



UNIVERSITY OF GOTHENBURG
DEPT OF LANGUAGES AND LITERATURES

ENGLISH

**A Linguistic Introduction to the Discursive
Conventions of *Twitter***

Andrew Symes

Advanced research essay, linguistics
Spring term 2012
Supervisor: Joe Trotta
Examiner: Jennifer Herriman

Title: *A Linguistic Introduction to the Discursive Conventions of 'Twitter'*

Author: Andrew Symes

Supervisor: Joe Trotta

Abstract: This paper provides an inductively-derived overview of some of the pertinent linguistically-related, empirically-observable discursive phenomena to be found on Twitter, a popular microblogging site used to publish and exchange messages constrained to 140 characters. A corpus of 11,187 'tweets' derived from 100 public Twitter users over a 48-hour period forms the basis for a Computer-Mediated Discourse Analysis approach to the study of user habits, the communicative function of tweets, and three principal "conventions" that help users manage their discourse: use of the '@ symbol', 'retweeting' and 'hashtagging'. The findings reveal that Twitter is used in diverse ways such that users neither constitute a homogenous mass, nor can be easily categorised according to their habits. However, Twitter serves as a key medium for inter-user communication, the maintenance of social relationships, and the exhibitionistic practice of identity performance.

Keywords: Twitter, Microblogging, Computer-Mediated Communication, New Media, Web 2.0, Social Media, Linguistics

Contents

1. Introduction.....	1
2. Background.....	3
2.1. Computer-mediated & new media communication.....	3
2.2. Web 2.0, user-generated content & social media.....	6
2.3. Microblogging.....	8
2.4. Twitter.....	9
2.4.1. Features & conventions.....	9
2.4.2. Popularity.....	12
3. Aims & methodology.....	14
3.1. Research aims.....	14
3.2. Ethical considerations.....	14
3.3. Data.....	14
3.4. Analytical approach.....	15
3.4.1. Computer-mediated discourse analysis.....	15
3.4.2. Content analysis.....	16
3.4.3. Software.....	16
4. Findings & discussion.....	17
4.1. Basic usage.....	17
4.2. Communicative function of tweets.....	17
4.3. @ symbol.....	20
4.4. Retweeting.....	23
4.5. Hashtagging.....	28
5. Conclusion.....	32
6. References.....	34
7. Appendices.....	37
Appendix A. User statistics.....	37
Appendix B. Tweet frequency graphs.....	39
Appendix C. Functions of the @ sign.....	40
Appendix D. Most frequent hashtags.....	41

1. Introduction

As a result of the enormous impact of emergent communication technologies, behavioural and social norms have gradually altered (cf. Baron, 2008). In recent years, the Internet in particular has forged a position at the very heart of modern society; over a third of the world's population is now estimated to be online (ITU, 2011). This rapid and often unpredictable evolution of the Internet is said to represent one of the greatest challenges to contemporary scholars (Schneider & Foot, 2005). As textual language is one of the most pervasive and visible manifestations of 'new media technologies', e.g. the Internet, smart phones and mobile telephony, among others, these media are of particular interest to linguists, and an ever-increasing body of research is devoted to the study of 'Language Online', 'Language on the Internet' or 'Internet Linguistics' (see Section 2.1¹). A particular challenge is that the Internet allows users to circumvent the traditional gatekeepers of the published word; user-generated content has emerged as one of the defining features of a fundamentally different online environment, i.e. 'Web 2.0' (for a definition and discussion, see Section 2.2).

Amongst the vast *milieu* of diffuse internet uses, so-called online social media (see Section 2.2) have recently emerged to be the most popular family of applications since the launch of the World Wide Web² in the early 1990s (Kaplan & Haenlein, 2010). Well-known examples of such sites include the social networking site *Facebook*, the video sharing platform *YouTube*, the image hosting site *Flickr*, and the microblogging application *Twitter*, the focus of the research conducted in the present study. Launched in October 2006, Twitter has indisputably become the most popular microblogging application available. The service enables users to post messages – 'tweets' – about their activities, opinions and status at any moment using small elements of user-generated content of up to 140 characters such as short sentences, individual images, or video links to a public virtual audience (Section 2.3.1 provides a detailed account of Twitter's formal features and conventions). Its rise to prominence has been meteoric; although Twitter does not consistently publish usage statistics, the number of registered users has grown at a substantial rate, from an estimated 94,000 in early 2007 (Java et al., 2007) to approximately 500 million, of which 140 million were considered "active", as of February 2012 (MediaBistro, 2013; Guardian, 2013). Its popularity, discussed further in Section 2.3.2) stems principally from its light-weight framework, i.e. that users are not burdened by the need for significant investments of time, thought or cost (Java et al., 2007), and its open network, which allows users to freely contact others without any technical requirements, and usually without social expectations, for reciprocity (Marwick & boyd³, 2010).

¹ The term *Internet* is somewhat misleading as some related forms of communication take place "offline", e.g. via intranets and mobile telephony. *Internet* is therefore extended to include these related communication technologies.

² The Web and the Internet have a part-to-whole relationship: the larger entity, the Internet, is the entire technological infrastructure; the Web is an extensive software subset dedicating to broadcasting HTML pages.

³ This unconventional orthography adheres to boyd's own "political irritation at the importance of capitalisation", <http://www.danah.org/aboutme.html> [accessed: 16-02-2013]

Twitter's exponential growth partly justifies this study; while previous research focusing on the medium is hardly scarce, the communicative features and conventions found on the platform can be expected to have evolved as the user-base has increased, with the appropriation of online literary practices contributing to an increasingly diverse environment. The aforementioned study by Java et al. (2007), for example, however useful and insightful, cannot be considered representative of Twitter use in 2012 having been based on so few users compared with current volumes. Therefore, whilst nevertheless important to the understanding of platform, these early accounts provide only diachronic snapshots of Twitter at a particular period in its development and continued study of the application is necessary if it is to be fully understood.

Aside from its popularity, the types of communication afforded by the application make it well worthy of research. As Zappavigna (2011:790) points out, tweets constitute "interesting cases in making meaning within constrained environments", while Twitter has an open and public network that represents a challenging context in which to negotiate social relationships, both individually and with a broader audience.

The aim of this study is to provide an inductively derived, preliminary assessment of some of the pertinent linguistically-related, empirically-observable discursive phenomena to be found on Twitter based on an examination of the communicative habits of 100 users over a 48-hour period. The study is split into 5 research "modules". First, the basic usage of tweets is examined in an attempt to establish how much and how often users 'tweet'. Second, a macro-level overview of the communicative functions of tweets is sought via categorisation of the different purposes that tweets serve. The final three modules concern three discursive Twitter "conventions" – uses of the '@ symbol', 'retweeting' and 'hashtagging' (Section 2.3.1 provides further details) – that have emerged since Twitter was launched. Each of these modules is guided by two principal questions: how prevalent are these functions and what purposes do they serve?

The paper is organised as follows: in Section 2, consideration is given to background concepts that will contextualise the study of Twitter, namely 'computer-mediated or new media communication', 'Web 2.0', 'user-generated content' and 'social media', as well as introductions to 'microblogging' as a practice and Twitter as an application, including considerations to its immense popularity; in Section 3, the research aims and methodology are described, including the data set, its collection and the approaches employed to analyse it; in Section 4 the findings of the analysis are presented and discussed; and finally, Section 5 concludes this study with a summary and some closing comments.

2. Background

2.1. Computer-mediated & new media communication

The term ‘computer-mediated communication’ (CMC) emerged in the 1980s to describe the digital means employed to create and transmit messages, and encompasses a variety of technologies such as email, forums, virtual reality role-playing games, chat and instant messaging (Baron, 2008). After initially being restricted to users in public and commercial institutions, textual CMC rapidly rose to domestic prominence after homes were brought “online” in the late 1980s and early 1990s, and has flourished ever since (Herring, 2010). As Georgakopoulou (2011:93) points out, “it is hardly an exaggeration to claim that CMC has truly revolutionised social interaction, at least in technologically advanced societies.”

Although communication is becoming increasingly multimodal and the Internet semiotically diverse (see *Web 2.0* below), most CMC remains fundamentally text-based; messages are “typed” on an input device such as a keyboard or keypad, and read as text on a screen, typically by a recipient at a different location (Herring, 2001). Indeed, it is this “textual trace” which makes online social activities more accessible to social scientific scrutiny and theory than is the case with ephemeral spoken communication (Herring, 2004). Despite the enormous amount of technological progress in the decade following its publication, Herring’s definition of CMC remains pertinent:

Text-based CMC takes a variety of forms [...] whose linguistic properties vary depending on the kind of messaging system used and the social and cultural embedding particular instances of use. However, such forms have in common that the activity that takes place through them is constituted primarily – in many cases, exclusively – by visually presented language. These characteristics of the medium have important consequences for understanding the nature of computer-mediated language. They also provide a unique environment, free from competing influences from other channels of communication and from physical context, in which to study verbal interaction and the relationship between discourse and social practice. (2001:612)

However, the term CMC is problematic and requires further consideration. Herring herself (2001) makes a clear distinction between CMC as a broader interdisciplinary study, and computer-mediated *discourse* (CMD), a specialisation which focuses on language and language use in computer-networked environments. Baron (2008), meanwhile, points to the emergence of devices that cannot be classed as computers, such as mobile phones, and thus offers “electronically mediated communication” (EMC) as a more appropriate label. Indeed, Twitter is a platform which does not necessitate access to a computer⁴. Baron’s reference to *electronic*, however, surely makes this label too broad; should, for example, all communication via television be considered under the same field of research? Furthermore, Crystal (2011) dismisses the use of *communication*, criticising it for being too broad, and considering it to blur the distinction between language and other forms of com-

⁴ Krishnamurthy et al. (2008) found that circa 7.5% of tweets are sent from mobile phones (and 61.7% from the Web), while a report into social media usage in the US (Nielsen, 2012) found that 39% of users access sites via mobile phones.

munication. In many respects, this distinction between communication and language perspective is valid, but it conflicts with the researching of environments which are fundamentally multi-modal and visually rich; for example, much user-generated content, one of the foundations of Web 2.0 (see below), concerns the convergence of text and images (see Trotta & Danielsson, 2011). Crystal's focus on language online, meanwhile, led him to initially champion the "pop-linguistic" term 'Netspeak' as an alternative to CMC (2006; 2011), but he has since abandoned the term due to its failure to adequately account for the linguistic idiosyncrasies of language found online, which is portrayed as a homogeneous variety. Crystal now favours 'Internet Linguistics' as "the most convenient name for the scientific study of all manifestations of language in the electronic medium" (2011:2). Focusing on the Internet, however, excludes from this field of research communicative forms with shared properties which function offline, such as text messaging on mobile phones and intranet platforms, seemingly rendering the label somewhat inappropriate.

A further alternative is the interdisciplinary term 'new media', as adopted by the journals *New Media & Society* and *Convergence: The International Journal of Research into New Media Technologies*. Determining what constitutes new media, and by extension "old media", represents an obvious problem, particularly when technologies evolve at a tremendous rate. Nevertheless, though imperfect, new media proves attractive as it emphasises the organic advancement in the ways we utilise emerging technologies. Furthermore, its genericness provides a forum in which can converge a multitude of related disciplines sharing these technologies as a common focus of research. This study therefore advocates its wider adoption as an appropriate title for a collective research profile, and henceforth substitutes the term new media communication (NMC) for predecessors such as CMC where appropriate. Specific terms such as 'new media technologies', 'new media communication', and 'new media linguistics⁵/semiotics/discourse' augment an organisation of research with a consistent common identity.

The themes and subjects of new media linguistics (NML) vary widely. Many of the "first wave" of NML studies have hitherto been devoted to mapping the formal features of NMC (e.g. spelling and orthography), and contrasting NMC with the prototypical features of speech and writing (Androutsopoulos, 2006; Thurlow & Mroczek, 2011); the general consensus, despite some debate, is that NMC is essentially a mixed modality, i.e. a hybrid combination of written and spoken features (cf. Baron, 2010; Crystal, 2011; Georgakopoulou, 2011). Furthermore, linguistic descriptions often accentuate the distinction between synchronous (e.g., e-chat, instant messaging) and asynchronous (e.g., email, texting, blogs) modes of communication (Androutsopoulos, 2006). Baron (2008; 2010), for example, suggests that the two parameters according to which NMC can be defined structurally are synchronicity and audience scope, i.e. the contrast between one-to-one (i.e. between two people) and one-to-many (i.e. involving multiple recipients) interactions. These para-

⁵ A term already adopted by Thurlow & Mroczek (2011)

digmatic distinctions produce four classes of NMC, although Baron concedes that in practice users often cross category lines (2010):

Table 1. Types of computer-mediated communication (Baron, 2010:1)

	<i>synchronous</i>	<i>asynchronous</i>
<i>one-to-one</i>	instant messaging	email, texting on mobile phones
<i>many-to-many</i>	chat, computer conferencing	blogs, social networking sites

However, although analyses of formal textual features remain integral, some NMC research has been criticised for perpetuating Internet language myths, such as the popular misconceptions of its negative impact on offline language (Thurlow & Bell, 2009; Crystal, 2011), and of it being distinct, homogenous and indecipherable to “outsiders” (Androutsopoulos, 2006); “Internet research often suffers from a premature impulse to label online phenomena in broad terms, e.g., all groups of people interacting online are ‘communities’; the language of the Internet is a single style or ‘genre’” Herring (2004:1). Androutsopoulos (2006) points in particular to Crystal’s attempts (2006) to define language on the Internet as a unique variety, i.e. Netspeak. Baron (2010), meanwhile, highlights issues with: generalisations made across different genres of NMC, despite usage patterns showing considerable disparity; the ahistorical perspective which ignores the evolution of usage patterns; the opacity of the “offline” data (i.e. of spoken and written language) to which NMC is compared; and the preoccupation of NMC research with many-to-many rather than dyadic communication. Furthermore, Thurlow & Mroczek (2011:28) urge caution against “making overextended claims and wild predictions about the stability or durability of the technolinguistic changes of the moment.”

