change-oriented youth voices on YouTube', *Journal of Youth Studies*. Taylor & Francis, 20(1), pp. 47–62. doi: 10.1080/13676261.2016.1184242.

Catford, J. C. (2010) *A Practical Introduction to Phonetics*. 2nd edn. Oxford, UK: Oxford University Press.

Cham, B. Q. (2016) 'Sixty Years of Speech: A Study of Language Change in Adulthood', *Lifespans and Styles*, 2(1), pp. 17–26. doi: 10.2218/ls.v2i1.2016.1427.

Chambers, J. K. and Trudgill, P. (1998) *Dialectology*. 2nd edn. Cambridge, UK: Cambridge University Press.

Chapman, J. (2020) 'Jim Chapman'. YouTube. Available at: https://www.youtube.com/jimchapman.

Chariatte, N. (2015) 'Internet-mediated Phonetics', *Entrehojas : Revista de Estudios Hispánicos*, 5(1)(Article 2).

Childs, B. (2016) 'Who I am and who I want to be: Variation and representation in a messaging platform', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter, pp. 261–280.

Choi, G. Y. and Behm-Morawitz, E. (2017) 'Giving a new makeover to STEAM: Establishing YouTube beauty gurus as digital literacy educators through messages and effects on viewers', *Computers in Human Behavior*. Elsevier Ltd, 73, pp. 80–91. doi: 10.1016/j.chb.2017.03.034.

Christiansen, M. S. (2015) '"A ondi queras": Ranchero identity construction by U.S. born Mexicans on Facebook', *Journal of Sociolinguistics*, 19(5), pp. 688–702. doi: 10.1111/josl.12155.

Cieri, C. (2011) 'Making a Field Recording', in Di Paolo, M. and Yaeger-Dror, M. (eds) *Sociophonetics: A student's guide*. New York: Routledge, pp. 24–35.

Cieri, C. and Yaeger-Dror, M. (2018) 'Alternative Sources of Panel Study Data: Opportunities, Caveats and Suggestions', in Evans Wagner, S. and Buchstaller, I. (eds) *Panel Studies of Variation and Change*. New York, NY: Taylor & Francis, pp. 53–72.

Clark, H. H. and Carlson, T. B. (1982) 'Hearers and speech acts', *Language*, 58(2), pp. 332–373. doi: 10.1353/lan.1982.0042.

Clarke, I. and Grieve, J. (2019) 'Stylistic variation on the Donald Trump Twitter account: A linguistic analysis of tweets posted between 2009 and 2018', *PLoS ONE*, 14(9), p. e0222062. doi: 10.1371/journal.pone.0222062.

Clopper, C. G. (2013) 'Experiments', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 151–161.

Coats, S. (2016) 'Grammatical feature frequencies of English on Twitter in Finland', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter, pp. 179–210.

Coats, S. (2019) 'Articulation Rate in American English in a Corpus of YouTube Videos', *Language and Speech*. doi: 10.1177/0023830919894720.

Cochrane, L. (2017) 'Topshop seeks smooth transition with London fashion week show', *The Guardian*, 17 September. Available at: https://www.theguardian.com/fashion/2017/sep/17/topshop-seeks-smooth-transition-with-london-fashion-week-show.

Collingridge, R. (2018) 'Alfie Deyes Leaves Gleam Futures', *TenEighty*, 22 January. Available at: https://teneightymagazine.com/2018/01/22/alfie-deyes-leaves-gleam-futures/.

Companies House (2013a) 'ZOE SUGG LIMITED (Company number 08399657)', 12 February. Available at: https://find-and-update.company-information.service.gov.uk/company/08399657.

Companies House (2013b) 'ZOELLA PRODUCTS LIMITED (Company number 08788394)', 25 November. Available at: https://beta.companieshouse.gov.uk/company/08788394.

Companies House (2014) 'CREW LIVE LIMITED (Company number 09060659)', 28 May. Available at: https://beta.companieshouse.gov.uk/company/09060659.

Companies House (2016a) 'PIPPIN PRODUCTIONS LTD (Company number 10270331)', 11 July. Available at: https://beta.companieshouse.gov.uk/company/10270331.

Companies House (2016b) 'ZS BEAUTY LTD (Company number 10276906)', 13 July. Available at: https://beta.companieshouse.gov.uk/company/10276906.

Companies House (2016c) 'ZS LIFESTYLE LTD (Company number 10276977)', 13 July. Available at: https://beta.companieshouse.gov.uk/company/10276977.

Coupland, N. (2001) 'Dialect stylization in radio talk', *Language in Society*, 30(3), pp. 345–375. doi: 10.1017/S0047404501003013.

Coupland, N. (2003) 'Sociolinguistic authenticities', *Journal of Sociolinguistics*, 7(3), pp. 417–431. doi: 10.1111/1467-9481.00233.

Coupland, N. (2007) *Style: Language Variation and Identity*. Cambridge University Press (Key Topics in Sociolinguistics).

Coupland, N. (2013) *The Handbook of Language and Globalization*. Edited by N. Coupland. Oxford, UK: Wiley-Blackwell.

Cukor-Avila, P. and Bailey, G. (2013) 'Real Time and Apparent Time', in Chambers, J. K. and Schilling, N. (eds) *The handbook of Language Variation and Change*. 2nd edn. Chichester, UK: Wiley-Blackwell, pp. 239–262.

Cutillas-Espinosa, J. A., Hernández-Campoy, J. M. and Schilling-Estes, N. (2010) 'Hypervernacularisation and speaker design: A case study', *Folia Linguistica*, 44(1), pp. 31–52. doi: 10.1515/flin.2010.002.

Cutler, C. (2016) '"Ets jast ma booooooooooooo": Social meanings of Scottish accents on YouTube', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter.

D'Arcy, A. and Young, T. M. (2012) 'Ethics and social media: Implications for sociolinguistics in the networked public', *Journal of Sociolinguistics*, 16(4), pp. 532–546. doi: 10.1111/j.1467-9841.2012.00543.x.

December, J. (1997) 'Notes on defining computer mediated communication', *CMC Magazine*.

De Decker, P. and Nycz, J. (2011) 'For the Record : Which Digital Media Can be Used for Sociophonetic Analysis?', *University of Pennsylvania Working Papers in Linguistics*, 17(2), pp. 51–59.

De Decker, P. and Nycz, J. (2013) 'The Technology of Conducting Sociolinguistic Interviews', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. Oxon, UK: Routledge, pp. 118–126.

Deyes, A. (2020) 'Alfie Deyes'. YouTube. Available at: https://www.youtube.com/alfiedeyes.

Ding, Y., Du, Y., Hu, Y., Liu, Z. and Wang, L. (2011) 'Broadcast yourself: understanding YouTube uploaders', *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 361–370. doi: 10.1145/2068816.2068850.

Docherty, G. J. and Foulkes, P. (1999) 'Derby and Newcastle: instrumental

phonetics and variationist studies', in Foulkes, P. and Docherty, G. J. (eds) *Urban Voices: Accent Studies in the British Isles*. London, UK: Routledge, pp. 47–71.

Dovchin, S. (2015) 'Language, multiple authenticities and social media: The online language practices of university students in Mongolia', *Journal of Sociolinguistics*, 19(4), pp. 437–459. doi: 10.1111/josl.12134.

Drager, K. (2014) 'Experimental Methods in Sociolinguistics', in Holmes, J. and Hazen, K. (eds) *Research Methods in Sociolinguistics: A Practical Guide*. Chichester: Wiley-Blackwell, pp. 58–73.

Drager, K. (2018) *Experimental Research Methods in Sociolinguistics*. London: Bloomsbury Academic.

Dryden, L. (2015) 'The 5 Most Important YouTuber Crews You Need To Know About', *PopBuzz*, 16 April. Available at: https://www.popbuzz.com/5-youtuber-friend-groups/.

Eckert, P. (1989) *Jocks and burnouts : social categories and identity in the high school*. Teachers College Press.

Eckert, P. (2003) 'Sociolinguistics and authenticity: An elephant in the room. Elephants in the room.', *Journal of Sociolinguistics*, 7(3), pp. 392–397.

Eckert, P. (2008) 'Variation and the indexical field', *Journal of Sociolinguistics*. John Wiley & Sons, Ltd (10.1111), 12(4), pp. 453–476. doi: 10.1111/j.1467-9841.2008.00374.x.

Eckert, P. (2012) 'Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation', *Annual Review of Anthropology*, 41(1), pp. 87–100. doi: 10.1146/annurev-anthro-092611-145828.

Eckert, P. (2016) 'Third Wave Variationism', *Oxford Handbooks Online*, February, pp. 1–18. doi: 10.1093/oxfordhb/9780199935345.013.27.

Ede, L. and Lunsford, A. (1984) 'Audience addressed/audience invoked: The role of audience in composition theory and pedagogy', *College Composition and Communication*, 35(2), pp. 155–171. doi: 10.2307/358093.

Eisenstein, J. (2015) 'Systematic patterning in phonologically-motivated

orthographic variation', *Journal of Sociolinguistics*, 19(2), pp. 161–188. doi: 10.1111/josl.12119.

Eisenstein, J., O'Connor, B., Smith, N. A. and Xing, E. P. (2014) 'Diffusion of lexical change in social media', *PLoS ONE*, 9(11), pp. 1–13. doi: 10.1371/journal.pone.0113114.

Van Engen, K. J., Baese-Berk, M., Baker, R. E., Choi, A., Kim, M. and Bradlow, A. R. (2010) 'The Wildcat Corpus of Native-and Foreign-accented English: Communicative Efficiency across Conversational Dyads with Varying Language Alignment Profiles', *Language and Speech*, 53(4), pp. 510–540. doi: doi.org/10.1177/0023830910372495.

Erik, S. (2014) 'Whose social meaning? Age and the indexical field: evidence from perception and conversational style in Manchester', in *ESRC End of Award Report*. Swindon, UK: ESRC.

Fabricius, A. (2002) 'Ongoing change in modern RP: Evidence for the disappearing stigma of t-glottalling', *English World-Wide*, 23(1), pp. 115–136. doi: 10.1075/eww.23.1.06fab.

Fabricius, A. H. (2000) *T-Glottalling Between Stigma and Prestige : a Sociolinguistic Study of Modern RP*. Copenhagen Business School.

Facebook (2019a) *Facebook*. Available at: www.facebook.com (Accessed: 22 October 2019).

Facebook (2019b) *Facebook Live*. Available at: https://www.facebook.com/facebookmedia/solutions/facebook-live (Accessed: 22 October 2019).

Feagin, C. (2013) 'Entering the Community: Fieldwork', in Chambers, J. K. and Schilling, N. (eds) *The Handbook of Language Variation and Change*. 2nd edn. Chichester, UK: Wiley-Blackwell, pp. 19–37.

Feltwell, T., Wood, G., Long, K., Brooker, P., Schofield, T., Petridis, I., Barnett, J., Vines, J. and Lawson, S. (2017) "I've been manipulated!": Designing Second Screen Experiences for Critical Viewing of Reality TV, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems,* pp. 2252-2263. Available at https://dl.acm.org/doi/10.1145/3025453.3025833

Fiesler, C., Beard, N. and Keegan, B. C. (2020) 'No Robots, Spiders, or Scrapers: Legal and Ethical Regulation of Data Collection Methods in Social Media Terms of Service', in *Proceedings of the Fourteenth International AAAI Conference on Web and Social Media*. Available at: www.aaai.org.

Fiesler, C. and Proferes, N. (2018) '"Participant" Perceptions of Twitter Research Ethics', *Social Media + Society*, Jan-Mar, pp. 1–14.

De Force, F. (2020) 'Fleur De Force'. YouTube. Available at: https://www.youtube.com/user/FleurDeForce.

Foulkes, P. and Docherty, G. (2006) 'The social life of phonetics and phonology', *Journal of Phonetics*, 34(4), pp. 409–438. doi: 10.1016/j.wocn.2005.08.002.

Foulkes, P., Docherty, G. and Jones, M. (2010) 'Analysing stops', in Di Paolo, M. and Yaeger-Dror, M. (eds) *Sociophonetics: A Student's Guide*. London, UK: Routledge, pp. 58–71.

Foulkes, P., Scobbie, J. M. and Watt, D. J. L. (2010) 'Sociophonetics', in Hardcastle, W. J., Laver, J., and Gibbon, F. E. (eds) *Handbook of Phonetic Sciences*. 2nd edn. Oxford: Blackwell Publishing, pp. 703–754.

Frazier, P. A., Tix, A. P. and Barron, K. E. (2004) 'Testing moderator and mediator effects in counseling psychology research', *Journal of Counseling Psychology*, 51(1), pp. 115–134. doi: 10.1037/0022-0167.51.1.115.

Frobenius, M. (2011) 'Beginning a monologue : The opening sequence of video blogs', *Journal of Pragmatics*, 43(3), pp. 814–827. doi: 10.1016/j.pragma.2010.09.018.

Frobenius, M. (2014) 'Audience design in monologues: How vloggers involve their viewers', *Journal of Pragmatics*, 72, pp. 59–72. doi: 10.1016/j.pragma.2014.02.008.

García-Rapp, F. (2016) 'The digital media phenomenon of YouTube beauty gurus: The case of Bubzbeauty', *International Journal of Web Based Communities*, 12(4), pp. 360–375. doi: 10.1504/IJWBC.2016.080810.

García-Rapp, F. (2017a) '"Come join and let's BOND": authenticity and legitimacy building on YouTube's beauty community', *Journal of Media*

Practice. Taylor & Francis, 18(2–3), pp. 120–137. doi: 10.1080/14682753.2017.1374693.

García-Rapp, F. (2017b) *The Digital Media Phenomenon of YouTube Beauty Gurus : The Case of Bubzbeauty TESI DOCTORAL UPF / 2017*. Universitat Pompeu Babra, Barcelona.

García-Rapp, F. (2019) 'Trivial and normative ? Online fieldwork within YouTube's beauty', *Journal of Contemporary Ethnography*, 48(5), pp. 619–644.

García-Rapp, F. and Roca-Cuberes, C. (2017) 'Being an online celebrity: Norms and expectations of YouTube's beauty community', *First Monday*, 22(7). Available at: https://doi.org/10.5210/fm.v22i7.7788.

Gatica-Perez, D., Sanchez-Cortes, D., Tri Do, T. M., Jayagopi, D. B. and Otsuka, K. (2018) 'Vlogging over time: Longitudinal impressions and behavior in YouTube', *ACM International Conference Proceeding Series*, pp. 37–47. doi: 10.1145/3282894.3282922.

Geere, M., Everett, J. and MacLeod, A. (2015) '"My Vocal Cords are Made of Tweed": Style-Shifting as Speaker Design', *Lifespans and Styles*, 1(0), p. 12. doi: 10.2218/ls.v1i0.2015.1179.

Georgakopoulou, A. (2015) 'Sharing as rescripting: Place manipulations on YouTube between narrative and social media affordances', *Discourse, Context and Media*. Elsevier, 9, pp. 64–72. doi: 10.1016/j.dcm.2015.07.002.

Georgakopoulou, A. and Spilioti, T. (2016) *Routledge Handbook of Language and Digital Communication*. London: Routledge.

Gil-Lopez, T., Shen, C., Benefield, G. A., Palomares, N. A., Kosinski, M. and Stillwell, D. (2018) 'One size fits all: Context collapse, self-presentation strategies and language styles on facebook', *Journal of Computer-Mediated Communication*, 23(3), pp. 127–145. doi: 10.1093/jcmc/zmy006.

Giles, H., Coupland, N. and Coupland, J. (1991) 'Accomodation theory: Communication, context, and consequence', *Contexts of accomodation: Developments in applied sociolinguistics*, pp. 1–68. doi: 10.1017/CBO9780511663673.001.

Giles, H., Taylor, D. M. and Bourhis, R. (1973) 'Towards a theory of

interpersonal accommodation through language: some Canadian data', *Language in Society*, 2(02), pp. 177–192. doi: 10.1017/S0047404500000701.

Di Gioacchino, M. and Crook Jessop, L. (2010) 'Uptalk: towards a quantitative analysis', *Toronto Working Papers in Linguistics*, 33, pp. 1–15.

Gleam Futures (2019) *Gleam futures*. Available at: https://www.gleamfutures.com/ (Accessed: 18 December 2019).

Godwin, R. (2015) 'Her cosy YouTube videos about shopping have spawned 20 million followers, a make-up range and a best-selling novel: Why every teenage girl wants to be Zoella', *Mail Online*, 25 October. Available at: http://www.dailymail.co.uk/home/you/article-3280866/I-don-t-feel-like-celebrity-just-feel-like-teenage-girl-wants-Zoella.html.

Goffman, E. (1981) *Forms of Talk*. Philadelphia: University of Pennsylvania Press.

Gold, E., Ross, S. and Earnshaw, K. (2018) 'The "West Yorkshire Regional English Database": Investigations into the Generalizability of Reference Populations for Forensic Speaker Comparison Casework', in *Proceedings of Interspeech 2018*. Hyderabad, pp. 2748–2752.

Gonçalves, B. and Sanchez, D. (2014) 'Crowdsourcing Dialect Characterization through Twitter', *PLoS ONE*, 9(11). doi: 10.1371/journal.pone.0112074.

Gordon, E., Campbell, L., Hay, J., Maclagan, M., Sudbury, A. and Trudgill, P. (2004) *New Zealand English: Its Origins and Evolution*. Cambridge: Cambridge University Press.

Gordon, M. J. (2013) *Labov: A Guide for the Perplexed*. Bloomsbury Academic (Guides for the Perplexed).

gov.uk (2019) 'Guidance: Exceptions to copyright', *gov.uk*, 5 August. Available at: https://www.gov.uk/guidance/exceptions-to-copyright.