Nevertheless, linguistic disciplines have begun to recognise the need to explore new avenues of research in order to demythologise the purported homogeneity and highlight the diversity of language use in NMC. In an overview of discourse-pragmatic research, an area within which this current study falls, Georgakopoulou (2011:93) points to the progress made “from treating everything that takes place on the medium as an undifferentiated whole to acknowledging and exploring distinctions amongst computer-mediated discourses that are related to register, style, and genre, or, equally, to system specifications”. From a sociolinguistic perspective, renewed emphasis is being placed on the interplay of technological, social, and contextual factors in the shaping of new media language practices, and the role of linguistic variability in the formation and performance of online social interaction and identities (Androutsopoulos, 2006). Further selected themes central to the current body of linguistic NMC research include: social interaction and interpersonal relations; expressive aspects, such as playfulness, humour and wit; online communities; self-representation and identity performance; online ethnography, including gender; language variation; multilingualism and language choice; connecting online and offline practices; and the hybridity of NMC genres (see Androutsopoulos, 2006; Danet & Herring, 2007; Georgakopoulou, 2011).

2.2. Web 2.0, user-generated content & social media

The term ‘Web 2.0’ was popularised following an influential conference of the same name hosted by the communications entrepreneur Tim O’Reilly in 2004. Although the label 2.0 suggests a new “updated” version of the Web, it does not refer to any single technological advancement, but rather to incremental changes to the ways the Web is used (Wikipedia, 2013a). However, despite the inexistence of any straightforward distinction between “old” and “new” Webs, Web 2.0 environments are said to share technological, sociological, and structural features that clearly separate them from earlier developmental stages (Androutsopoulos, 2011): while “Web 1.0” sites of the mid-1990s were typically single-authored, static and limited users to the passive consumption of content, Web 2.0 sites allow users to interact and collaborate in a social media dialogue in a virtual community (Herring, 2012; Wikipedia, 2013a). Moreover, Web 2.0 refers to the ways in which online content, applications, ideas and knowledge are no longer created and published by individuals, but are instead continuously modified by large communities of users in an iterative, participatory and collaborative process (Bruns, 2008; Kaplan & Haenlein, 2010).

The notion of Web 2.0 is, however, contested; according to the Internet’s inventor, Tim Berners-Lee, the Internet was intended from conception be a “collaborative medium, a place where we [could] all meet and read and write” (Wikipedia, 2013a), while some critics claim it to be a mere marketing buzzword which implies a revolution in web content and use, rather than a more accurate gradual shift (Bruns, 2008; Herring, 2012). Thurlow (2012:5), meanwhile, criticises the “mythology” of Web 2.0, maintaining that “presentism” invariably engenders a distinct lack of consideration for “historicity and precedent”, leading most accounts of Web 2.0 to cite exaggerated, dichotomised characterisations of the “old” and “new”; the “newness” of new media is typically a fabrication, and is almost always a deeply ideological discursive construction (Thurlow, 2012). This issue is addressed by Herring (2012), who introduces a three-part categorisation of online discourse phenomena: ‘familiar’ aspects of NMC carried over from an era prior to Web 2.0; ‘reconfigured’ aspects have been adapted by Web 2.0 environments; ‘emergent’ aspects did not exist, or were not publicly visible, prior to Web 2.0. Herring maintains that the majority of online phenomena, contrary to the impression that everything on the Web today is new and different, can be classified as ‘familiar’: for example, textual language remains the predominant channel of communication.

The term ‘user-generated content’ (UGC) is used to describe the various forms of public media content created by end-users, and can be seen as the sum of the ways in which people utilise ‘social media’ (see below; Kaplan & Haenlein, 2010). Kaplan & Haenlein (2010) stipulate three defining requirements of UGC: first, it must be published on either a public website or a social networking site accessible to a selected group of people, second, it must show a degree of creative effort; and third, it must be created outside of professional routines. This accessibility of localised, bottom-up production and distribution of online content is alternatively referred to as ‘participation’, which

contributes towards the concept ‘online convergence’, the fusion of formerly distinct technologies and modes of communication in integrated digital environments (Androutsopoulos (2011); see also Jenkins, 2008). While UGC was indeed available prior to the Web 2.0 era, e.g. via blogs and discussion boards, the combination of technological (e.g. increased broadband availability and hardware functionality), economic (e.g. wider availability of creative tools), and social factors (e.g. the rise of a generation of “digital natives”⁶, i.e. younger age users with substantial technical knowledge and willingness to engage online) make contemporary UGC intrinsically different (Kaplan & Haenlein, 2010). The majority of UGC, whether it be text, audio, video or static images, is ripe for social scientific research as it constitutes human discourse; Herring (2012) refers to the discourse specifically found in Web 2.0 environments as ‘convergent media computer-mediated discourse’, or ‘Discourse 2.0’.

2.2.1. Social media

In contrast to Web 2.0, which is a broader concept that constitutes an ideological and technological platform, ‘social media’ refers to a group of Internet-based applications that facilitate the production – or ‘produsage’ (Bruns, 2008) – of UGC (Kaplan & Haenlein, 2010). Social media sites are configured using ‘social software’, defined by Coates (2003) as “a particular sub-class of software-prosthesis that concerns itself with the augmentation of human, social, and/or collaborative abilities through structured mediation [which] may be distributed or centralised, top-down or bottom up/emergent).”

Different types of social media include collaborative projects such as *Wikipedia*, blogs such as *Blogger*, microblogs (see below), social networking sites⁷ such as *Facebook*, *Google+* and *LinkedIn*, user-generated media content communities such as *Pinterest*, *Achan*, *Flickr*, and *YouTube*, and virtual social and gaming worlds such as *Second Life* and *World of Warcraft*. Furthermore, social media has become one of the most powerful sources for news (Wikipedia, 2013b).

Kaplan & Haenlein (2010) categorise the principal types of social media using ‘media research’ and ‘social processes’ theories. The media-related component utilises ‘social presence’ and ‘media richness’ theories: social presence postulates that media, influenced by the intimacy and immediacy of the medium, differ in the degree of social presence, i.e. the acoustic, visual and physical contact that can be achieved between two communication partners; media richness states that media differ in the degree of richness they possess, i.e. the amount of information they allow to be transmitted in a given time interval. The social dimension concerns the concepts of ‘self-presentation’ and ‘self-disclosure’: self-presentation states that in any type of social interaction, people have the desire to influence the impressions other people form of them, either to gain reward or to project personal identity; this is achieved through self-disclosure, the conscious or unconscious revelation of person-

⁶ Bruns (2008) uses the term “Generation Content”

⁷ For a concise definition and history, see boyd & Ellison (2007)

al information and a critical step in the development of relationships. Social media can thus be classified accordingly:

Table 2. Classification of Social Media (from Kaplan & Haenlein, 2010:62)

		social presence/media richness		
		low	medium	high
self-presentation/ self-disclosure	low	collaborative projects (e.g., <i>Wikipedia</i>)	content communities (e.g., <i>YouTube</i>)	virtual game worlds (e.g., <i>World of Warcraft</i>)
	high	blogs	social networking sites (e.g., <i>Facebook</i>)	virtual social worlds (e.g., <i>Second Life</i>)

2.3. Microblogging

Descendent from “away messages” in instant messaging (see Baron, 2008), microblogging is a relatively new form of social media. The most notable services include *Twitter*, *Tumblr*, *Cif2.net*, *Plurk*, *Jaiku* and *identi.ca*, while other social network sites such as *Facebook*, *MySpace*, *LinkedIn* and *Google+* also provide their own microblogging feature, known more commonly as ‘status updates’ (Wikipedia, 2013c). As the name suggests, microblogging is comparable to “traditional” blogging⁸; Herring et al.’s (2004:1) somewhat broad definition of a blog – that blogs are “frequently modified web pages in which dated entries are listed in reverse chronological order” – certainly encompasses the microblog, while both Miller & Shepherd (2004) and Kaplan & Haenlein (2010) also recognise the centrality of dated ‘posts’. However, in contrast to traditional blogging, microblogs encourage shorter posts of small elements of user-generated content, or “micro-content”, such as short textual units, individual images, or video links, which enable users to easily broadcast and share information about their activities, opinions and status at any moment via a range of Internet-based technologies such as mobile phones, instant message clients and the Web (Java et al., 2007; Krishnamurthy et al., 2008; Kaplan & Haenlein, 2011). The reduced requirements of users’ time and thought investment for content generation, allows frequent updates within a single day (Java et al., 2007); in contrast, the average interval between entries on traditional blogs has been estimated to be five days (Herring et al., 2004). Microblogging thus provides a faster, mobile, light-weight, and easy-to-use mode of communication. Using the same classification of social media as discussed above, Kaplan & Haenlein (2011) characterise microblogs as having a high degree of self-presentation/self-disclosure, and a medium-to-low degree of social presence/media richness, and place them between traditional blogs and social networking sites on the continuum of social media.

⁸ For genre analyses of blogging see, for example, Miller & Shepherd (2004), Herring, et al. (2004) and Myers (2010)

2.4. Twitter

2.4.1. Features & conventions

Twitter is indisputably the most popular microblogging application available. Users send textual messages – henceforth referred to as ‘tweets’ – limited to 140 characters⁹ to a web interface on which they are presented to a virtual audience. Figure 1 shows an example of a tweet sent by the official account attributed to the Dalai Lama¹⁰, as presented on Twitter’s own webpage:

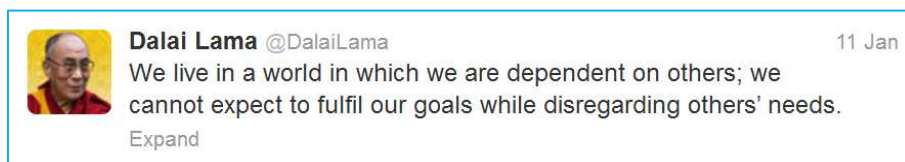


Figure 1. An example of a “tweet”

The visual appearance of a tweet differs depending on the channel used; a tweet is thus “a text with multiple expression plane realisations or, in other words, with no single stable visual or typographic form” (Zappavigna, 2011:792).

An important distinction between Twitter and social networking sites such as Facebook is its ‘directed friendship model’: Twitter accounts are typically open for users to ‘follow’, and in turn, each user has the potential to accumulate a group of ‘followers’ but there is no technical requirement, and usually no social expectation, for reciprocity (Marwick & boyd, 2010). Indeed, connections are often asymmetric: the account for the popular recording artist Katy Perry, for example, has in excess of 31 million followers, but in turn follows only 118 users¹¹. Participants employ heterogeneous strategies for deciding which accounts to follow: some follow hundreds or even thousands of diverse accounts, some follow only a few personal acquaintances, while others follow celebrities and strangers of interest (boyd et al., 2010). By default, tweets are made public, meaning they appear on individual users’ microblogs, and can be accessed via internal search functions, external web-based search engines and direct links. Thus, anyone, with or without a registered Twitter account, can access the public tweets. However, to control which users are granted access, users can make their account private, and have the option of sending private 140-character direct messages to a follower.

The central feature of Twitter, which users encounter upon logging in, is the Twitter ‘feed’, a stream of constantly updated tweets posted by those they follow listed in reverse chronological order. Figure 2 displays the author’s own Twitter feed:

⁹ The figure of 140 arose because the application was originally designed to utilise mobile phone technology, which features SMS text messages limited to 160 characters, with twenty characters reserved for usernames. Though the service has evolved beyond SMS technology to include 3rd party web and desktop clients, this limitation has persisted and has been re-narrated as a distinguishing feature (boyd et al., 2010).

¹⁰ <https://twitter.com/DalaiLama> [accessed: 14-01-2013]

¹¹ <https://twitter.com/katyperry> [accessed: 24-01-2013]

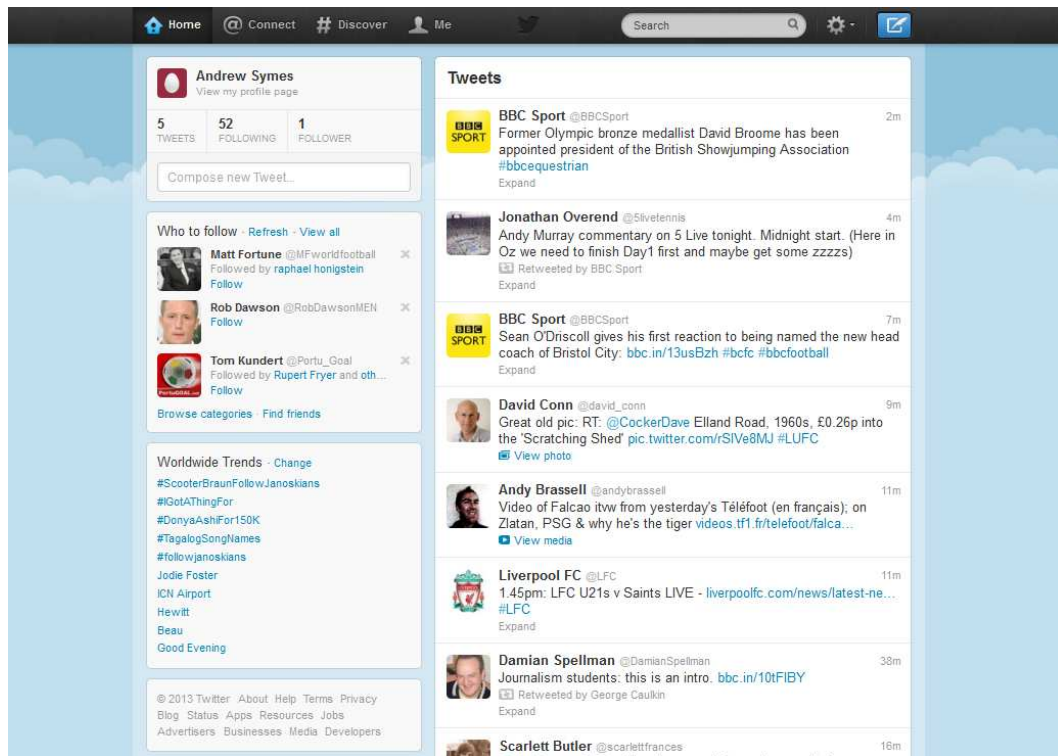


Figure 2. An example of a Twitter feed

In contrast to other forms of communication, there is no communal expectation that tweets be responded to or even acknowledged, as implied by the metaphor of ‘twittering’ continuously like a bird; nevertheless, a social need exists among Twitter users to engage with other voices (Zappavigna, 2011). To cope with the constraints of formulating messages confined to 140 characters, a series of conventions afforded by the technology available have been established in the Twitter community. Through the creative use of punctuation, users have developed strategies to reference and interact with others (see @ symbol below), to tag or label common topics (see *Hashtagging* below), and to propagate messages (see *Retweeting* below). Zappavigna (2011:790) suggests that “these expansions in typography meaning potential are part of a community-driven movement towards Twitter becoming a form of ‘public conversation’ [which is] multiparty, temporarily fluid and highly intertextual.”

2.4.1.1. @ SYMBOL

The first of these conventions, which stems from an older Internet Relay Chat practice (boyd et al., 2010), is the appropriation of the @ character to prefix a username in order to reference specific users:

- [1]. @user8: @addressee While I like the new facility their management of the media leaves so much to be desired. I'd give the host school a "F".
- [2]. @user4: Just watched LOL for the first time and it's now one of my fave films, absolutely love @mileycyrus¹²

¹² Miley Cyrus (@mileycyrus) is a well-known personality in contemporary popular culture, making it unnecessary to mask her identity.

In Twitter, it is multifunctional: as a form of addressivity, i.e. to direct messages to specific users (see [1]); and as an oblique reference to other users (see [2]) (Honeycut & Herring, 2009; boyd et al., 2010). In initial position, as with [1], the @ character typically indicates that the username which follows it is being addressed in the tweet, the structure functioning as a form of address; In medial or final positions, as with [2], its function is typically to draw attention to another user rather than explicitly directing an address (Zappavigna, 2011). Regardless of where the @user marker appears syntactically, the message will appear in the referenced user's 'mentions' feed.

2.4.1.2. RETWEETING

'Retweeting' is the process of republishing part or all of a tweet from another user on one's own Twitter feed, either in its original form or with modifications and/or added content. In doing so, tweet content is introduced to new audiences (Marwick & boyd, 2010). Structurally, it resembles email forwarding.

[3]. @user69: RT @originaluser: Freshman year of high school was the best.

Although the most common way of signifying a 'retweet' is by preceding the username with the character combination RT, retweeting strategies are varied and inconsistent, and retweets are rarely formatted as cleanly as [3], which may result in the text and meaning of messages changing: "there is no consistent syntax to indicate a retweet, attribution is inconsistent, the 140-character limitation and other factors prompt users to alter the original message, and adding commentary [either before or after the message] is prevalent," (boyd et al., 2010:2). Retweet processes are iterative; a retweet can contain several RTs and @s if the sender wants to credit several participants involved at different stages. They may, however, be altered to eschew any reference to the original source, casting doubt on origin and authorship.

2.4.1.3. HASHTAGGING

The 'hashtag' convention consists of prefixing a keyword/phrase with the symbol #. Although hashtags function in a variety of different ways, they are ostensibly used to mark the semantic topic of a tweet (see [4]) or to group tweets together by, for example, referencing an event or text-based meme (see [5]).

[4]. @user85: So happy that I GEDified the Motorola #Xoom last night now running #Android404 and awaiting #Jellybean

[5]. @user76: #20PeopleIThinkArePretty @addressee

Twitter's automated framework assigns hashtags a hyperlink directing users to search results for tweets using the same hashtag, enabling users to easily view and participate in on-going discourse. Example [4] features three unique tags which enter the message into three concurrent conversations based on the topic represented by respective the tags. This use of hashtags is a form of 'inline' metadata, i.e. "data about data" integrated into the linguistic structure of tweets (Zappavigna, 2011:791). Indeed, such metadata has become the defining feature of Web 2.0 (Pesce, 2006, cited in

Bruns, 2008). This practise parallels the use of tags as a strategy to categorise decentralised user-generated content in ‘produsage’ contexts in order to manage its diverse distribution; descriptive keywords are added to discourse to enable fellow users to negotiate it more easily (Bruns, 2008; boyd et al., 2010). Bruns (2008:172) likens this shared practise to “annotation at a distance”. In doing so, Twitter users enter into the social realm of collaborative tagging, or ‘folksonomy’. Derived from *taxonomy*, folksonomies are fluid structures of knowledge categorisation developed by the wider collaborative community of knowledge users (Bruns, 2008).

2.4.2. Popularity

As Kaplan & Haenlein (2011) point out, it is somewhat counterintuitive that an application limited to the exchange of predominantly text-based messages of 140 characters or less should prove so popular. Twitter’s rapid ascension to prominence, therefore, requires consideration.

From a pragmatic perspective, Twitter offers a cost-free, flexible, and easy-to-use means of disseminating information to a potentially substantial audience; it also utilises readily accessible technology, requires neither subscription fees nor the divulgence of private data, and the burden of time and thought investment on users is reduced compared to a medium such as blogging. Furthermore, while ostensibly a broadcast medium, Twitter offers dialogic potential and has the ability to facilitate conversation between proximally distant interlocutors (Marwick & boyd, 2011). Twitter has also become a key source of eye-witness accounts during newsworthy events, which often bypass the traditional gatekeepers of corporate news agencies, and has been cited as an influential medium through which social action can be instigated; for example, it has been credited with playing a crucial role for the Arab Spring (Kassim, 2012) and Occupy Wall Street activists. The journalism industry has also wholeheartedly embraced the medium, however, using it to report on unfolding stories such as courtroom developments and sports events. Furthermore, Twitter has emerged as both a key business channel, allowing companies to engage directly with customers and other parties, and as a critical channel to propagate media and ideas (O’Reilly & Milstein, 2012).

A significant part of the appeal of Twitter, however, is the role it plays in contemporary celebrity culture. Due to its open framework, it enables unparalleled access, whether genuine or artificial, to public figures and celebrities (Marwick & boyd, 2011); of the top twenty most followed Twitter users, sixteen can be considered celebrities¹³. As Marwick & boyd (2011) argue, Twitter fulfils a key role in the practise of celebrity (or “micro-celebrity”); through the appearance and performance of “backstage” access, particularly the supply of “in the know” information, first-person pictures, and opinionated statements, celebrity practitioners attempt to appeal to fan communities by creating a sense of intimacy between participant and follower, while visible interactions with others of similar status give the impression of candid, uncensored access to the people behind the personas. Alt-

¹³ <http://twitaholic.com/> [accessed: 24-01-2013].

though this access is not entirely new online – Miller & Shepherd (2004) point to the weakening boundary between the public and the private in their genre analysis of the blog – the scale and immediacy are historically unrivalled.

Twitter is thus an environment characterised by virtual – or mediated – exhibitionism and voyeurism (Kaplan & Haenlein, 2011). Central to exhibitionism is the social psychology of self-disclosure, which functions intrinsically to provide a heightened understanding of self through communicating with others and confirmation that personal beliefs fit with social norms, and extrinsically to turn personal information into a commodity and to manipulate the opinions of others through calculated revelations (Calvert, 2000). Mediated voyeurism concerns the consumption of revealing images of and information about others' apparently revealed and unguarded lives, often, yet not always, for the purpose of entertainment, through the mass media and the Internet (Calvert, 2000). The social forces that promote mediated voyeurism include the pursuit of truth or authenticity in an increasingly media-saturated world, the desire for vicarious experiences and excitement, and the need to be involved in the surrounding world, if only through observation (Miller & Shepherd, 2004). "Both voyeurism and exhibitionism have been morally neutralised, and are on their ways to becoming ordinary modes of being, [...] inscribed in our mediated discourse" (Miller & Shepherd, 2004).

Related to these concepts, are the notions of 'ambient awareness' or 'ambient intimacy' (Kaplan & Haenlein, 2011; O'Reilly & Milstein, 2012). Ambient awareness/intimacy describes a form of peripheral social awareness which is engendered by a relatively constant and lightweight, yet meaningful connection with one's social circle via social media; users experience near omnipresent knowledge, which may lead to increased effectiveness, stronger social relationships and improved well-being (Kaplan & Haenlein, 2011; O'Reilly & Milstein, 2012; Wikipedia, 2013d). Thus, while tweets and status updates may function in isolation, they often contribute to a larger body of discourse which may depict something very different. Despite temporal and proximal differences, posts can engender a strong feeling of closeness and intimacy; the ability to inform friends and family, or indeed the world, of current activities and feelings at a particular moment regardless of physical location is thus one of the key characteristics of Twitter (Kaplan & Haenlein, 2011).

3. Aims & methodology

3.1. Research aims

The primary motivation for this study is to conduct a broad analysis of the pertinent linguistically-related, empirically-observable discursive phenomena emanating from current Twitter practices. The results will serve as the basis for the pursuit of more detailed future research with Twitter as the principal medium in focus. As such this study adopts an inductive approach.

The study is split into 5 research “modules”:

1. Basic usage
2. Communicative functions
3. @ symbol
4. Retweeting
5. Hashtaging

The first module examines the basic usage of tweets and seeks to establish how much and how often users tweet. The second seeks to categorise and thus provide a macro-level overview of the different communicative purposes that tweets serve. The final three modules each concern one of the three discursive conventions discussed in Section 2.3. Each of these modules is guided by two principal questions: how prevalent are these conventions and what purposes do they serve? Module 3 also examines the prevalence and characteristics of ‘interactions’, while modules 4 and 5 analyse the structure of their respective phenomena. Further details regarding the specifics of the various analyses conducted, where necessary, are accounted for as part of the respective *Findings & discussion* sections which follow.

3.2. Ethical considerations

The ease with which the Internet facilitates social research has led to prominent debate over the ethics of online research (see Hine, 2005), particularly regarding covert non-participant observation methods (see Sanders, 2005). Nevertheless, Twitter is an unquestionably public platform and upon subscribing, users must agree to terms of service (Twitter, 2013) which make this abundantly clear. As messages analysed were taken from public users only, and from accounts that are free-to-view rather private, it was considered ethically sound to pursue such a line of enquiry, on the basis that user identities, links and any other sensitive information would not be published. Usernames and links have therefore been replaced by alternative text; only users who overtly use Twitter to reach a public audience, such as celebrities and journalists, were exempt from this practice.

3.3. Data

Tweets were collected using a free-to-use online script developed by Martin Hawksey, which runs via a Google Spreadsheet¹⁴. Users are simply required to specify a number of parameters including the search terms, period and number of desired results, and the data collection runs automatically,

¹⁴ <http://mashe.hawksey.info/2012/01/twitter-archive-tagsv3/> [Still available as of 25-01-2013]

extracting detailed data downloadable in spreadsheet format. For the purposes of this piece of research, the search term used was “from:@username OR to:@username”, where the sequence @username was replaced by a genuine user name for each data query, which facilitated the collection of messages both sent and received by respective users.

Tweets analysed were sent during the 48-hour period 00:00 Sunday 22th July to Monday 23:59 23rd July 2012. This period was chosen to include both a traditional weekend day and a working day with the aim of preventing any potential skew in the data. The 100 Twitter users who comprise the data set were identified randomly by using the public Twitter timeline¹⁵, and their microblogs given a preliminary scan. Only accounts belonging to members of the general public were selected; accounts belonging to, for example, companies, media groups, and celebrities were purposefully ignored. This study therefore considers only a particular user profile, and does not offer a holistic view of Twitter. A further prerequisite for inclusion was the apparent use of English as the primary language of communication on their Twitter feed; no attempts were made to choose only native speakers, and tweets containing foreign languages were included within the data. These were excluded from content analyses, but were included in the generic quantitative analyses. As Twitter’s light-weight framework does not oblige users to provide demographic information upon registering, such considerations played no part when selecting potential participants.

The resulting corpus contains a total of 11,187 tweets

3.4. Analytical approach

3.4.1. Computer-mediated discourse analysis

This paper adopts the ‘Computer-Mediated Discourse Analysis’ (CMDA) approach to researching online interactive behaviour, developed by Susan Herring. It adapts methods from language-focused disciplines such as linguistics, communication and rhetoric for the analysis of computer-mediated communication (Herring, 2004). Herring’s approach is summarised briefly below.