Grabe, E. (2004) 'Intonational variation in urban dialects of English spoken in the British Isles', in Gilles, P. and Peters, J. (eds) *Regional Variation in Intonation*. Berlin, Germany: De Gruyter, pp. 9–31.

Grabe, E., Post, B. and Nolan, F. (2001) 'The IViE Corpus'. Department of

Linguistics, University of Cambridge.

Grieve, J., Montgomery, C., Nini, A., Murakami, A. and Guo, D. (2019) 'Mapping Lexical Dialect Variation in British English Using Twitter', *Frontiers in Artificial Intelligence*, 2(July), pp. 1–18. doi: 10.3389/frai.2019.00011.

Grieve, J., Nini, A. and Guo, D. (2018) 'Mapping Lexical Innovation on American Social Media', *Journal of English Linguistics*, 46(4), pp. 293–319. doi: 10.1177/0075424218793191.

Guy, G., Horvath, B., Vonwiller, J., Daisley, E. and Rogers, I. (1986) 'An intonational change in progress in Australian English', *Language in Society*, 15(1), pp. 23–51.

Guy, G. and Vonwiller, J. (1984) 'The meaning of an intonation in Australian English', *Australian Journal of Linguistics*, 4(1), pp. 1–17.

H&M (2020) *H&M*.

Hall-Lew, L. and Boyd, Z. (2017) 'Phonetic Variation and Self-Recorded Data', 23(2).

Hall-Lew, L., Coppock, E. and Starr, R. L. (2010) 'Indexing Political Persuasion: Variation in the Iraq Vowels', *American Speech*, 85(1), pp. 91–102. doi: 10.1215/00031283-2010-004.

Hall-Lew, L., Friskney, R. and Scobbie, J. M. (2017) 'Accommodation or political identity: Scottish members of the UK Parliament', *Language Variation and Change*, 29(3), pp. 341–363. doi: 10.1017/S0954394517000175.

Hammersley, M. and Atkinson, P. (2007) *Ethnography: Principles in practice*. 3rd edn. London, UK: Routledge.

Hansen, M., Fabriz, S. and Stehle, S. (2015) 'Cultural Cues in Students' Computer-Mediated Communication: Influences on E-mail Style, Perception of the Sender, and Willingness to Help', *Journal of Computer-Mediated Communication*, 20(3), pp. 278–294. doi: 10.1111/jcc4.12110.

Harley, D. and Fitzpatrick, G. (2009a) 'Creating a conversational context through video blogging: A case study of Geriatric1927', *Computers in Human Behavior*. Elsevier Ltd, 25(3), pp. 679–689. doi: 10.1016/j.chb.2008.08.011.

Harley, D. and Fitzpatrick, G. (2009b) 'YouTube and intergenerational communication: The case of Geriatric1927', *Universal Access in the Information Society*, 8(1), pp. 5–20. doi: 10.1007/s10209-008-0127-y.

Harrenstien, K. (2009) 'Automatic captions in YouTube', *Google Official Blog*, 19 November. Available at: https://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html (Accessed: 7 October 2020).

Harrington, J., Palethorpe, S. and Watson, C. (2000) 'Monophthongal vowel changes in Received Pronunciation: An acoustic analysis of the Queen's Christmas broadcasts', *Journal of the International Phonetic Association*, 30(1–2), pp. 63–78. doi: 10.1017/S0025100300006666.

Hayes, A. F. (2018) *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*. 2nd edn. London: Guilford Press.

Hayes, A. F. (2020) 'PROCESS'.

Hazen, K. (2014) 'An Historical Assessment of Research Questions in Sociolinguistics', in Holmes, J. and Hazen, K. (eds) *Research Methods in Sociolinguistics: A Practical Guide*. Chichester, UK: John Wiley & Sons, Ltd, pp. 7–22.

Van Herk, G. (2013) 'Working with and Preserving Existing Data', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. London, UK: Routledge, pp. 165–168.

Herring, Susan C (1996) *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Edited by S. C. Herring. Philadelphia: Benjamins, John.

Herring, Susan C. (1996) 'Introduction', in *Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives*. Amsterdam: John Benjamins Publishing Company, pp. 1–13.

Herring, S. C. (2001) 'Computer-mediated discourse', in Schiffrin, D., Tannen, D., and Hamilton, H. E. (eds) *The Handbook of Discourse Analysis*. Malden, Massachusetts: Blackwell, pp. 612–634.

Herring, S. C. (2004) *Computer-mediated discourse analysis: An approach to researching online behavior*, *Designing for Virtual Communities in the Service of Learning*. doi: 10.1017/CBO9780511805080.016.

Herring, S. C. (2007) 'A Faceted Classification Scheme for Computer-Mediated Discourse', *Language@Internet*, 4.

Herring, S. C., Stein, D. and Virtanen, T. (2013) *Pragmatics of computer-mediated communication*. Edited by S. C. Herring, D. Stein, and T. Virtanen. Berlin, Germany: Walter de Gruyter.

Highfield, T. and Leaver, T. (2016) 'Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji', *Communication Research and Practice*. Routledge, 2(1), pp. 47–62. doi: 10.1080/22041451.2016.1155332.

Hine, C. (2000) *Virtual Ethnography*. London, UK: SAGE Publications.

Hine, C. (2005) 'Virtual Methods and the Sociology of Cyber-Social-Scientific Knowledge', in Hine, C. (ed.) *Virtual Methods: Issues in Social Research on the Internet*. Oxford: Berg, pp. 1–16.

Hine, C. (2015) *Ethnography for the internet: Embedded, embodied and everyday*. London, UK: Bloomsbury Academic.

Hinsliff, G. (2019) 'Cheap and cheerful: why there's more to Primark's success than you thought', *The Guardian*, 28 May.

Horvath, B. M. (2013) 'Ways of observing: Studying the interplay of social and linguistic variation', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 7–16.

Huang, Y., Guo, D., Kasakoff, A. and Grieve, J. (2016) 'Understanding U.S. regional linguistic variation with Twitter data analysis', *Computers, Environment and Urban Systems*. Elsevier Ltd, 59, pp. 244–255. doi: 10.1016/j.compenvurbsys.2015.12.003.

Huffaker, D. A. and Calvert, S. L. (2005) 'Gender, Identity, and Language Use in Teenage Blogs.', *Journal of Computer-Mediated Communication*, 10(2), pp. 0–0. doi: https://doi.org/10.1111/j.1083-6101.2005.tb00238.x.

Hung, A. C. Y. (2017) 'Hanging Out on Xbox Live: How Teens Enter and Open Conversations in Party Chats', *Language@Internet*, 14(3).

Hunter Johnston, L. (2014) 'Zoella: Yes, using a ghostwriter matters when your whole brand is built on being authentic', *Independent*, 8 December.

IBM Corp (2019) 'SPSS'.

Ilbury, C. (2019) '"Sassy Queens": Stylistic orthographic variation in Twitter and the enregisterment of <scp>AAVE</scp>', *Journal of Sociolinguistics*, p. josl.12366. doi: 10.1111/josl.12366.

Instagram (2019) *Instagram*. Available at: https://www.instagram.com/ (Accessed: 22 October 2019).

International Phonetic Association (1999) *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge, UK: Cambridge University Press.

Ivković, D. (2013) 'The Eurovision Song Contest on YouTube: A Corpus-based Analysis of Language Attitudes', *Language@Internet*, 10(10), pp. 1–25. Available at: http://www.languageatinternet.org/articles/2013/Ivkovic.

Jaccard, J. (2013) 'Theory Construction, Model Building, and Model Selection', in Little, T. D. (ed.) *The Oxford Handbook of Quantitative Methods in Psychology, Vol. 1: Foundations*. Oxford: Oxford University Press, pp. 82–104.

Jeffries, L. (2011) 'The revolution will be soooo cute: YouTube "hauls" and the voice of young female consumers', *Studies in Popular Culture*, 33(2), pp. 59–75. doi: 10.2307/23416384.

Jenks, C. and Firth, A. (2008) 'Synchronous voice-based computer-mediated communication', in Herring, S. C., Stein, D., and Virtanen, T. (eds) *Pragmatics of Computer-Mediated Communication*. De Gruyter Mouton, pp. 217–244.

Jerslev, A. (2016) 'In the time of the microcelebrity: Celebrification and the YouTuber Zoella', *International Journal of Communication*, 10, pp. 5233–5251.

Johnson, K. (2012) *Acoustic and Auditory Phonetics*. 3rd edn. Chichester: Wiley-Blackwell.

Jones, G. M. and Schieffelin, B. B. (2009) 'Talking text and talking back: "my

BFF jill" from boob tube to youtube', *Journal of Computer-Mediated Communication*, 14(4), pp. 1050–1079. doi: 10.1111/j.1083-6101.2009.01481.x.