The essential objectives of discourse analysis are to: first, identify demonstrable discursive patterns which may not be immediately obvious to observers or participants; second, provide insight into both linguistic and non-linguistic speaker choices, as conditioned by cognitive and social factors; and third, investigate whether, and to what extent, new media technologies shape the communication that takes place through them. Five discourse analysis paradigms commonly employed in CMDA research are text analysis, conversation analysis, pragmatics, interactional sociolinguistics and critical discourse analysis. However, rather than any single theory or method, the CMDA approach provides “a methodological toolkit and a set of theoretical lenses through which to make observations and interpret the results of empirical analysis” (Herring, 2004:4). Furthermore, most

¹⁵ https://twitter.com/public_timeline (now offline); <http://twitspy.com/> fulfils a similar function [accessed: 25-01-03]

CMDA research does not take as its point of departure a paradigm, but observations about online discourse, making it an inductive rather than deductive, or theory-driven approach.

The CMDA approach is modelled on five domains of language, organised in a hierarchy from micro to macro-linguistic phenomena: 1) structure, including typographical and orthographical issues, morphology, and syntax; 2) meaning, meanings of words, speech acts and macrosegments; 3) interaction management, including turn-taking, topic development, and coherence; 4) social phenomena, including expressions of play, conflict, power, and group membership; 5) participation patterns, as measured by frequency and length of posted messages. The work conducted as part of the present study pertains mainly to domains 1, 4 and 5.

3.4.2. Content analysis

This study employs the “counting and coding” paradigm of classical content analysis, the basic methodological apparatus of CMDA (Herring, 2004). Used to make objectified inferences from a focal text to its social context, this hybrid method bridges statistical formalism and the qualitative analysis of the materials by considering the “kinds’, ‘qualities’ and ‘distinctions’ in the text before any quantification takes place” (Bauer, 2000:132). Here, content analysis is used quantitatively to establish an overview of the principal communicative function of tweets (Section 4.2), and qualitatively to classify the most prevalent trends of retweeting (4.4) and hashtagging (4.5) practises.

Making definitive judgments about the communicative intent of language is notoriously difficult. For example, Austin’s speech act theory maintains that utterances perform three simultaneous acts: ‘locutionary’, the basic act of speaking; ‘illocutionary’, the speaker’s intention; and ‘perlocutionary’, the ultimate effect on the addressee (Huang, 2007). Language acts are therefore multifaceted. Being disconnected from the context in which these tweets were exchanged means that the content analysis was susceptible to an inherent degree of subjectivity; tweets were coded according to what was considered the most likely semantic interpretation from an array of possibilities.

While each hashtag was analysed individually, the single-code analysis employed in investigating communicative functions and retweets did not take into account the likely plurality of content meaning, and considered tweets as singular communicative acts despite them containing multiple sentences. Consequently, results should be treated with a degree of caution. Furthermore, the coding categories in all three content analyses offer only broad overviews of the respective phenomena. Nevertheless, this analysis can be considered a type of pilot study whose goal is a better generic understanding of Twitter discourse, thus paving the way for more detailed research in the future.

3.4.3. Software

Excel was the programme used most extensively to analyse, code and count the data; the filtering, formula, and conditional formatting tools were particularly utilised. The concordance software AntConc was also used to identify certain frequently occurring words, word clusters and patterns, which were then copied into Excel for analysis.

4. Findings & discussion

In the following sections, the results of the study are presented, discussed and evaluated in the context of selected research papers on microblogging and Twitter, as well as instant messaging, text messaging, and blogging, three genres of NMC which ostensibly appear closest to microblogging.

4.1. Basic usage

The statistics in Table 3 provide an overview of tweeting practices of the 100 users analysed. A more detailed breakdown of individual user activity can be found in appendix A, and graphs charting tweets sent contra tweets received can be found in appendix B.

Table 3. Basic user statistics

	<i>frequency</i>	<i>range</i>	<i>mean</i>	<i>median</i>
<i>total tweets</i>	11,187	1 – 765	111.9	59.5
<i>outbound¹⁶ tweets</i>	8,965	1 – 609	89.7	47,0
<i>inbound tweets</i>	2,222	0-297	22.2	8
<i>active period* (hh:mm:ss)</i>	-	01:05:29 - 47:55:01	37:01:43	40:57:36
<i>average time lag between tweets sent*</i>	-	00:03:43 - 16:38:36	01:22:34	00:43:46

* Concerning only tweets sent for users who sent a minimum of 2 tweets.

Tweets are posted with variable frequency. At the lower end of the scale, 11 users either received or sent a combined total of less than 10 tweets over the 48 hour period. In contrast, a similar number (12) posted in excess of 200 tweets, with 4 users posting in excess of 400; on average these 4 users sent a message every 5 minutes and 3 seconds. However, this group of extremely “prolific” tweeters skew the data somewhat, as demonstrated by the disparate mean and median values for each of the variables; on average, a tweet is posted every 24 minutes and 47 seconds¹⁷. Both the mean and median values for the active period suggest that users typically contribute to Twitter for sustained periods, and, as supported by average and medium lag time values of approximately an hour, post regularly within the time frame.

Due to methodological differences, providing a robust comparison of the data extrapolated here with the findings of other NMC studies was ultimately unachievable. While such data may well exist, given the scope and research aims of the current study, sourcing it was considered a low priority and not pursued.

4.2. Communicative function of tweets

The main content theme of each outbound tweet (inbound tweets were excluded to prevent a data skew) was identified and coded to give a macro-level overview of the communicative function of tweets. Due to time constraints, a cap of 200 tweets per user was introduced, providing a total cor-

¹⁶ ‘Outbound’ tweets are those sent by the 100 users under analysis; ‘inbound’ tweets are those received.

¹⁷ Calculated by dividing the mean value for tweets sent by the mean value for active period; using the entire 48-hour period produces a value of 32 minutes and 7 seconds.

pus of 7441 tweets. The coding schemata was developed using, as a point of departure, similar studies conducted by Honeycutt & Herring (2009) and Lee (2011), which were supplemented by a ‘grounded theory’ approach (see Oktay, 2012) to encompass emergent trends. Descriptions and examples of the dominant communicative functions identified¹⁸ are provided below. Functions are listed in order of their prevalence:

- **Retweet** (25.%): use of the RT convention (see section 2.4.1.2)
 - [6]. @user40: RT @original_sender: You do not truly love a band or musician, until you're willing to blow all your savings to see them live or meet them.
 - [7]. @user83: no . I'll say Ricky Tan n Rush Hour 2 “@original_sender: The saddest death in a movie by far is when G-BABY died in HARD BALL!”
- **Twitter interaction** (24.9%): messages directed at fellow Twitter users (via the @ symbol)
 - [8]. @user12: @addressee are you going to the beach?
 - [9]. @user53: @addressee Wow, sounds interesting :D now I can't sleep :P tell me more about it!
- **Self-experience** (10.8%): comments concerning the user’s own self, besides those deemed to represent “current state”
 - [10]. @user26: had fun during practice today finally playing well again #gv
 - [11]. @user37: Every time I use my phone while I'm in the bed, I drop it on my face
- **Opinion & judgement** (10.7%): subjective or evaluative comments (regarding topics other than the user)
 - [12]. @user04: Genuinely think my niece will grow up to be a comedian, she's hilarious for being only two years old
 - [13]. @user62: Workout shorts are heaven
- **Current state** (8.5%): comments pertaining to the user’s current, or extremely recent, activity, state or mood
 - [14]. @user41: I have no energy
 - [15]. @user75: Chilling with my Bestie talking about some of everything!
- **Link** (3.7%): links to external Internet content
 - [16]. @user31: Fresh Mozzarella Pasta Casserole for #SundaySupper <http://t.co/xxxxxx> via @original_sender
 - [17]. @user59: Photo: <http://t.co/xxxxxx>
- **Fabricated text** (3.2%): song lyrics, famous sayings, quotes, etc.
 - [18]. @user71: The superior man wishes to be slow in his speech and earnest in his conduct.
 - [19]. @user93: abc EASY AS 123
- **Others’ experience** (2.1%): non-subjective or evaluative comments about others
 - [20]. @user41: My mum never even bothers to check if I'm alive ever, she just texts me from time to time asking if I want food hahah god sake
 - [21]. @user91: Tasha just said "I don't know how to open up this fancy popcorn" LMAO
- **External interaction** (2.0%): messages directed at an specific but unstated recipient, and general greetings

¹⁸ 239 tweets either did not fall into one of the categories or could not be accurately analysed due to obscurity of meaning.

[22]. @user42: Good Morning Everyone !!!!

[23]. @user22: When you told me I was beautiful I actually felt beautiful.

- **Metacommentary** (1.5%): comments about Twitter or using Twitter

[24]. @user76: When people tweet for 5rts.. I tweet back, my tweet forever gets ignored-.-

[25]. @user23: Taniyah Just told me I got tweet watcher

- **Humour & play** (1.4%): messages with no other obvious intent than to amuse readers

[26]. @user31: #greattobeaguy You can lean down to pick something up without having to worry about your shirt hanging open.

[27]. @user52: I'm going to call you "Monday" because no one likes you! #InsultOfTheDay

- **Public commentary** (1.1%): reports on public events, including weather updates

[28]. @user60: Its raining hard over here in vegas with lightning and its 95 degrees!

[29]. @user76: Tomorrow - 2 year anniversary of the formation of a band that changed millions of girls' lives and brought those girls together as family <3

- **Exhortation** (1.0%): messages which direct or encourage others to act

[30]. @user25: someone let me use their pool #please

[31]. @user12: someone do something with me! :)

- **Initiate interaction** (0.8%): messages directed at a general audience which seek a response

[32]. @user74: how do you cure a blocked nose :(

[33]. @user06: Time to book and plan vacation. What to do? Where to go? #procrastinator

Tweets have evolved far beyond providing a response to the original prompt of “What’s happening?”; they now serve a wide range of communicative functions, far wider than the macro-level coding schemata used here suggests. People now use Twitter to engage in dialogue, develop social relationships, exchange ideas, partake in debates, instigate business, and more. While Krishna-murthy et al. (2008) identify three groups of user, – those who broadcast tweets (‘broadcasters’), those who exhibit reciprocity in their relationships (‘acquaintances’), and those who follow many more users than they have followers (‘miscreants’) – following Twitter’s exponential growth since such early studies, it is apparent that the medium fulfils users’ individual needs or goals. Thus, contrary to popular misconceptions, Twitter users do not constitute a homogenous mass

Nonetheless, Twitter is used more extensively for certain purposes than others; the two largest categories (*retweets* and *Twitter interaction*) combined constitute over half (50.1%) of all tweets sent. They are characteristically similar as they both directly contribute to the collective Twitter discourse, either by interacting with fellow users or replicating their content. These two categories are further supplemented by the smaller categories of *exhortation* (1.0%), *initiate interaction* (0.8%), *general interaction* (2.0%), and *metacommentary* (1.5%), as they all explicitly seek to instigate or comment on Twitter discourse, thereby contributing to a highly interactive environment. This confirms the fallacy of the view that microblogging is principally monologic, and, as Zappavigna points out, (2011:803) “criticism of Twitter as a service facilitating inane and frequent status updates about users’ activities seems to have missed the social point of twittering.”

A second common use which unites several sub-categories of communicative function (*self-experience, opinion and judgement* and *current state*, together forming 30.0% of the sample) concerns the activities and sensibilities of individual users, i.e. what Java et al. (2007) label as *daily chatter*. Tweets are often acutely intimate, revealing and at times sexual in nature, which suggests that identity performance is a principal motivation for engaging in Twitter discourse, as it is in many other online environments. Such high degrees of self-presentation and self-disclosure contribute to the categorisation of Twitter, much like blogs, as a perfect environment for virtual exhibitionism, and, predicated on the assumption that performance requires an audience, voyeurism (Kaplan & Haenlein, 2010 & 2011; see Section 2.3.2). Furthermore, such tweets contribute to the “ambient intimacy” of social media, i.e. that being peripherally aware of fellow users’ activities and well-being can help engender strong feelings of closeness and intimacy. Therefore, what might be considered inane chatter serves an important social function.

Although methodological differences prevent direct comparison, daily chatter is also a consistent theme identified in other studies of social media. Honeycutt & Herring (2009), for example, report that tweets reporting users’ own experiences comprise the most common function, while Lee (2011:118) observes that “communicating mundane and day-to-day topics [seems] to be a persistent function of short new media messaging [i.e. microblogging, texting and away messages].”

Another category which contributes to the performance of identity is *humour and play*, but only 1.4% of tweets were defined as such, and the category would thus appear to misrepresent the Internet at large. However, many instances of playfulness can be found integrated within tweets categorised elsewhere, in particular in the form of hashtags (see also the emoticons used in [31] and [32]); thus these results should not be interpreted as suggesting that Twitter is a humour-free domain. Indeed, like many other new media contexts, playfulness is a core activity (Lee, 2011).

Of less importance appears to be the “offline” world; although Java et al. (2007) identify *reporting news* as one of only four main “user intentions”, only 1.1% of tweets were devoted to *public commentary* and 2.1% to *others’ experience*, although these categories will be represented amongst other sub-categories. Furthermore, one must take into account that media organisations are excluded from the current analysis, and as a major presence on Twitter, Java et al.’s observations are likely to be entirely valid if this study had adopted a more holistic approach.