Jones, T. (2015) 'Toward a description of African American vernacular english dialect regions using "black twitter"', *American Speech*, 90(4), pp. 403–440. doi: 10.1215/00031283-3442117.

Jones, T. (2016) 'Tweets as graffiti: What the reconstruction of Vulgar Latin can tell us about Black Twitter', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter, pp. 43–68.

Jun, S.-A. and Fletcher, J. (2014) 'Methodology of studying intonation: from data collection to data analysis', in Jun, S.-A. (ed.) *Prosodic Typology II: The Phonology of Intonation and Phrasing*. Oxford, UK: Oxford University Press, pp. 493–519.

Kaver, L. (2017) 'One billion captioned videos', *YouTube Official Blog*, 16 February. Available at: https://blog.youtube/inside-youtube/one-billion-captioned-videos (Accessed: 7 October 2019).

Kelly, E. (2017) 'YouTube fans reckon Alfie Deyes has left Team Gleam', *Metro.co.uk*, 27 November. Available at: https://metro.co.uk/2017/11/27/youtube-fans-reckon-alfie-deyes-has-left-team-gleam-7111933/.

Kelly, M. (2019) 'Google will pay $170 million for YouTube's child privacy violations', *The Verge*, 4 September. Available at: https://www.theverge.com/2019/9/4/20848949/google-ftc-youtube-child-privacy-violations-fine-170-milliion-coppa-ads (Accessed: 24 August 2020).

Kiesling, S. F. (2005) 'Variation, stance and style: word-final -er, high rising tone, and ethnicity in Australian English', *English World-Wide*, 26(1), pp. 1–42.

Kim, C., Reddy, S., Stanford, J. N., Wyschogrod, E. and Grieve, J. (2019) 'Bring on the crowd! Using online audio crowd-sourcing for large-scale new England dialectology and acoustic sociophonetics', *American Speech*, 94(2), pp. 151–194. doi: 10.1215/00031283-7251252.

Kirkham, S. (2013) *Ethnicity, social practice and phonetic variation in a*

Sheffield secondary school. University of Sheffield.

Kirkham, S. and Moore, E. (2016) 'Constructing social meaning in political discourse: Phonetic variation and verb processes in Ed Miliband's speeches', *Language in Society*, 45(1), pp. 87–111. doi: doi:10.1017/S0047404515000755.

Knight, R.-A. (2012) *Phonetics: A coursebook*. Cambridge: Cambridge University Press.

Kozinets, R. (2010) *Netnography*. London, UK: SAGE Publications.

Krippendorff, K. (2004) *Content Analysis: An introduction to its methodology*. 2nd edn. London: Sage Publishing.

Labov, W. (1963) 'The Social Motivation of a Sound Change', *Word*, 19, pp. 273–309.

Labov, W. (1966) *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.

Labov, W. (1972) 'Some principles of linguistic methodology', *Language in Society*, 1(1), pp. 97–120. doi: 10.1017/S0047404500006576.

Labov, W. (2006) *The Social Stratification of English in New York City*. 2nd edn. Cambridge University Press.

Labov, W. (2013) *The Language of Life and Death: The Transformation of Experience in Oral Narrative*. Cambridge: Cambridge University Press. doi: 10.1017/cbo9781139519632.002.

Ladd, D. R. (2008) *Intonational Phonology*. 2nd edn. Cambridge: Cambridge University Press.

Ladefoged, P. (2005) *A Course in Phonetics*. 5th edn. Thomson/ Wadsworth Publishers.

LaFave, N. (2016) 'Social factors and lexical frequency influencing English adjective gradation in speech and CMC', in Squires, L. (ed.) *English in computer-mediated communication: Variation, Representation, and Change*. Berlin, Germany: De Gruyter, pp. 301–326.

Landert, D. and Jucker, A. H. (2011) 'Private and public in mass media communication: From letters to the editor to online commentaries', *Journal of*

Pragmatics. Elsevier B.V., 43(5), pp. 1422–1434. doi: 10.1016/j.pragma.2010.10.016.

Lange, P. G. (2007) 'Publicly private and privately public: Social networking on YouTube', *Journal of Computer-Mediated Communication*, 13(1), pp. 361–380. doi: 10.1111/j.1083-6101.2007.00400.x.

Lave, J. and Wenger, E. (1991) *Situated Learning: Legitimate Peripheral Participation*. Cambridge, UK: Cambridge University Press.

Lawson, R. (2011) 'Patterns of linguistic variation among Glaswegian adolescent males', *Journal of Sociolinguistics*, 15(2), pp. 226–255. doi: 10.1111/j.1467-9841.2011.00477.x.

Lee, S. (2017) 'Style-Shifting in Vlogging : An Acoustic Analysis of "YouTube Voice"', *Lifespans and Styles*, 3(1)(Article 4). doi: 10.2218/ls.v3i1.2017.1826.

Leemann, A., Kolly, M. J. and Britain, D. (2018) 'The English Dialects App: The creation of a crowdsourced dialect corpus', *Ampersand*. Elsevier Ltd, 5, pp. 1–17. doi: 10.1016/j.amper.2017.11.001.

Levon, E. (2013a) 'Ethnographic Fieldwork', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 69–79.

Levon, E. (2013b) 'Ethnography and recording interaction', in Podesva, R. J. and Sharma, D. (eds) *Research Methods in Linguistics*. Cambridge, UK: Cambridge University Press, pp. 195–215.

Levon, E. and Holmes-Elliott, S. (2013) 'East end boys and west end girls: /s/-fronting in Southeast England', *University of Pennsylvania Working Papers in Linguistics*, 19(2), pp. 111–120. Available at: http://repository.upenn.edu/cgi/viewcontent.cgi?article=1309&context=pwpl.

Lewis, K., Kaufman, J., Gonzalez, M., Wimmer, A. and Christakis, N. (2008) 'Tastes, ties, and time: A new social network dataset using Facebook.com', *Social Networks*, 30(4), pp. 330–342. doi: 10.1016/j.socnet.2008.07.002.

Liikkanen, L. A. and Salovaara, A. (2015) 'Music on YouTube: User engagement with traditional, user-appropriated and derivative videos', *Computers in Human Behavior*. Elsevier Ltd, 50, pp. 108–124. doi:

10.1016/j.chb.2015.01.067.

Lim, H. L. and Sudweeks, F. (2014) *Innovative Methods and Technologies for Electronic Discourse Analysis*. Hershey, Pennsylvania, USA: Information Science Reference.

Lindsey, G. (2019) *English After RP: Standard British Pronunciation Today*. Cham, Switzerland: Palgrave Macmillan.

Linguistic Society of America (2009) *Linguistic Society of America Ethics Statement*. Available at: https://www.linguisticsociety.org/resource/ethics.

Lister, M., Dovey, J., Giddings, S., Grant, I. and Kelly, K. (2009) *New media: a critical introduction*. 2nd edn. Oxon, UK: Routledge. doi: 10.1177/026858090702200234.

Livingstone, S. (2005) 'On the relation between audiences and publics', in Livingstone, S. (ed.) *Audiences and Publics: When Cultural Engagement Matters for the Public Sphere*. Bristol, UK: Intellect, pp. 17–42.

Llamas, C. (2007) '"A place between places": Language and identities in a border town', *Language in Society*, 36(4), pp. 579–604. doi: 10.1017/S0047404507070455.

Love, J. and Walker, A. (2013) 'Football versus football: Effect of topic on /r/ realization in American and English sports fans', *Language and Speech*, 56(4), pp. 443–460. doi: 10.1177/0023830912453132.

Love Productions (2015) 'The Great Comic Relief Bake Off'. BBC. Available at: https://www.bbc.co.uk/programmes/b053h69s.

Macaulay, R. K. S. (1977) *Language, Social Class and Education: A Glasgow Study*. Edinburgh: University of Edinburgh Press.

Macaulay, R. K. S. (2009) *Quantitative Methods in Sociolinguistics*. Basingstoke, UK: Palgrave Macmillan.

Mallinson, C., Childs, B. and Van Herk, G. (2013) *Data Collection in Sociolinguistics: Methods and Applications*. Oxon, UK: Routledge.

Mann, S. L. (2013) 'Vignette 3b. Conducting Research with Vulneraboe Populations', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data*

Collection in Sociolinguistics: Methods and Applications. Oxon, UK: Routledge, pp. 50–53.

Manovich, L. (2001) *The Language of New Media*. Cambridge, Massachusetts: MIT Press.

Marcus, G. E. (1995) 'Ethnography in/of the World System: The Emergence of Multi-Sited Ethnography', *Annual Review of Anthropology*, 24, pp. 95–117.

Markham, A. and Buchanan, E. (2012) 'Ethical Decision-Making and Internet Research : Recommendations from the AoIR Ethics Working Committee (Version 2.0)'.

Markham, A. N. (2006) 'Ethic as Method, Method as Ethic', *Journal of Information Ethics*, 15(2), pp. 37–55.

Marsh, S. (2006) 'The rise of the interrogatory statement', *The Times*, 28 March. Available at: http://www.thetimes.co.uk/tto/law/columnists/article2045829.ece.