4.3. @ symbol

An @ symbol – irrespective of function – was identified in a total of 6985 tweets (62.4%), at an average of 0.7 per tweet. The @ symbol was used 8229 times in total. However, given that all inbound tweets must necessarily contain an @ symbol to be included in the data sample, such tweets were excluded to examine microblogging literary practice fairly. Of outbound tweets only, 4830 (53.9%) contain an @ symbol, within which 5947 instances were identified, giving an average con-

sistent with the wider data sample of 0.7 instances per tweet. Therefore, just as it has become symbolic of wider online discourse, the @ symbol constitutes one of the defining features of Twitter.

4.3.1. FUNCTION

The @ symbol has played an integral role in the evolution of Twitter into a highly interactive environment. Twitter is a “noisy” environment due to the large volume of tweets and the speed with which they are posted, leading to a high degree of disrupted turn adjacency when users “converse”, much more than in a typical chatroom or discussion forum (Honeycutt & Herring, 2009). The @ symbol is therefore a useful strategy for relating one tweet to another and for making coherent exchanges possible.

Each instance of the @ symbol was examined individually in an attempt to categorise and quantify its principal function, as summarised in appendix C. The vast majority of the @ symbols fulfilled three main duties: to direct a tweet towards a particular user (49.4%) (e.g. [1]); to indicate the original author of a retweet (36.6%) (e.g. [3]); and to reference a user within the body of a tweet, with no explicit expectation of a response (10.5%) (e.g. [2]). Within the grouping *other* (0.6%), uses include substitutions for the preposition *at*, both in locative and temporal senses, forming part of an email or user name on another platform, and meta-references to the practice of using the @ symbol (e.g. “*I think you @'d the wrong person*”). None of these uses were sufficient in number to warrant a separate grouping, however.

Honeycutt & Herring (2009) identified 91.0% of the @ symbols in their data sample as instances of addressivity, and only 5.4% as references; how they dealt with retweet authorship is unclear. Nevertheless, the comparatively infrequent utilisation of the @ symbol beyond these three major uses in both studies suggests that users are aware of the distinct role it now plays in Twitter discourse, and use it with discretion to avoid ambiguity.

4.3.2. INTERACTIONS

The present investigation identified each of the ‘interactions’ the 100 users engaged in, and quantify the number of tweets that comprise them. An interaction was considered to be instigated when a minimum of two users each employed the @username of their counterpart(s) at any point during the period sampled. An interaction can therefore feature multiple conversations of different durations and semantic content. The prevalence of interactions is summarised in Table 4.

Table 4. Summary of the prevalence of interactions

	<i>total</i>	<i>range</i>	<i>per user</i>		
			<i>mean</i>	<i>median</i>	<i>mode</i>
<i>number of interactions</i>	565	0 - 37	6.5	4.0	1 (x20)
<i>number of tweets</i>	4681	2 - 181	8.3	4	2 (x118)
<i>interaction duration (hh:mm:ss)</i>	-	00:00:16 - 47:32:39	07:07:30	00:42:33	-
<i>average time lag between tweets</i>	-	00:00:16 - 21:15:14	01:18:34	00:10:05	-

Unlike NMC platforms such as instant messaging, email and contemporary text messaging, the Twitter architecture does not provide an explicit “window” in which interactions can automatically occur without interruption; although an @ symbol referencing a user directs messages to a specific page, replies may appear elsewhere. Nevertheless, this does not prevent Twitter users from frequently interacting and engaging in conversations using both the @ symbol and the retweets (see 4.4) to compose dialogic lines of communication; as Herring (2010) points out, communicators’ access to a persistent textual record enables an efficient strategy of discourse processing.

The characteristics of Twitter interactions and the ways in which users engage in them are, however, highly variable. Some users interact frequently with fellow users, while many others (20) engage with others only once; some users respond to messages rapidly, suggesting that they constantly monitor their Twitter notifications, while others take much longer to respond, which would suggest that for them Twitter is perhaps less critical. The shortest interaction lasted a mere 16 seconds, while the longest stretched across almost the entire time period sampled (47h:32m:39s). Interactions are often short and dyadic – 118 of the 565 interactions (20.9%) lasted only a solitary response, and can hardly be considered as conversations – yet may be lengthy and occur concurrently with other Twitter activities. The most extensive exchange of 181 tweets occurred over a period of 38h:23m:08s with an average lag time between tweets of 12m:48s.

Given that Twitter’s open network both permits any user to freely address any other, and affords users the luxury of being selective about which messages warrant response without significant adverse ramifications on social relationships, the extent to which these interactions are reciprocated was investigated. Out of 2222 inbound tweets, 2029 (91.3%) formed part of an interaction: 8.7% thus seemingly went unacknowledged¹⁹. This appears somewhat at odds with claims that responsiveness on Twitter is variable (Marwick & boyd, 2011). However, this number would probably be much higher if a broader spread of Twitter accounts were analysed to include celebrities, journalists, media accounts and other popular users as for these users the volume of inbound messages becomes difficult to manage.

These results contrast with those of Baron (2010), who finds instant messaging (IM) *conversations* on average to span 93 “transmission units” across duration of only 24 minutes, making IM, predominately, a near synchronous technology. Twitter *interactions*²⁰, measured over the entire 48-hour period (mean duration = 07h:07m:30s), average only 8.3 tweets per interaction, more than 10 times as short; the lag between tweets in an interaction on averages over 1 hour and 18 minutes, making Twitter an asynchronous medium. Indeed, the number of exchanges taken just to close IM conversations averages 7 (Baron, 2010). However, the average number of words in tweets²¹ is almost double (10.0) than that of IM (5.4). Hence, although Twitter interactions are shorter and

¹⁹ No data is available detailing responses to individual messages

²⁰ Demarcating Twitter ‘conversations’ from ‘interactions’ and analysing them separately would widen this disparity

²¹ Established as part of an abandoned syntactic analysis of a sub-corpus of 1711 tweets

spread over a longer period of time than those of IM, each tweet is longer, with users taking the extra time to compose and post more substantial messages. This is likely to derive from a combination of the technological affordances, or limitations, attributed to each platform – IM interactions invariably occur via computers with a keyboard, a much more expedient input device than the cumbersome keypad of a mobile phone, as used by many Twitter users, – the different intended functions of the two media – IM is a tool designed specifically for messaging, while Twitter is used more diversely and the medium less immediate, or “intrusive”, – and the awareness that tweeting excessively may be considered “bad practice” as it clogs followers’ feeds.

Furthermore, this regular use of Twitter to engage in direct interactions with fellow users apparently distances the practice of microblogging from regular blogging; while blogs exhibit some similarities with conversations, such as the use of discourse markers (Myers, 2010), and are frequently characterised as socially interactive and community-like in nature (Herring et al., 2004), blogs demonstrate little of the conversational potential often claimed for them: “communication in weblogs may entail an exchange of messages between addresser and addressee, but no exchange of messages is sufficient to constitute weblogs as conversation” (Peterson, 2011:15). Baron (2008) also points to blogs being used *instead* of personal conversation, and suggests that this may be due to them being an unobtrusive ‘pull’ technology rather than a ‘push’ technology such as Twitter, which “shows up uninvited on your electronic doorstep” (2008:113). Furthermore, while bloggers are empowered to control the “volume” of interpersonal communication, i.e. to decide which messages warrant response (Baron, 2008), a greater awareness and indeed desire appears to exist among Twitter users that posts may be responded to; arousing a response or validation through a retweet appears to be a primary motivation for some users, which if achieved constitutes, for some users, a sort of “badge of honour”.

4.4. Retweeting

Before considering the results below, an important caveat must first be addressed: verifying that retweets are *bona fide* examples proved impossible using the current methodology. Retweets were identified and coded where explicit conventions were employed by users, for example, by preceding the message with the acronym RT or by enclosing a message in quotation marks. Users are, however, free to amend a retweet so that it appears as an original message. The analysis, therefore, was conducted having put complete faith in users’ online behaviour being ethical.

Analysed in this section were the tweets sent by the 100 users only; at least a single retweet was posted by 87 users, and a total of 2259 posts were identified as retweets, representing a significant proportion of the 8965 outbound tweets (25.2%), and confirming the centrality of retweeting practices to Twitter discourse.

4.4.1. STRUCTURE

The majority of retweets (1861, or 82.4%) – referred to here as ‘direct’ retweets – are posted verbatim, i.e. modifications are restricted to the addition of the @username of the original author and retweet markers. This behaviour is unlike that found in other new media environments; responding to an email or forum post, for example, may include quoting original content but rarely occurs *without* any new content (Crystal, 2011). Retweets with added content, or ‘modified’ retweets, numbered only 348 (15.4%). Added content is typically short, unsurprising given the 140-character constraint, and responds to or comments on the content of the message being retweeted:

- [34]. @user39: Lmao RT @original_author: Daughter up eatin onion & garlic toasted ritz chips. Breath smelling like death & vampire protection.
- [35]. @user53: Have some ice cold beer? RT @original_author: Does anybody here know what to do if a bear attacks?

Of the 50 remaining retweets, 47 were classified as ‘via retweets’, which are tweets which accredit their source using the marker ‘via’, or ‘v’ when abbreviated:

- [36]. @user38: "Guns don't kill people - Americans kill people." Michael Moore responds to the Aurora shootings: via @NewStatesman

Whether these should be classified as retweets is, however, debatable as the ‘via’ marker may be used only to credit the source of information despite the form of the retweet being entirely different. Nevertheless, the conventions mirrors the functionality of retweeting and these examples were therefore retained.

The strategies for indicating a retweet form a relatively narrow group. By far the most common strategy in the data (featuring in 1817, or 80.4%, of retweets) is to precede the original message and its author’s user name with the abbreviation RT (short for retweet). The @username is then usually followed by a colon to distinguish the author of a retweet from a potential addressee, as exemplified by [34] and [35]. Indeed, this is the default strategy employed by some of the Twitter software platforms available when producing retweets. However, the results garnered by the data extraction process are not entirely reflective of the current online reality; while users were previously forced to copy messages manually, the Twitter homepage now enables users to publish a retweet with a single mouse click and then truly presents a retweet “verbatim”, complete with the original author’s avatar and username, with the name of the re-poster given only at the bottom of the tweet:



Figure 3. An example of a ‘direct’ retweet as represented on Twitter’s homepage

Consequently, the semiotic and linguistic output of a direct retweet is dependent on the software being used. While this may initially seem trivial, it may significantly influence the interpretation of a message; a retweet which simply references the @username of a user with substantial social capital is unlikely to have the same impact as one which appears to come directly from the source.

Nonetheless, this issue applies only to direct retweets; any other strategies or modifications employed in constructing a retweet require manual input, and result in it being presented as though it were the user's own. This extra burden may indeed explain the substantial gap between direct and modified retweets. Further retweeting strategies observed include the alternative abbreviations MT (meaning modified tweet) and QT (meaning quoted tweet), the placing of the original message within a pair of punctuation marks, such as " ", " ", and « », and the marker 'via' (see example [36]). Where additional text is added to the retweet, it can either precede or follow the original message: in cases where the abbreviations demarcate the original messages, additional commentary typically precedes them; commentary in conjunction with punctuation pairs is more variable as the visual impact of the symbols makes comments in either position equally distinguishable. No retweets were found to feature commentary on both sides of an original comment simultaneously.

It can be further noted that users display an overwhelming degree of uniformity with regards to which retweeting strategy they employ, tending not to vary between the use of abbreviations or punctuation pairs. This may be a result of the constrictions placed upon their tweets by the software they favour, or it may be a more conscious effort to construct a consistent online identity.

4.4.2. FUNCTION

Considering users' reasons for retweeting necessitates interacting with the users themselves, and thus, without any informant testing or questionnaires, it is not possible within the scope of the present work to evaluate such motivations. That being said, boyd et al.'s (2010) research into conversational aspects of retweeting did extract such views from Twitter users, and provides a non-exhaustive list of retweet motivations. Using this list and their subsequent discussion as a guide, a number of trends which emanate from the data were identified:

- **To amplify or spread** tweets to new audiences (in particular humour)

[37]. @user10: RT @original user: Twilight won multiple awards at the Teen Choice Awards last night. In related news, teens still have awful taste in movies.

- **To spread links** to content of general interest (e.g. news stories and articles)

[38]. @user38: "@original_author: Ethnic Cleansing of the Rohingya in Burma continues (videos, coverage) <http://t.co/6lkcUCSD>"

- **To spread information** relevant to a user's interests, and potentially like-minded followers

[39]. @user14: RT @justinbieber: tomorrow. TEEN CHOICE AWARDS on FOX. watch. #AsLongAsYOULoveMe

- **To comment** on a tweet by adding new content, often to initiate a new conversation

[40]. @user05: “@original_author: I feel like I should be doing something” // come London

- **To converse** in the public domain

[41]. @user77: “@original_author: @user77 haha yes! But whatsupppp how are youuu. Did you move yet?” I've been good wbu and no not yet so ucc it is

- **To publicly agree** with someone or offer approval

[42]. @user78: RT @original_author: Colorado guys know how to treat a girl.

- **To highlight contrastive sensibilities**, and to highlight abuse

[43]. @user89: Eww “@original_author: I love when guys speak to me in Spanish ☹”

- **To spread content as an act of friendship, or loyalty**, by drawing attention

[44]. @user76: RT @Harry_Styles: Please vote for One Direction!! :D

- **To publicly appreciate another user’s attention**

[45]. @user83: Aww thank you sooo much! ☺ I love yours too! ♥ RT @original_author: @user83 I love your display icon ?

- **For self-gain**, either to gain followers or reciprocity from more visible participants

[46]. @user88: RT @original_author: i'm going to try to follow all you back! retweet if you want one!

- **To encourage social action** (e.g. sign a petition or vote for a candidate)

[47]. @user04: RT @ original_author: Spread the word people!! <http://t.co/xxxxxx>

While retweeting is ostensibly a simple act of copying and rebroadcasting, a wide range of diffuse functions have emerged; those listed above are only a selection of the most common. Many such functions supplement the @ symbol and make retweeting central to interactional practices (retweets formed 14.4% of the exchanges discussed in 4.3); “regardless of why users embrace retweeting, through broadcasting messages, they become part of a broader conversation” (boyd et al. 2010:10). As opposed to a directed message via the @ symbol, which although public occurs within a bounded group²², a retweet is published on a user’s feed and thus relays its content to a non-participatory but interconnected audience; despite making no active contribution, through ambient awareness (see Section 2.3.2) others may feel part of a conversation, particularly when such conversations become conspicuous by their magnitude. Thus, as boyd et al. (2010:1) point out, “the practice contributes to a conversational ecology in which conversations are composed of a public interplay of voices that give rise to an emotional sense of shared conversational context.”

This intertextual and heteroglossic interplay of voices leads to substantial ambiguity about authorship, ownership, attribution and conversational fidelity (boyd et al. 2010), particularly when

²² Directed messages, i.e. featuring an @user address, between two users both “followed” by a third-party will show up in the latter’s “stream”

messages are replicated verbatim. Although the syntactic content of a retweet may remain unchanged, the information and possibly semantic contents may be altered. Pronouns, for example, are a particular source of ambiguity:

[48]. @user81: RT @original_author: I actually wish I hadd summer skool rather then be bored at home

Does [48] mean only to relay the original author's wish to attend summer school, or does @user81 endorse the sentiments of the message and also wish to attend summer school? A retweet posted verbatim indicating disagreement, meanwhile, communicates this disposition without any explicit linguistic content; to interpret such retweets correctly, background information regarding the posting user is required or such a "silent" disagreement can be easily misconstrued as endorsement, or indeed *vice versa*. These problems are exacerbated in the case of embedded retweets, of which 137 instances were found in the data set, particularly when punctuation or author attribution is inconsistent. Interpreting retweets thus represents a significant challenge for audiences.

Retweeting also raises issues about the reception of content; users may be unwittingly exposed to content of which they disapprove, for example. Unlike other types of Internet content, where users can choose what to view, Twitter's classification as a 'push' technology allows content to intrude on users' virtual personal space. Furthermore, retweeting a message might expose it to an audience unintended or undesired by the original author; users must give careful consideration before posting, particularly those with a large number of followers, or those posting to such users, as unsavoury messages will quickly reach a large audience.

A particular trend of retweeting practices can be collectively considered "ego tweets" (boyd et al., 2010), i.e. users refer to themselves within a retweet (see also [45]).

[49]. @user74: RT @original_author: "@user74: HOW CAN I BE SO ILL IN THIS BEAUTIFUL WEATHER :("<<

In [49], @user74's retweet features an embedded retweet of his/her original message; neither of the two turns add explicit communicative content. 393 ego tweets were identified, with the @username of the posting user featuring as the author (48 instances; see [49]), addressee (303; see [41] and [45]) or referent (42) of a retweet. This seemingly narcissistic behaviour would appear to validate the assertions of blogging and microblogging as exhibitionistic acts (see 2.3.2). Ego retweets may, however, be seen as giving credit to fellow users' tweets, a seemingly integral part of becoming a popular and widely followed Twitter user. Indeed, a recurring observation from the retweeting data suggests that it is a vital interpersonal strategy for enhancing social relations; by reposting to their own group of followers, users often offer their approval of the content of the original tweet, resulting in stronger social bonds with the original poster and an increased likelihood of the "favour" being reciprocated. Furthermore, as previously pointed out, arousing a response or validation through a retweet appears at times to be a highly prized achievement.

4.5. Hashtagging

A total of 703 (7.8%) outbound tweets were identified as containing one or several hashtags. At first glance, this may seem a somewhat surprisingly low figure considering the purported centrality of hashtagging as a Twitter practice. On consideration, however, given that the majority of tweets concern, broadly speaking, interaction and daily chatter, it is perhaps no surprise as these types of message may not be expected to require organisational metadata (see Section 2.3.1). Identified were 908 individual hashtags, of which 615 (67.7%) are unique. Multiple hashtags were found in 124 tweets (17.9% of tweets with hashtags). A list of the most popular hashtags can be found in appendix D. Of the 100 accounts under scrutiny, 79 were found to feature at least one hashtag, suggesting that the practice is widespread if not used extensively as anticipated. The highest number of hashtags employed by any one single user was 98 (@user11), while 8 was the highest number of hashtags in a single tweet.

4.5.1. STRUCTURE

Although hashtags were also found in initial (16.8%) and medial (15.1%) syntactic positions, the overwhelming majority (66.4%) were found in the final position. When they occurred initially, they often introduced a “micro-meme” (see below). A small percentage of tweets (1.6%) consist solely of hashtags and thus lack a syntactic position.

Many tags are comprised of multiple words, and even whole sentences. The mean character length was 10.7, the median 10 and the mode 9. The longest tag in the data set featured 64 characters: *#iknowwhatyoumeanbutiwillpretendtoactinnocentandtypeshitlikethis*. They are thus often formed using substantial grammatical structures rather than being restricted to single lexical items. Furthermore, tags often function in tandem, with the whole being greater than the sum of its parts:

[50]. @user23: I'm tired of her, #that's #What #she #do #every #Sunday <http://t.co/xxxxxx>

[51]. @user73: lmao rick ross just walked into my church #dying #toofunny #lookalike #notsomuch #butstillfunny

In [50] the tags function to emphasise the message rather than to provide any metadata. In [51] the user includes several successive “turns” in discourse within the same tweet, creating a condensed, yet complex unit of communication. The repetition of hashtags often serves to “scale up” their impact to the point of humorous hyperbole (Zappavigna, 2011).

4.5.2. FUNCTION

The functions of hashtags are extremely varied and are no longer restricted to the organisational purposes for which they were originally introduced; due to the specific purposes and narrow scope of this investigation, a comprehensive account of these functions is not possible. Outlined below, however, is a selection of the most prominent, reoccurring functions within the data, identified using content analysis and grounded theory methods.

Topical

The original function of the hashtag, these examples mark the semantic topic of the tweet, allowing others using the same tag to engage in open, asynchronous, ambient discourse via automated hyperlinks and internal search features. This practise is reminiscent of ‘collaborative tagging’ or ‘folksonomy’ (Bruns, 2008), user-led strategies of user-generated content categorisation (see 2.4.1.3).

[52]. @user38: Syria's refugee figures : 400'000 outside the country, 1 Million displaced inside the country #Syria.

[53]. @user61: Swaggy new logo i jus designed for the #creepvantour2012 #wickedwitch #emeraldgang #magicspellz <http://t.co/xxxxxx>

Topic categorisations include general issues and current events, such as example [52], denote events, or even refer to a group identity or community, as in [53].

Micro-meme

Memes²³ represent one of the quintessential components of user-generated content on Web 2.0. Twitter memes, or “micro-memes” (Huang, et al., 2010), are typically, though not exclusively, found in the initial syntactic position and prompt users to make light-hearted comments on a common theme:

[54]. @user64: #ThingsBoysSayAfterRejection Its not you its me , i just dont wanna ruin our friendship

Other micro-meme hashtags in the data set include #20peopleithinkarepretty, #greattobeaguy, #4wordsafterabreakup, #favoritelinesinclass and #20peoplewhoithinkarehandsome. The most popular micro-memes often form a Twitter “trend”, i.e. feature on a list of the most popular and fastest-growing words or phrases currently being tweeted.

Marking online discursive conventions

On Twitter, certain online discursive conventions, in particular abbreviations, are often marked by the hashtag symbol. Examples include: #oomf (one of my followers), #np (now playing), #lrt (last retweet), #nfb (now following back), #nw (now watching), and #tmlt (to my last tweet). The hashtags are highly unlikely to be intended for organisational purposes, and instead, they may represent attempts at disambiguation, i.e. to avoid possible interpretations such as spelling mistakes, or possibly as markers of identity, i.e. users want to demonstrate their mastery of Twitter discourse, thus signifying membership of a broad community or in-group.

Extralinguistic

Hashtags may concern strategies for representing extralinguistic features of discourse, in particular for adding emphasis to a single word (see [56]) or for signifying other nonverbal components of communication.

²³ A meme is “an idea, behaviour, style, or usage that spreads from person to person within a culture” (Merriam-Webster, 2013)

- [55]. @user04: Genuinely think i'm going to end up spewing in work or something, feel like utter shit #bleughh
- [56]. @user90: I feel your struggle fellas. As much as #some of us have high standards and expect to be treated like Queens it's a 2 way street
- [57]. @user96: Just woke up to the cutes text :) #wipethesmileoffmyface

Contextual

Twitter is a highly intertextual environment, with users often having to refer to background information in preceding tweets to make sense of current ones. As a strategy to combat the character restraints placed on messages, users often use hashtags to contextualise the main proposition of their tweets by offering additional informational, albeit in elliptical form.

- [58]. @user52: I just want to eat copious amounts of pizza and cuddle. #terribleday
- [59]. @user04: Really want 23rd of august to hurry up, so sick of this country right now #ZANTE #33DAYS

The truncated nature of these hashtags may lead to high degrees of ambiguity.

Emphatic

Such hashtags strengthen or confirm the proposition they accompany.

- [60]. @user31: The heart is like a parachute it works if u open it....#word
- [61]. @user42: Just looking out for you #NoLie

Focusing

Here the hashtag is used to isolate and give prominence to the key term(s) within a proposition. It differs from the *topical* function as these messages cannot be interpreted as being contributions to a larger conversation.

- [62]. @user32: This a big one #decision
- [63]. @user41: In the #gym

Humour

A significant proportion of hashtags represent various attempts at humour, of which irony is particularly prominent. Users appear to be conscious of the tacit “rules” of hashtagging, yet deliberately and blatantly violate them by publishing creative sequences unique to the context of the tweet.

- [64]. @user11: @addressee you're just a frigging hater, young lady! #hopoffmyjock #idontwannatweet everysecondlikeyou ;)
- [65]. @user07: Nicest run! Singing out loud to my own songs like a right nutter #imsocool

Evaluation

Perhaps the largest class of hashtags convey subjective and emotional reflections upon or associated with the main body of the tweet. As with [51] the users in the examples below include supplementary elliptical comments which expand the meaning of tweets, creating a condensed, complex unit of communication. In some cases, the hashtag(s) may convey more than the main body of the tweet.

- [66]. @user06: Gorgeous day in the bay with @obrienstours #soexcited
- [67]. @user81: Call me crazy, but I miss dressing in 5 layers for cold weather=/ #neverhappy

Hashtagging on Twitter is an emergent activity, yet has been too readily characterised as being a convention confined to marking the semantic topic of a tweet; the data collected here suggests in fact that this practice is now in the minority. Instead, hashtags have been appropriated to enable users to add creative, elliptical comments which give their tweets increased degrees of complexity and texture. In particular, hashtags often provide contextual, humorous and evaluative commentary on the message they accompany. The apparent presumptive misunderstanding of hashtagging practices on Twitter necessitates an in-depth, holistic analysis of how and why they are used.

5. Conclusion

By examining the communicative behaviour of 100 Twitter users over a 48-hour period, this study has provided a preliminary assessment of some of the pertinent linguistically-related discursive phenomena to be found on the medium through 5 separate research modules: in Section 4.1 an analysis of the basic habits of Twitter users showed that tweets are posted with variable frequency, with some using the medium only intermittently, while others utilise it frequently within the space of only an hour; in Section 4.2, a macro-level overview of the communicative functions of tweets was developed, which established that Twitter supports inter-user dialogue, the maintenance of social relationships and the performance of identity; Section 4.3 found that the @ symbol is used extensively (in 53.9% of outbound tweets) as a strategy to facilitate the coherent exchange of messages, contributing to a highly interactive environment in which interactions can occur frequently and over lengthy periods; Section 4.4 ascertained that the retweet function (identified in 25.2% of outbound tweets) fulfils numerous functions, most notably to contribute to a wider conversation and strengthen interpersonal relationships, and poses significant challenges in terms of the interpretation of such messages; and in Section 4.5, hashtags (found in 7.8% of outbound tweets), which were ostensibly introduced to explicitly semantic topics, were found to fulfil a number of diffuse functions seemingly unaccounted for by previous studies, and play a vital role in the creation of complex units of discourse.

For the average user, Twitter is no incidental communicative medium; instead, it plays an integral role in many individuals' communicative and social behaviours, and thus constitutes an important and widely-used addition to the family of new media technologies. However, that is not to say that the medium is necessarily revolutionary; many characteristics – the use of text, hyperlinks, content reproduction, et cetera - accord with Herring's classification (2011) of certain Web 2.0 discourse phenomena as 'familiar' or 'reconfigured'. New media technologies are usually embedded into the banal practices of everyday life (Thurlow & Mroczek, 2011), and may be best understood as 'prosthetic extensions' of people's abilities and lives, akin to a hearing aid or paper clip (cf. McLuhan, 1964). Proclamations regarding Twitter's anthropological influence – such as Lee's assertion (2011:118) that status updates have become “a crucial aspect of everyday life” - must therefore be considered with a degree of scepticism; this is very much a Western world generalisation, and fails to mention that it is NOT a crucial aspect of everyday life for the two-thirds of the world's population without Internet access, nor indeed for many competent Internet users.