Marwick, A. E. (2013) *Status update: Celebrity, publicity and branding in the social media age.* New Haven, Connecticut: Yale University Press.

Marwick, A. E. and boyd, danah (2010) 'I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience', *New Media & Society*, 13(1), pp. 114–133. doi: 10.1177/1461444810365313.

Mather, P. A. (2012) 'The Social Stratification of /r/ in New York City: Labov's Department Store Study Revisited', *Journal of English Linguistics*, 40(4), pp. 338–356. doi: 10.1177/0075424211431265.

Mathisen, A. G. (1999) 'Sandwell, West Midlands: ambiguous perspectives on gender patterns and models of change', in Foulkes, P. and Docherty, G. (eds) *Urban Voices: Accent Studies in the British Isles*. Oxon, UK: Taylor & Francis, pp. 107–123.

McDonald, K. (2018) *Cu[?]ing the crease: Style-shifting between vlogs and tutorials – A study of glottal replacement of /t/ in the female YouTube voice*. University of Glasgow.

McGoogan, C. (2017) 'How YouTube's golden girl Zoella lost her shine', *The Telegraph*, 15 November.

McGregor, J. (2005) *High rising tunes in Australian English*. Macquarie University.

McGregor, R. L. (1979) *High-rising tone in non-question forms in Sydney Australian English*. Macquarie University.

Mees, I. and Collins, B. (1999) 'Cardiff: A real-time study of glottalisation', in Foulkes, P. and Docherty, G. (eds) *Urban Voices: Accent Studies in the British Isles*. London: Routledge, pp. 185–202.

Mendoza-Denton, N. (2008) *Homegirls: Language and culture practice among latina youth gangs*. Massachusetts: Blackwell.

Mendoza-Denton, N. (2016) 'Norteño and Sureño Gangs, Hip Hop, and Ethnicity on YouTube', in Alim, H. S., Rickford, J., and Ball, A. (eds) *Raciolinguistics: How Language Shapes Our Ideas About Race*. Oxford, UK: Oxford University Press, pp. 135–150.

Meyerhoff, M. (1992) 'High rising terminal, eh?', *New York Times*, 19 January.

Milroy, J. and Milroy, L. (1978) 'Belfast: change and variation in an urban vernacular', in Trudgill, P. (ed.) *Sociolinguistic patterns in British English*. London: E. Arnold, pp. 19–36.

Milroy, J. and Milroy, L. (1985) 'Linguistic change, social network and speaker', *Journal of Linguistics*, 21(2), pp. 339–384.

Milroy, J., Milroy, L., Hartley, S. and Walshaw, D. (1994) 'Glottal stops and Tyneside glottalization: Competing patterns of variation and change in British English', *Language Variation and Change*, 6, pp. 327–357.

Milroy, L. (2007) 'Off the shelf or under the counter? On the social dynamics of sound changes', in Cain, C. M. and Russom, G. (eds) *Managing Chaos: Strategies for Identifying Change in English*. New York, New York, USA: Mouton de Gruyter, pp. 149–172.

Milroy, L. and Gordon, M. (2003) *Sociolinguistics: Method and Interpretation*. Oxford, UK: Blackwell Publishing.

Milroy, L. and Milroy, J. (1992) 'Social network and social class: Toward an integrated sociolinguistic model', *Language in Society*, 21(1), pp. 1–26. doi:

10.1017/S0047404500015013.

Mocanu, D., Baronchelli, A., Perra, N., Gonçalves, B., Zhang, Q. and Vespignani, A. (2013) 'The Twitter of Babel: Mapping World Languages through Microblogging Platforms', *PLoS ONE*, 8(4). doi: 10.1371/journal.pone.0061981.

Montgomery, C. (2017) 'Maps and Mapping in (Perceptual) Dialect Geography', in Montgomery, C. and Moore, E. (eds) *Language and a Sense of Place*, pp. 147–170.

Montgomery, C. and Moore, E. (2017) *Language and a Sense of Place: Studies in Language and Region*. Cambridge, UK: Cambridge University Press.

Moore, J. (2019) 'Primark won't easily surrender its position as the face of fast fashion with market share growing', *The Independent*, 17 January. Available at: https://www.independent.co.uk/news/business/comment/primark-sales-fast-fashion-high-street-retail-parliamentary-environmental-audit-committee-christmas-a8732256.html.

Neuendorf, K. A. (2017) *The Content Analysis Guidebook*. 2nd edn. London: Sage Publishing.

Nguyen, D., Dogruöz, A. S., Rosé, C. P. and de Jong, F. (2016) 'Computational sociolinguistics: A survey.', *Computational Linguistics*, 42, pp. 537–593.

Nissenbaum, H. (2004) 'Privacy is one of the most enduring social issues associated with Washington Law Review National Network to End Domestic Violence and the American Civil', *Washington Law Review*, 79(1), pp. 101–139. doi: 10.1109/SP.2006.32.

Nolan, F. (2006) 'Intonation', in Aarts, B. and McMahon, A. (eds) *Handbook of English Linguistics*. Oxford: Blackwell, pp. 433–458.

O'Reilly, K. (2009) *Key Concepts in Ethnography*. London, UK: SAGE Publications.

Oakley, T. (2014) 'Universal Day with the British Crew | Playlist 2014 Day 2'. YouTube. Available at: https://www.youtube.com/watch?v=YmsDkZaavjM.

Ogden, R. (2017) *An Introduction to English Phonetics*. 2nd edn. Edinburgh

University Press.

Ohala, J. (1983) 'Cross-language use of pitch: an ethological view', *Phonetica*, 40, pp. 1–18.

Onita, L. (2019) 'Can fast-fashion pioneer Primark survive the growing demand for ethical clothing?', *The Telegraph*, 7 December.

Oppenheim, M. (2016) *YouTube star Zoella is reportedly earning £50,000 a month*, *The Independent*. Available at: http://uk.businessinsider.com/zoella-reportedly-earns-50000-a-month-2016-3 (Accessed: 20 July 2018).

Ostendorf, M., Price, P. and Shattuck-Hufnagel, S. (1996) *Boston University Radio Speech Corpus*. Available at: https://catalog.ldc.upenn.edu/LDC96S36 (Accessed: 21 August 2020).

Ouirdi, M. El, El Ouirdi, A., Segers, J. and Henderickx, E. (2014) 'Social Media Conceptualization and Taxonomy: A Lasswellian Framework', *Journal of Creative Communications*, 9(2), pp. 107–126. doi: 10.1177/0973258614528608.

Panzarino, M. (2013) 'Google Dumps Video Responses From YouTube Due To Dismal .0004% Click-Through Rate', *techcrunch*, 28 August. Available at: https://techcrunch.com/2013/08/27/google-dumps-video-responses-from-youtube-due-to-dismal-0004-click-through-rate/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAKSnlXQ1iS_uDcqCPzY6QRjmzWH8i2wCuh5fc1YqZXRBDyUI9c75YEFlQfE.

Paolillo, J. C. (2001) 'Language variation on Internet Relay Chat: A social network approach', *Journal of Sociolinguistics*, 5(2), pp. 180–213. doi: 10.1111/1467-9481.00147.

Park, S., Shim, H. S., Chatterjee, M., Sagae, K. and Morency, L.-P. (2014) 'Computational Analysis of Persuasiveness in Social Multimedia', in *International Conference on Multimodal Interaction*. ACM Press, pp. 50–57. doi: 10.1145/2663204.2663260.

Parkin, J. (2006) 'Stomach-churning gossip about the school loos. Slang that's, like, totally infuriating. Oh, the joys (and, yes, benefits) of family mealtimes', *Daily Mail*, 29 June, p. 49.

Parry, K. W. (2011) 'Constant Comparison', in Lewis-Beck, M. S., Bryman, A., and Liao, T. F. (eds) *The SAGE Encyclopedia of Social Science Research Methods*. Thousand Oaks,: SAGE Publications, p. 181. doi: 10.4135/9781412983907.n376.

Patterson, A. N. (2018) 'YouTube Generated Video Clips as Qualitative Research Data: One Researcher's Reflections on the Process', *Qualitative Inquiry*, 24(10), pp. 759–767. doi: 10.1177/1077800418788107.

PaulApproves (2011) *Under 301 club*, *Urban Dictionary*. Available at: https://www.urbandictionary.com/define.php?term=Under 301 club (Accessed: 7 January 2020).

Pavalanathan, U. and Eisenstein, J. (2015) 'Audience-modulated variation in online social media', *American Speech*, 90(2), pp. 187–213. doi: 10.1215/00031283-3130324.

Pentland, L. (2020) 'Louise Pentland'. YouTube. Available at: https://www.youtube.com/user/Sprinkleofglitter.

PewDiePie (2018) *CURB YOUR MEME COMPILATIONS*, *YouTube*. Available at: https://www.youtube.com/watch?v=mq_mgCb_Xtw.

Pfrehm, J. (2018) *Technolingualism. The Mind and the Machine*. London: Bloomsbury Academic.

Pierrehumbert, J. (1980) *The Phonetics and phonology of English Intonation*. MIT.

Pihlaja, S. (2011) 'Cops, popes, and garbage collectors: metaphor and antagonism in an atheist/Christian YouTube video thread', *Language@Internet*, 8, pp. 1–17.

Podesva, Robert J. (2007) 'Phonation type as a stylistic variable: The use of falsetto in constructing a persona', *Journal of Sociolinguistics*. John Wiley & Sons, Ltd (10.1111), 11(4), pp. 478–504. doi: 10.1111/j.1467-9841.2007.00334.x.

Podesva, Robert J (2007) 'Three sources of stylistic meaning', *Texas Linguistic Forum (Proceedings of the Symposium About Language and Society – Austin 15)*, 15, pp. 134–143. Available at:

http://studentorgs.utexas.edu/salsa/proceedings/2007/Podesva.pdf.

Podesva, R. J. (2011) 'The California Vowel Shift and Gay Identity', *American Speech*, 86(1), pp. 32–51. doi: 10.1215/00031283-1277501.

Podesva, R. J., Roberts, S. J. and Campbell-Kibler, K. (2001) 'Sharing resources and indexing meanings in the production of gay styles', in Campbell-Kibler, K., Podesva, Robert J, Roberts, S. J., and Wong, A. (eds) *Language and Sexuality: Contesting Meaning in Theory and Practice*. CSLI Publications, pp. 175–189. Available at: http://www.ling.ohio-state.edu/~kbck/Podesva_ea02.pdf.

Podesva, R. J. and Sharma, D. (2013) *Research Methods in Linguistics*. Cambridge, UK: Cambridge University Press.

Pope, J., Meyerhoff, M. and Ladd, D. R. (2007) 'Forty years of language change on Martha's Vineyard', *Language*, 83(3), pp. 615–627. doi: 10.1353/lan.2007.0117.

Porter, A. J. and Hellsten, I. (2014) 'Investigating participatory dynamics through social media using a multideterminant "frame" Approach: The case of climategate on YouTube', *Journal of Computer-Mediated Communication*, 19(4), pp. 1024–1041. doi: 10.1111/jcc4.12065.

Postil, J. (2017) 'The Diachronic Ethnography of Media: From Social Changing to Actual Social Changes', *Moment Journal*, 4(1), pp. 19–43. doi: 10.17572/mj2017.1.1943.

Postill, J. and Pink, S. (2012) 'Social media ethnography: The digital researcher in a messy web', *Media International Australia*, 145, pp. 123–134.

QSR International (no date a) *Import from YouTube*. Available at: http://help-nv11.qsrinternational.com/desktop/procedures/import_from_youtube.htm (Accessed: 21 August 2020).

QSR International (no date b) *NVivo*. Available at: https://www.qsrinternational.com/nvivo-qualitative-data-analysis-software/home (Accessed: 21 August 2020).

Queen, R. (2013) 'Working with Performed Language: Movies, Television, and Music', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. New York: Routledge, pp. 217–227.

Redi, M., O'Hare, N., Schifanella, R., Trevisiol, M. and Jaimes, A. (2014) '6 seconds of sound and vision: Creativity in micro-videos', *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 4272–4279. doi: 10.1109/CVPR.2014.544.

Reed, P. E. (2020) 'Place and language: Links between speech, region, and connection to place', *Wiley Interdisciplinary Reviews: Cognitive Science*, (January), pp. 1–11. doi: 10.1002/wcs.1524.

Reinikainen, H., Munnukka, J., Maity, D. and Luoma-aho, V. (2020) '"You really are a great big sister"– parasocial relationships, credibility, and the moderating role of audience comments in influencer marketing', *Journal of Marketing Management*. Routledge, 36(3–4), pp. 279–298. doi: 10.1080/0267257X.2019.1708781.

Richards, H. (2008) *Mechanisms, motivations and outcomes of change in Morley (Leeds) English.* University of York.

Robinson, L. and Schulz, J. (2009) 'New avenues for sociological inquiry: Evolving forms of ethnographic practice', *Sociology*, 43(4), pp. 685–698. doi: 10.1177/0038038509105415.

Rodríguez, S. (2019) 'The Baptist Pastor Persona : A sociophonetic case study of vowel stability across a lifespan Results of CMNE Features', in *New Ways of Analyzing Variation 48*.

Rossi, B. (2020) 'Barbara Rossi'. YouTube. Available at: https://www.youtube.com/user/ThePersianbabe.

RStudio Inc. (2019) 'RStudio'.

Rymes, B. and Leone-Pizzighella, A. (2018) 'YouTube-based accent challenge narratives : Web 2.0 as a context for studying the social value of accent', *International Journal of the Sociology of Language*, (250), pp. 0–23. doi: 10.1515/ijsl-2017-0058.

Sankoff, G. (2018) 'Before There Were Corporas: The Evolution of the Montreal French Project as a Longitudinal Study', in Evans Wagner, S. and Buchstaller, I. (eds) *Panel Studies of Variation and Change*. New York, NY: Taylor & Francis, pp. 21–52.

Santos Muñoz, A. (2016) 'Attending Multi-Party Videoconference Meetings: The Initial Problem', *Language@Internet*, 13(3).

Sayers, D. (2014a) 'Model answers: A rejoinder to the responses to the mediated innovation model', *Journal of Sociolinguistics*, 18(2), pp. 272–276. doi: 10.1111/josl.12076.

Sayers, D. (2014b) 'The mediated innovation model: A framework for researching media influence in language change', *Journal of Sociolinguistics*, 18(2), pp. 185–212. doi: 10.1111/josl.12069.

Schilling-Estes, N. (1998) 'Investigating "self-conscious" speech : The performance register in Ocracoke English', *Language in Society*, 27(1), pp. 53–83.

Schilling, N. (2013a) 'Investigating stylistic variation', in Chambers, J. and Schilling, N. (eds) *The Handbook of Language Variation and Change*. 2nd edn. Chichester: Wiley-Blackwell, pp. 327–349.

Schilling, N. (2013b) *Sociolinguistic Fieldwork*. Cambridge: Cambridge University Press.

Schleef, E. (2013) 'Glottal replacement of /t/ in two British capitals: Effects of word frequency and morphological compositionality', *Language Variation and Change*, 25(2), pp. 201–223. doi: 10.1017/S0954394513000094.

Schuler, D. (1994) 'Social Computing', *Communications of the ACM*, p. 29. doi: 10.1145/175222.175223.

Seargeant, P., Tagg, C. and Ngampramuan, W. (2012) 'Language choice and addressivity strategies in Thai-English social network interactions', *Journal of Sociolinguistics*, 16(4), pp. 510–531. doi: 10.1111/j.1467-9841.2012.00540.x.

Senft, T. M. (2008) *Camgirls: Celebrity & community in the age of social networks*. New York, NY: Peter Lang.

Shannon, C. E. and Weaver, W. (1998) *The mathematical theory of communication*. Urbana: University of Illinoid Press.

Shapp, A. A., LaFave, N. and Singler, J. V. (2014) 'Ginsburg v. Ginsburg: A Longitudinal Study of Regional Features in a Supreme Court Justice' s Speech',

University of Pennsylvania Working Papers in Linguistics, 20(2), pp. 149--158. Available at: http://repository.upenn.edu/pwpl/vol20/iss2/17/.

Sharma, B. K. (2012) 'Beyond social networking: Performing global Englishes in Facebook by college youth in Nepal', *Journal of Sociolinguistics*, 16(4), pp. 483–509. doi: 10.1111/j.1467-9841.2012.00544.x.

Sharma, D. (2011) 'Style repertoire and social change in British Asian English', *Journal of Sociolinguistics*, 15(4), pp. 464–492.

Sheffield, E. (2014) *The Vogue Interview: Zoella*, *Vogue*.

Shoemark, P., Sur, D., Shrimpton, L., Murray, I. and Goldwater, S. (2017) 'Aye or naw, whit dae ye hink? Scottish independence and linguistic identity on social media', *15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017 - Proceedings of Conference*, 1, pp. 1239–1248. doi: 10.18653/v1/e17-1116.

Siebenhaar, B. (2006) 'Code choice and code-switching in Swiss-German Internet Relay Chat rooms', *Journal of Sociolinguistics*, 10(4), pp. 481–506. doi: 10.1111/j.1467-9841.2006.00289.x.

Silverman, K., Beckman, M., Pitrelli, Jo., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J. and Hirschberg, J. (1992) 'ToBI: a standard for labeling English prosody', in *Proceedings of the International Conference on Spoken Language Processing*. Alberta, Canada, pp. 867–870.

Silverstein, M. (2003) 'Indexical order and the dialectics of sociolinguistic life', *Language and Communication*, 23(3–4), pp. 193–229. doi: 10.1016/S0271-5309(03)00013-2.