Contrary to popular misconceptions, Twitter users neither constitute a homogenous mass, nor can be easily categorised according to their habits. Tweets fulfil a wide range of diffuse communicative purposes, having evolved far beyond providing a conduit through which users respond to the original prompt of “What's happening?” Through widespread use of the @ and retweet functions, prominent uses of the medium include using it to engage in dialogue, to exchange ideas and infor-

mation, and to develop social relationships, contributing to a highly interactive environment and dismissing the fallacy of microblogging being based on inane monologic chatter. Tweets exhibit high degrees of self-presentation and self-disclosure, which may result in the medium being categorised as exhibitionistic, but may also be considered as a key contributor to the ambient intimacy of social media, which can engender strong feelings of closeness and intimacy.

With regards to the specific constraints of the medium, the 140-character format is certainly a Twitter-specific restraint, but it need not be seen as a limitation and may even be seen as an advantage; the brevity of messages allows them to be produced, consumed, and shared without the need for significant investment in thought, time and effort, which, in turn, engenders a fast and fluid interactive environment (cf. boyd et al., 2010). Nevertheless, tweets represent complex and challenging linguistic units; they are highly intertextual, at times ambiguous, and often multi-faceted. Retweets may obscure meaning and raise questions regarding authorship, ownership, attribution and conversational fidelity, while hashtagging, an emergent practice which necessitates further study, can introduce additional levels of complexity through elliptical commentary.

Twitter is a highly dynamic environment which is perhaps only beginning to settle down after a short embryonic period during which it has grown exponentially. Aside from the character limitation, there is little evidence to suggest that tweets constitute a single written genre, but a hybrid of genre features identified in different text types, from instant messaging to texting and blogging (cf. Lee, 2011). Indeed, a key aspect of new media communication is the concept of online convergence, which suggests that users seldom employ the same set of genre conventions in all instances. From this perspective, attempts to conceptualise a distinct language variety such as “Twitterspeak” would be futile.

6. References

- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and Computer-Mediated Communication. *Journal of Sociolinguistics*, 10(4) (pp.419-438).
- Androutsopoulos, J. (2011). From Variation to Heteroglossia in the Study of Computer-Mediated Discourse. In C. Thurlow and K. Mroczek (Eds.), *Digital Discourse: Language in the New Media* (pp.277-298). Oxford: Oxford University Press.
- Baron, N. S. (2008). *Always On: Language in an Online and Mobile World*. Oxford: Oxford University Press.
- Baron, N. S. (2010). Discourse Structures in Instant Messaging: The Case of Utterance Breaks. *Language@Internet*, 7:4.
- Bauer, M. W. (2000). Classical Content Analysis: A Review. In M. W. Bauer & G. Gaskell (Eds.), *Qualitative Researching with Text, Image and Sound: A Practical Handbook* (pp.131-151). London: Sage Publications
- boyd, d. m. & Ellison, N.B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13(1):11.
- boyd, d. m., Golder, S. & Lotan, G. (2010). Tweet Tweet Retweet: Conversational Aspects of Retweeting on Twitter. *Proceedings of HICSS-42*, Kauai, HI: IEEE Computer Society. January 5-8, 2010.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and Beyond: From Production to Producersage*. New York: Peter Lang Publishing.
- Calvert, C. (2000) *Voyeur nation : media, privacy, and peering in modern culture*. Boulder, Colo.: Westview Press.
- Coates, T. (2003). *My working definition of social software...*
http://www.plasticbag.org/archives/2003/05/my_working_definition_of_social_software/
[accessed: 22-01-2013]
- Crystal, D. (2006). *Language and the Internet*. 2nd edition. Cambridge: Cambridge University Press.
- Crystal, D. (2011). *Internet Linguistics: A Student Guide*. Oxon: Routledge.
- Danet, B. & Herring, S.C. (2007). Introduction: Welcome to the Multilingual Internet. In B. Danet & S.C. Herring (Eds.), *The Multilingual Internet: Language, Culture and Communication Online* (pp.3-42). Oxford: Oxford University Press.
- Georgakopoulou, A. (2011). Computer-mediated communication. In J. O. Östman, & J. Verschueren (Eds.), *Pragmatics in Practice* (pp.93-110). Amsterdam: John Benjamins Publishing Company.
- Guardian, (2013). *Twitter now has 10m users in UK*.
<http://www.guardian.co.uk/technology/2012/may/15/twitter-uk-users-10m>
[accessed: 23-01-2013]
- Herring, S.C. (2001). Computer-mediated discourse. In D. Schiffrin, D. Tannen, & H. E. Hamilton (Eds.), *The handbook of discourse analysis* (pp.612-634). Oxford: Blackwell.
- Herring, S. C. (2004). Computer-mediated discourse analysis: An approach to researching online behavior. In S. A. Barab, R. Kling, & J. H. Gray (Eds.), *Designing for Virtual Communities in the Service of Learning* (pp. 338-376). New York: Cambridge University Press.
- Herring, S. C. (2008). Language and the Internet. In W. Donsbach (Ed.), *International Encyclopedia of Communication*. (pp. 2640-2645) Oxford: Blackwell.
- Herring, S. C. (2010). Computer-Mediated Conversation: Introduction and Overview. *Language@Internet*, issue 7, article 2.
- Herring, S. C. (2013). Discourse in Web 2.0: Familiar, Reconfigured, and Emergent. In D. Tannen & A. M. Tester (Eds.), *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and new media*. Washington, DC: Georgetown University Press.

- Herring, S. C., Scheidt, L. A., Bonus, S., and Wright, E. (2004). Bridging the gap: A genre analysis of weblogs. *Proceedings of the 37th Hawai'i International Conference on System Sciences (HICSS-37)*. Los Alamitos: IEEE Computer Society Press.
- Hine, C. (2005). Virtual Methods and the Sociology of Cyber-Social-Scientific Knowledge. In C. Hine (Ed.), *Virtual Methods: Issues in Social Research on the Internet* (pp. 1-13). Oxford: Berg.
- Honeycutt, C. & Herring, S. C. (2009). Beyond microblogging: Conversation and collaboration via Twitter. *Proceedings of the Forty-Second Hawai'i International Conference on System Sciences (HICSS-42)*. Los Alamitos, CA: IEEE Press.
- Huang, Y. (2007). *Pragmatics*. Oxford: Oxford University Press.
- Huang, J., Thornton, K. M., Efthimiadis., E. N. (2010) *Conversational Tagging in Twitter*. Seattle: University of Washington.
- ITU (International Telecommunications Unions). (2011). *The World in 2011: ICT Facts and Figures*. Retrieved from: <http://www.itu.int/ITU-D/ict/facts/2011/material/ICTFactsFigures2011.pdf> [accessed: 16-02-2013]
- Java, A., Finin, T., Song, X. & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. *Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop 2007*, (pp. 56–65). New York: ACM.
- Jenkins, H. (2008). *Convergence culture : where old and new media collide*. New York: New York University Press.
- Kaplan, A. M. & Haenlein, M. (2010). Users of the World, Unite! The Challenges and Opportunities of Social Media. *Business Horizons*, 53(1) (pp.59-68).
- Kaplan, A. M. & Haenlein, M. (2011). The Early Bird Catches the News: Nine Things You Should Know about Micro-blogging. *Business Horizons*, 54 (pp.105-113).
- Kassim, S. (2012). Twitter Revolution: How the Arab Spring was Helped by Social Media. *PolicyMic*. <http://www.policymic.com/articles/10642/twitter-revolution-how-the-arab-spring-was-helped-by-social-media> [accessed: 16-02-2013]
- Krishnamurthy, B., Gill, P., & Arlitt, M. (2008) A Few Chirps About Twitter. *Proceedings of the first workshop on online social networks* (pp.19-24).
- Lee, C. K. M. (2011). Micro-blogging and Status Updates on Facebook: Texts and Practices. In C. Thurlow & K. Mroczek (Eds.), *Digital Discourse: Language in the New Media* (pp.110-130). Oxford: Oxford University Press.
- Marwick, A. E. & boyd, d. (2010). I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience. *New Media Society* (published online). Retrieved from: <http://nms.sagepub.com/content/early/2010/06/22/1461444810365313>
- Marwick, A. E. & boyd, d. (2011). To See and Be Seen: Celebrity Practice on Twitter. *Convergence: The International Journal of Research into New Media Technologies*, 17(2) (pp.139-158).
- MediaBistro,(2013). *Twitter To Surpass 500 Million Registered Users On Wednesday*. http://www.mediabistro.com/alltwitter/500-million-registered-users_b18842 [accessed: 23-01-2013]
- McLuhan, M. (1964). *Understanding Media: the Extensions of Man*. New York: McGraw Hill.
- Merriam-Webster Dictionary. (2013). *Meme*. <http://www.merriam-webster.com/dictionary/meme> [accessed: 17-02-2013]
- Miller, C. R. & Shepherd, D. (2004). *Bloggng as Social Action: A Genre Analysis of the Weblog*. Retrieved from: http://blog.lib.umn.edu/blogosphere/bloggng_as_social_action_a_genre_analysis_of_the_weblog.html [accessed:20-07-2012]
- Myers, G. (2010). *The Discourse of Blogs and Wikis*. London: Continuum International Publishing Group.
- Nielsen. (2012) *Social Media Report 2012*. <http://blog.nielsen.com/nielsenwire/social/2012/> [accessed: 16-02-2013]

- Oktaý, J. S. (2012) *Grounded Theory*. Oxford: Oxford University Press.
- O'Reilly, T. & Milstein, S. (2012). *The Twitter Book*. 2nd edition. Sebastopol: O'Reilly Media, Inc.
- Peterson, E. E. (2011). How Conversational are Weblogs? *Language@Internet*, 8:8.
- Sanders, T. (2005). Researching the Online Sex Work Community. In C. Hine (Ed.), *Virtual Methods: Issues in Social Research on the Internet* (pp. 1-13). Oxford: Berg.
- Schneider, S. M., & Foot, K. A. (2005). Web Sphere Analysis: An Approach to Studying Online Action. In C. Hine (Ed.), *Virtual Methods: Issues in Social Research on the Internet* (pp. 157-170). Oxford: Berg.
- Thurlow, C. & Bell, K. (2009) Against Technologization: Young People's New Media Discourse as Creative Cultural Practice. *Journal of Computer-Mediated Communication*, 14(4) (pp.1038-1049).
- Thurlow, C. & Mroczek, K. (2011) *Digital Discourse: Language in the New Media*. Oxford: Oxford University Press.
- Thurlow, C. (2012). Fakebook: Synthetic Media, Pseudo-Sociality and the Rhetorics of Web 2.0. In D. Tannen & A. M. Tester (Eds.), *Discourse 2.0: Language and new media*. Washington, DC: Georgetown University Press.
- Trotta, J. & Danielsson, M. (2011) Image-based Online Communication: Observations on the Status of Images as Linguistic Constituents in Computer-mediated Communication. *McLuhan Galaxy Conference Proceedings: Understanding Media, Today*, (pp. 389-400)
- Twitter. (2013) *Twitter Privacy Policy*. <https://twitter.com/privacy> [accessed: 09-02-2013]
- Wikipedia. (2013a). *Web 2.0*. http://en.wikipedia.org/wiki/Web_2.0 [accessed: 20-01-2013]
- Wikipedia. (2013b). *Social Media*. http://en.wikipedia.org/wiki/Social_media [accessed: 22-01-2013]
- Wikipedia. (2013c). *Microblogging*. <http://en.wikipedia.org/wiki/Microblogging> [accessed: 12-01-2013]
- Wikipedia. (2013d). *Ambient Awareness*. http://en.wikipedia.org/wiki/Ambient_awareness [accessed: 24-01-2013]
- Zappavigna, M. (2011). Ambient Affiliation: A Linguistic Perspective on Twitter. *New Media & Society*, 13(5) (pp.788-806)

Software

- Anthony, L. (2011) *AntConc*. Version 3.2.4w. Available free for download from: <http://www.antlab.sci.waseda.ac.jp/software.html> [accessed: 08-01-2013]