Singh, A. (2017) 'Hello World Live: organisers of YouTubers' convention apologise to disappointed fans after flood of complaints', *The Telegraph*, 30 October. Available at: https://www.telegraph.co.uk/news/2017/10/30/hello-world-live-organisers-youtubers-convention-apologise-disappointed/.

Smart, N. (2020) 'Niomi Smart'. YouTube. Available at: https://www.youtube.com/user/niomismart.

Smith, J. and Holmes-Elliott, S. (2017) 'The unstoppable glottal: tracking rapid change in an iconic British variable 1', *English Language and Linguistics*, 22(3),

pp. 1–33. doi: 10.1017/S1360674316000459.

Snell, J. (2010) 'From sociolinguistic variation to socially strategic stylisation', *Journal of Sociolinguistics*, 14(5), pp. 630–656. doi: 10.1111/j.1467-9841.2010.00457.x.

Snell, J. (2017) 'Enregisterment, Indexicality and the Social Meaning of Howay: Dialect and Identity in North-East England', in Montgomery, C. and Moore, E. (eds) *Language and a Sense of Place: Studies in Language and Region*. Cambridge, UK: Cambridge University Press, pp. 301–324.

Sonderegger, M., Bane, M. and Graff, P. (2017) 'The medium-term dynamics of accents on reality television', *Language*, 93(3), pp. 598–640. doi: https://doi.org/10.1353/lan.2017.0038.

Stanley, J. A. and Renwick, M. E. L. (2016) 'Phonetic shift /ɔr/ phonemic change? American English mergers over 40 years', in *The 15th Conference on Laboratory Phonology (LabPhon15)*.

Stanton, K. (2006) *A study of high rising terminals in various dialects of English*. University of Canterbury.

Stoddart, J., Upton, C. and Widdowson, J. D. A. (1999) 'Sheffield dialect in the 1990s: Revisiting the concept of NORMs.', in Foulkes, P. and Docherty, G. (eds) *Urban Voices: Accent Studies in the British Isles*. London: Routledge, pp. 72–89.

Strelluf, C. (2019) 'Anymore, it's on Twitter: Positive anymore, American regional dialects, and polarity licensing in tweets', *American Speech*, 94(3), pp. 313–351. doi: 10.1215/00031283-7587883.

Stuart-Smith (no date) 'Is TV a contributory factor in accent change in adolescents?' Economic and Social Research Council.

Stuart-Smith, J. (1999) 'Glottals past and present: a study of T-glottalling in Glaswegian', *Leeds Studies in English*, 30, pp. 181–204.

Stuart-Smith, J. (2014) 'No longer an elephant in the room', *Journal of Sociolinguistics*, 18(2), pp. 250–261. doi: 10.1111/josl.12071.

Stuart-Smith, J., Sonderegger, M. and Mielke, J. (no date) 'SPeech Across

Dialects of English (SPADE): Large-scale digital analysis of a spoken language across space and time (2017-2020). HRC/CRSH Grant 869-2016-0006, NSF Grant SMA-1730479. (Digging into Data/Trans-Atlantic Platform).' ESRC Grant ES/R003963/1, NSERC/CRSNG Grant RGPDD 501771-16, SS. Available at: https://spade.glasgow.ac.uk/project-team/.

Stuart-Smith, J., Timmins, C. and Tweedie, F. (2007) '"Talkin" Jockney'? Variation and change in Glaswegian accent', *Journal of Sociolinguistics*, 11(2), pp. 221–260. doi: 10.1111/j.1467-9841.2007.00319.x.

Sugg, J. (2020) 'ThatcherJoe'. YouTube. Available at: https://www.youtube.com/user/ThatcherJoe.

Sugg, Z. (2012) 'What's In My Bag? (Winter Edition) | Zoella'. Zoella. Available at: https://www.youtube.com/watch?v=VG-N_CogQMY.

Sugg, Z. (2013a) 'Florida Day 1 | YouTuber Arrival', *More Zoella*. Available at: https://www.youtube.com/watch?v=q0RMeEFYPK0.

Sugg, Z. (2013b) 'Main Stage & Starships Duet', *More Zoella*. Available at: https://www.youtube.com/watch?v=KT3RiXdPlGs.

Sugg, Z. (2013c) 'Summer In The City 2013'. YouTube. Available at: https://www.youtube.com/watch?v=UocyyPoGGkk.

Sugg, Z. (2014a) *Girl Online*. Penguin.

Sugg, Z. (2014b) 'Meet Up, Poolside Chats & "Suggwich" | Florida Day Six', *More Zoella*. Available at: https://www.youtube.com/watch?v=8VgIKJB452Q.

Sugg, Z. (2014c) 'The Best And Last Day Of Vidcon', *More Zoella*. Available at: https://www.youtube.com/watch?v=pKTRPCAs0WU.

Sugg, Z. (2015a) 'An Amazing Weekend In Florida | Playlist Live 2015'. YouTube. Available at: https://www.youtube.com/watch?v=pWMziMQIy6Q.

Sugg, Z. (2015b) 'An Amazing Weekend In Florida | Playlist Live 2015', *More Zoella*. Available at: https://www.youtube.com/watch?v=pWMziMQIy6Q.

Sugg, Z. (2015c) 'Christmas Beauty Products'. YouTube. Available at: https://www.youtube.com/watch?v=8t6-G4hIvmc.

Sugg, Z. (2015d) *Girl Online: On Tour*. Penguin.

Sugg, Z. (2015e) 'I HAVE NEW BEAUTY PRODUCTS!' YouTube. Available at: https://www.youtube.com/watch?v=O-FznoiffDk.

Sugg, Z. (2016a) 'Autumn Bath & Body Works Haul | Zoella'. Zoella, 31 August. Available at: https://www.youtube.com/watch?v=nxs6DCR2SDM.

Sugg, Z. (2016b) 'FIRST LOOK AT NEW ZOELLA BEAUTY CHRISTMAS RANGE'. YouTube. Available at: https://www.youtube.com/watch?v=oZIisPrMMIU.

Sugg, Z. (2016c) *Girl Online: Going Solo*. Penguin.

Sugg, Z. (no date a) 'Zoella', *YouTube*. Available at: https://www.youtube.com/zoella (Accessed: 20 July 2018).

Sugg, Z. (no date b) 'Zoella'. YouTube. Available at: https://www.zoella.co.uk/ (Accessed: 20 July 2018).

Sutton, S. (2020) *Project: How to conduct sociophonetics research in online public video*, *ResearchGate*. Available at: https://www.researchgate.net/project/How-to-conduct-sociophonetics-research-in-online-public-video.

Tagliamonte, S. A. (2006) *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press.

Tagliamonte, S. A. (2012) *Variationist Sociolinguistics: Change, Observation, Interpretation*. Chichester, UK: Wiley-Blackwell.

Tagliamonte, S. A. (2016) 'So sick or so cool? the language of youth on the internet', *Language in Society*, 45(1), pp. 1–32. doi: 10.1017/S0047404515000780.

Tagliamonte, S. A. (2017) 'Changing Places: Tracking Innovation and Obsolescence across Generations', in *Language and a Sense of Place*. Cambridge, UK: Cambridge University Press, pp. 15–37.

Tagliamonte, S. A. and Denis, D. (2008) 'Linguistic Ruin? Lol! Instant Messaging and Teen Language', *American Speech*, 83(1), pp. 3–34. doi: 10.1215/00031283-2008-001.

Tagliamonte, S. and D'Arcy, A. (2004) 'The Quotative System in Canadian

Youth', *Journal of Sociolinguistics*, 8, pp. 493–514. doi: 10.1111/j.1467-9841.2004.00271.x.

Tagliamonte, S. and Hudson, R. (1999) 'Be like et al. beyond America: The quotative system in British and Canadian youth', *Journal of Sociolinguistics*, 3(2), pp. 147–172. doi: 10.1111/1467-9481.00070.

Tannen, D. and Trester, A. M. (2013) *Discourse 2.0: Language and new media*. Edited by D. Tannen and A. M. Trester. Washington: Georgetown University Press.

Tatman, R. (2017) '"I'm a spawts guay": Comparing the Use of Sociophonetic Variables in Speech and Twitter', *University of Pennsylvania Working Papers in Linguistics*, 18(January), pp. 1–10.

The Advertising Standards Agency and The Committee of Advertising Practice (2020) 'Recognising ads: Blogs and vlogs', 5 February. Available at: https://www.asa.org.uk/advice-online/recognising-ads-blogs-and-vlogs.html (Accessed: 8 May 2020).

Thurlow, C., Lengel, L. and Tomic, A. (2004) *Computer Mediated Communication. Social Interactions and the internet*. London: SAGE Publications.

Thurlow, C. and Mroczek, K. (2011) *Digital Discourse: Language in the new media*. Edited by C. Thurlow and K. Mroczek. Oxford: Oxford University Press.

TikTok (2019) *TikTok*. Available at: https://www.tiktok.com/en/ (Accessed: 22 October 2019).

Tollfree, L. (1999) 'South East London English: Discrete versus continuous modelling ofconsonantal reduction', in Foulkes, P. and Docherty, G. (eds) *Urban voices: accent studies in the British Isles*. London: Routledge, pp. 163–184.

Tolson, A. (2010) 'A new authenticity? Communicative practices on YouTube', *Critical Discourse Studies*, 7(4), pp. 277–289. doi: 10.1080/17405904.2010.511834.

Topshop (2018) 'Topshop's Best Design Collaborations Ever', *Topshop Blog*, 30 August. Available at: https://www.topshop.com/blog/2018/08/topshops-best-

design-collaborations-ever.

Townsend, L. and Wallace, C. (2016) 'Social Media Research: A Guide to Ethics'. University of Aberdeen. Available at: https://www.gla.ac.uk/media/Media_487729_smxx.pdf.

Trechter, S. (2013) 'Social Ethics for Sociolinguistics', in Mallinson, C., Childs, B., and Van Herk, G. (eds) *Data Collection in Sociolinguistics: Methods and Applications*. Oxon, UK: Routledge, pp. 33–45.

Trudgill, P. (1974) *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Trudgill, P. (1988) 'Norwich revisited: Recent linguistic changes in an English urban dialect.', *English World-Wide*, 9, pp. 33–49.

Trudgill, P. (1999) 'Norwich: Endogenous and exogenous change.', in Foulkes, P. and Docherty, G. (eds) *Urban Voices: Accent Studies in the British Isles*. Routledge, pp. 124–140.

Trudgill, P. (2000) *Sociolinguistics: An Introduction to Language and Society*. 4th edn. London, UK: Penguin Books.

Trudgill, P. (2014) 'Diffusion, drift, and the irrelevance of media influence', *Journal of Sociolinguistics*, 18(2), pp. 213–222. doi: 10.1111/josl.12070.

Twitch (2019) *Twitch*. Available at: https://www.twitch.tv/ (Accessed: 22 October 2019).

Twitter (2019) *Twitter*. Available at: www.twitter.com/ (Accessed: 22 October 2019).

Twitter (2020a) *Academic Research*. Available at: https://developer.twitter.com/en/solutions/academic-research (Accessed: 20 August 2020).

Twitter (2020b) *Twitter API*. Available at: https://developer.twitter.com/en/products/twitter-api (Accessed: 20 August 2020).

UK Data Service (2020) *UK Data Service*, *Webpage*. Available at: https://www.ukdataservice.ac.uk/ (Accessed: 21 August 2020).

Uldall, E. (1964) 'Dimensions of meaning in intonation', in Abercrombie, D., Fry, D. B., MacCarthy, P. A. D., Scott, N. D., and Trim, J. L. M. (eds) *In Honour of Daniel Jones: Papers contributed on the occasion of his eightieth birthday*. London: Longman, pp. 271–279.

Unknown (no date) *OnlineVideoConverterV3.0.*

Vanderweele, T. J. (2015) *Explanation in Causal Inference: Methods for Mediation and Interaction*. Oxford, UK: Oxford University Press.

Varis, P. (2016) 'Digital Ethnography', in Georgakopoulou, A. and Spilioti, T. (eds) *Routledge Handbook of Language and Digital Communication*. London: Routledge, pp. 55–68.

Vine (2019) *Vine*.

Vitak, J., Proferes, N., Shilton, K. and Ashktorab, Z. (2017) 'Ethics Regulation in Social Computing Research: Examining the Role of Institutional Review Boards', *Journal of Empirical Research on Human Research Ethics*, 12(5), pp. 372–382. doi: 10.1177/1556264617725200.

Vitak, J., Shilton, K. and Ashktorab, Z. (2016) 'Beyond the Belmont principles: Ethical challenges, practices, and beliefs in the online data research community', *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*, 27, pp. 941–953. doi: 10.1145/2818048.2820078.

Wallace, V. (2006) '"Whey Aye My Good Sir": Has Cheryl Fernandez-Versini's Accent Moved from Tyneside English to RP?', *Lifespans and Styles*, 2(2), pp. 10–19. doi: 10.2218/ls.v2i2.2016.1609.

Warren, P. (2016) *Uptalk: The phenomenon of rising intonation*. Cambridge: Cambridge University Press.

Watson, K. (2007) 'Liverpool English', *Journal of the International Phonetic Association*, 37(3), pp. 351–360. doi: 10.1017/S0025100307003180.

Wells, J. C (2006) *English Intonation. An Introduction*. Cambridge: Cambridge University Press.

Wells, J. C. (1982) *Accents of English 1: An introduction*. Cambridge, UK: Cambridge University Press.

Wesch, M. (2009) 'YouTube and You', *Explorations in Media Ecology*, pp. 19–34.

Whitmarsh, E. (2017) *Are Your Formants Facebook Famous? The Impact of Social Media Data Compression on Acoustic Features of Speech*. University of York, UK.

Wieling, M., Grieve, J., Bouma, G., Fruehwald, J., Coleman, J. and Liberman, M. (2016) 'Variation and change in the use of hesitation markers in Germanic languages', *Language Dynamics and Change*, 6(2), pp. 199–234.

Wilkinson, S. (2017) 'Zoella Sells Overpriced Advent Calendar, Because She Is A Business', *Grazia Daily*, 14 November. Available at: https://graziadaily.co.uk/life/opinion/zoella-advent-calendar-review/.

Williams, A. and Kerswill, P. (1999) 'Dialect levelling: change and continuity in Milton Keynes, Reading and Hull', *Urban Voices: Accent Studies in the British Isles*, (January 1999), pp. 141–162.

Williams, M. L., Burnap, P. and Sloan, L. (2017) 'Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation', *Sociology*, 51(6), pp. 1149–1168. doi: 10.1177/0038038517708140.

Witmer, D. F. and Katzman, S. L. (1997) 'On-Line Smiles: Does Gender Make a Difference in the Use of Graphic Accents?', *Journal of Computer-Mediated Communication*, 2(4), pp. 0–0. doi: 10.1111/j.1083-6101.1997.tb00192.x.

Wolfram, W. (1969) *A Sociolinguistic Description of Detroit Negro Speech*. Washington, DC: Center for Applied Linguistics.

YouTube (2019a) *YouTube Creator Academy*. Available at: https://creatoracademy.youtube.com/page/home (Accessed: 22 October 2019).

YouTube (2019b) *YouTube Kids*. Available at: https://www.youtube.com/kids/ (Accessed: 22 October 2019).

YouTube (2019c) *YouTube Music*. Available at: https://music.youtube.com/ (Accessed: 22 October 2019).

YouTube (2019d) *YouTube Premium*. Available at:

https://www.youtube.com/premium (Accessed: 22 October 2019).

YouTube (2019e) *YouTube Press*. Available at:
https://www.youtube.com/about/press/ (Accessed: 22 October 2019).

YouTube (2019f) *YouTube TV*. Available at: https://tv.youtube.com/welcome/
(Accessed: 22 October 2019).

YouTube (2020a) *API Reference*. Available at: https://developers-dot-devsite-
v2-prod.appspot.com/youtube/v3/docs (Accessed: 24 August 2020).

YouTube (2020b) *YouTube Engineering and Developers Blog*. Available at:
https://youtube-eng.googleblog.com/.

YouTube Creator Academy (2020) *'The Algorithm' - How YouTube Search &
Discovery Works*, *YouTube Creator Academy*. Available at:
https://creatoracademy.youtube.com/page/lesson/discovery?cid=get-
discovered&hl=en-GB (Accessed: 24 August 2020).

YouTube Great Britain (2010) *Terms of Service*, *Terms of Service*. Available at:
https://www.youtube.com/static?gl=GB&template=terms.

YouTube Great Britain (2020) *Get insights with YouTube Analytics*, *YouTube
Creator Academy*. Available at:
https://creatoracademy.youtube.com/page/course/analytics-series?hl=en-GB
(Accessed: 1 September 2020).

YouTube Team (2013) 'So long, video responses...Next up: better ways to
connect', *YouTube Official Blog*, 27 August.

Zhang, Q. (2008) 'Rhotacization and the "Beijing Smooth Operator":', *Journal
of Sociolinguistics*, 12(2), pp. 201–222. doi: 10.1386/jmpr.7.1.43/1.

Zimmer, M. (2010) '"But the data is already public": On the ethics of research in
Facebook', *Ethics and Information Technology*, 12(4), pp. 313–325. doi:
10.1007/s10676-010-9227-5.

Zimmer, M. and Proferes, N. J. (2014) 'A topology of twitter research:
Disciplines, methods, and ethics', *Journal of Information Management*, 66(3),
pp. 250–261. doi: 10.1108/AJIM-09-2013-0083.

Zoella (2019) *Zoella - Videos*, *YouTube*. Available at:

https://www.youtube.com/user/zoella280390/videos?view=0&sort=dd&flow=gri
d (Accessed: 12 December 2019).