7. Appendices

Appendix A. User statistics

<i>user ID</i>	<i>total tweets</i>	<i>tweets sent</i>	<i>tweets received</i>	<i>difference</i>	<i>tweets sent with hashtags</i>	<i>total hashtags used</i>	<i>interactions</i>	<i>retweets</i>	<i>active period (d:hh:mm:ss)</i>	<i>average lag time (d:hh:mm:ss)</i>
1	4	4	0	4	4	6	0	0	1 17:22:49	0 10:20:42
2	103	100	3	97	3	3	2	12	1 13:38:58	0 00:22:35
3	122	108	14	94	2	4	2	8	0 21:28:21	0 00:11:56
4	31	29	2	27	5	6	1	5	1 07:50:58	0 01:05:54
5	157	120	37	83	4	4	11	16	1 14:49:24	0 00:19:25
6	27	25	2	23	23	34	0	1	1 18:32:01	0 01:42:05
7	51	43	8	35	8	9	3	9	1 15:03:25	0 00:54:30
8	72	44	28	16	2	2	6	0	1 22:12:47	0 01:03:01
9	10	8	2	6	0	0	1	0	1 04:11:29	0 03:31:26
10	31	31	0	31	0	0	0	23	1 13:19:35	0 01:12:15
11	178	135	43	92	56	98	14	51	1 17:27:34	0 00:18:26
12	40	31	9	22	0	0	4	7	1 18:27:33	0 01:22:11
13	30	20	10	10	0	0	6	3	1 04:12:05	0 01:24:36
14	27	26	1	25	2	2	1	5	1 01:43:10	0 00:59:21
15	46	28	18	10	1	1	4	8	1 00:03:58	0 00:51:34
16	5	4	1	3	0	0	1	1	1 06:11:59	0 07:33:00
17	39	31	8	23	0	0	4	24	1 12:28:02	0 01:10:35
18	41	36	5	31	6	6	2	14	1 12:50:28	0 01:01:24
19	42	38	4	34	5	7	2	17	1 16:36:24	0 01:04:07
20	110	96	14	82	23	23	6	10	1 07:32:35	0 00:19:43
21	321	289	32	257	3	3	13	69	1 20:29:30	0 00:09:14
22	157	115	42	73	1	1	10	10	1 06:45:01	0 00:16:03
23	352	321	31	290	13	19	16	122	1 20:52:03	0 00:08:23
24	42	38	4	34	0	0	2	9	1 07:57:58	0 00:50:28
25	28	28	0	28	2	2	0	16	1 16:25:21	0 01:26:37
26	15	13	2	11	3	3	1	1	1 10:25:33	0 02:38:53
27	4	4	0	4	0	0	0	0	0 06:02:14	0 01:30:33
28	69	58	11	47	0	0	4	34	1 19:14:09	0 00:44:44
29	70	57	13	44	0	0	7	3	0 23:44:09	0 00:24:59
30	34	30	4	26	2	2	2	6	1 23:52:14	0 01:35:44
31	53	50	3	47	16	16	1	6	1 10:51:16	0 00:41:50
32	143	132	11	121	4	4	7	88	1 18:26:51	0 00:19:18
33	12	11	1	10	4	5	0	6	0 22:06:06	0 02:00:33
34	3	3	0	3	1	1	0	1	1 10:32:29	0 11:30:50
35	108	93	15	78	3	3	4	39	1 08:41:26	0 00:21:05
36	11	10	1	9	1	1	1	1	0 10:31:19	0 01:03:08
37	97	81	16	65	3	3	9	37	1 13:18:47	0 00:27:38
38	723	426	297	129	50	71	25	70	1 23:51:44	0 00:06:44
39	94	93	1	92	1	1	1	48	1 23:11:07	0 00:30:27
40	200	185	15	170	1	1	3	56	1 22:58:46	0 00:15:14
41	127	84	43	41	11	11	6	0	1 20:49:59	0 00:32:01
42	150	145	5	140	21	21	4	23	1 20:24:33	0 00:18:23
43	118	79	39	40	2	2	17	16	1 23:19:27	0 00:35:57
44	13	13	0	13	1	2	0	2	1 07:24:40	0 02:24:58
45	77	54	23	31	4	4	8	24	1 19:57:00	0 00:48:50
46	243	221	22	199	11	12	9	96	1 23:36:23	0 00:12:55
47	19	16	3	13	7	10	2	4	1 09:22:32	0 02:05:10
48	5	3	2	1	0	0	1	0	0 01:05:29	0 00:21:50
49	2	1	1	0	0	0	0	0	0 00:00:00	0 00:00:00

50	1	1	0	1	0	0	0	0	0 00:00:00	0 00:00:00
51	93	88	5	83	1	1	1	2	1 21:30:24	0 00:31:02
52	17	15	2	13	4	4	1	2	1 22:16:16	0 03:05:05
53	636	413	223	190	30	32	29	105	1 16:26:59	0 00:05:53
54	18	15	3	12	1	1	2	5	1 17:36:54	0 02:46:28
55	48	41	7	34	0	0	3	8	1 21:24:55	0 01:06:28
56	111	104	7	97	7	7	3	21	1 21:58:05	0 00:26:31
57	3	2	1	1	1	1	1	0	1 09:17:11	0 16:38:36
58	182	162	20	142	10	10	4	8	1 23:55:01	0 00:17:45
59	391	261	130	131	11	14	37	68	0 20:22:14	0 00:04:41
60	60	60	0	60	1	1	0	17	1 07:39:31	0 00:31:40
61	79	61	18	43	18	36	4	4	1 19:31:33	0 00:42:49
62	95	76	19	57	9	12	8	14	1 18:29:15	0 00:33:33
63	93	70	23	47	5	7	8	5	1 08:13:11	0 00:27:37
64	59	55	4	51	8	8	1	6	1 23:08:05	0 00:51:25
65	46	46	0	46	15	17	0	40	1 18:45:27	0 00:55:46
66	126	118	8	110	13	15	6	36	1 20:24:19	0 00:22:35
67	39	38	1	37	1	1	1	6	1 18:38:39	0 01:07:20
68	243	224	19	205	3	3	11	42	1 23:47:57	0 00:12:48
69	237	198	39	159	11	12	19	84	1 18:32:45	0 00:12:54
70	60	48	12	36	6	6	3	9	1 22:54:52	0 00:58:39
71	13	13	0	13	0	0	0	0	1 13:29:47	0 02:53:04
72	165	123	42	81	6	7	5	37	1 23:47:47	0 00:23:19
73	171	133	38	95	18	39	16	65	1 21:22:32	0 00:20:28
74	215	178	37	141	0	0	13	78	1 15:23:43	0 00:13:17
75	29	28	1	27	2	2	1	1	1 18:48:27	0 01:31:44
76	765	609	156	453	56	65	34	80	1 13:39:35	0 00:03:43
77	99	77	22	55	1	1	7	41	1 20:49:42	0 00:34:56
78	20	15	5	10	6	8	4	3	1 15:44:26	0 02:38:58
79	53	39	14	25	1	1	8	6	1 19:48:09	0 01:07:23
80	62	55	7	48	0	0	5	5	1 21:44:29	0 00:49:54
81	43	32	11	21	4	4	3	2	1 11:20:32	0 01:06:16
82	1	1	0	1	0	0	0	0	0 00:00:00	0 00:00:00
83	203	160	43	117	1	1	12	46	1 22:47:12	0 00:17:33
84	193	182	11	171	26	33	3	46	1 20:21:00	0 00:14:37
85	14	13	1	12	6	11	1	2	1 11:47:36	0 02:45:12
86	51	43	8	35	4	4	5	6	1 12:21:23	0 00:50:44
87	20	17	3	14	5	28	1	3	1 17:29:36	0 02:26:27
88	26	23	3	20	5	5	1	11	1 17:37:39	0 01:48:36
89	73	65	8	57	2	2	4	7	1 21:50:24	0 00:42:19
90	379	310	69	241	43	46	29	60	1 23:16:02	0 00:09:09
91	236	201	35	166	0	0	7	61	1 11:35:11	0 00:10:37
92	21	21	0	21	12	12	0	18	1 23:51:50	0 02:16:45
93	94	76	18	58	1	1	5	2	1 13:57:10	0 00:29:58
94	2	1	1	0	0	0	0	0	0 00:00:00	0 00:00:00
95	592	541	51	490	11	13	15	151	1 17:18:47	0 00:04:35
96	50	42	8	34	18	22	4	6	1 22:53:53	0 01:07:00
97	17	11	6	5	0	0	2	0	1 08:49:28	0 02:59:03
98	9	8	1	7	1	1	1	3	1 16:06:42	0 05:00:50
99	349	203	146	57	5	6	7	4	1 22:47:40	0 00:13:50
100	129	39	90	-51	5	5	7	32	1 18:45:50	0 01:05:47
TOTAL	11187	8965	2222	6743	703	908	565	2259	-	-
average	111,9	89,7	22,2	67,4	7,0	9,1	6,5	22,6	1 13:01:43	0 01:22:34
median	59,5	47,0	8,0	36,5	3,0	3,0	4,0	8,0	1 16:57:36	0 00:43:46
mode	4,0	13,0	-	10,0	-	-	1,0	-	-	-

Appendix B. Tweet frequency graphs

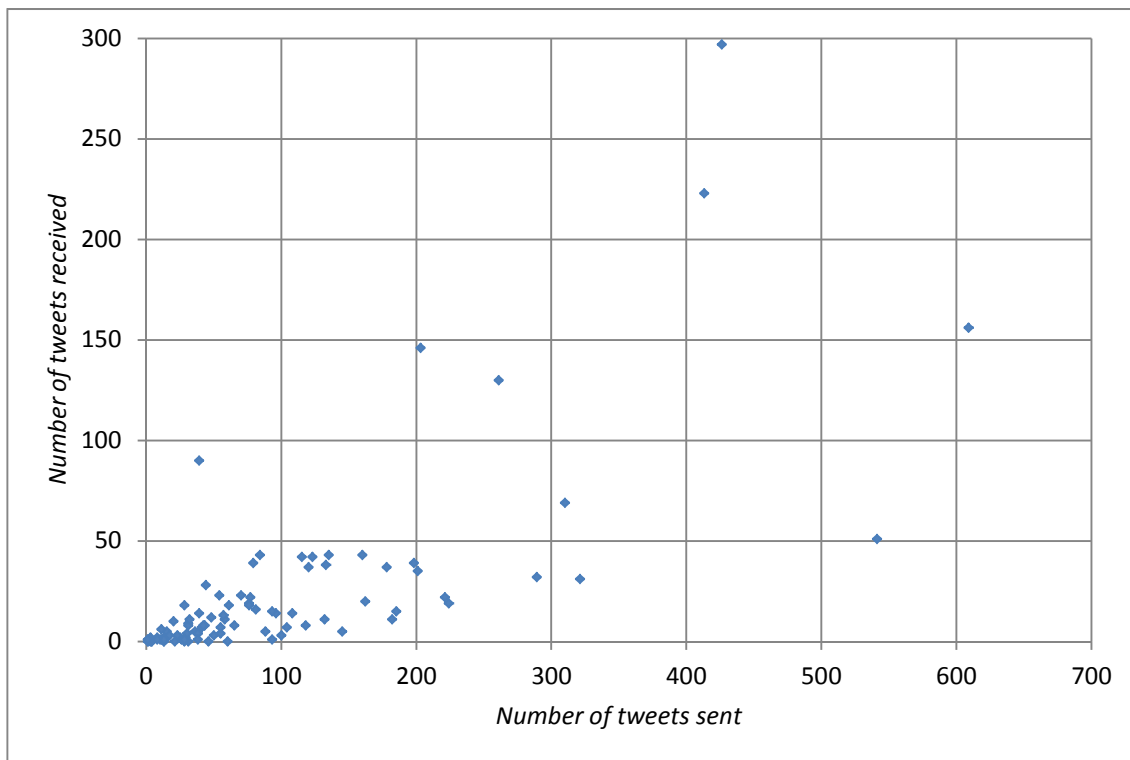


Figure B.1. Tweets sent measured against tweets received

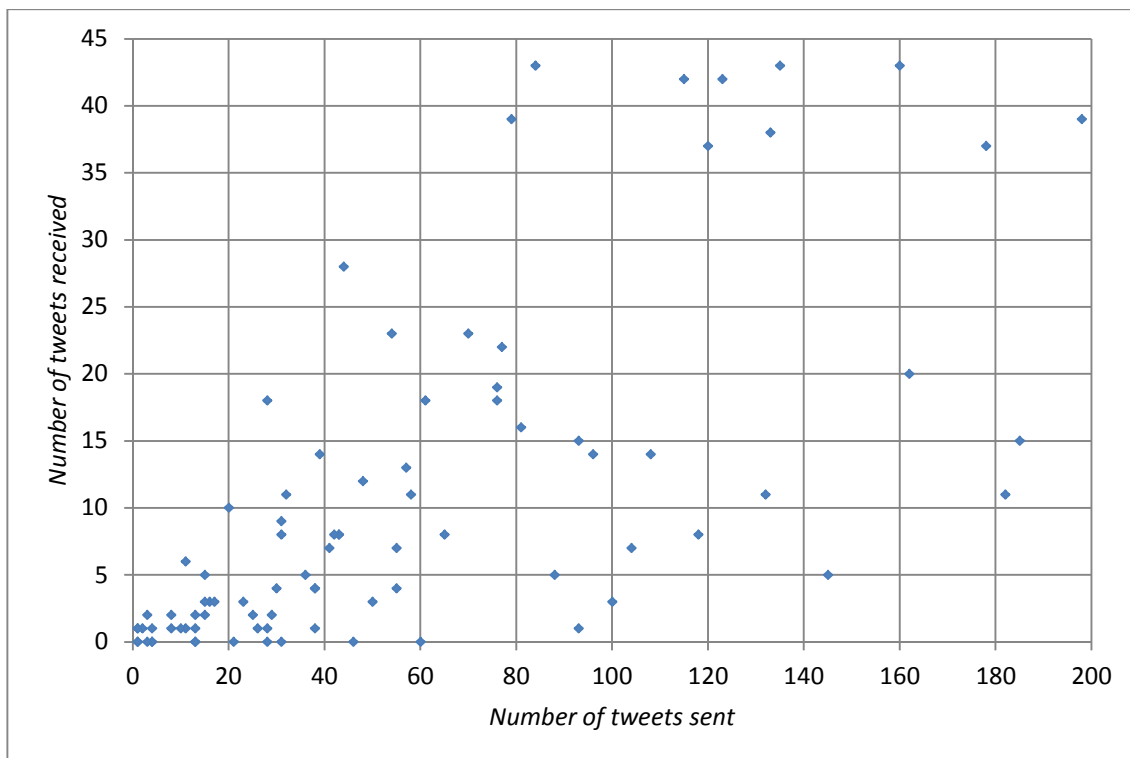


Figure B.2. Tweets sent measured against tweets received minus “prolific” users, i.e. whose tweets sent > 200, or tweets received > 90

Appendix C. Functions of the @ sign

	total freq.	total %	tweets sent freq.	tweets sent %
addressivity	5166	62.8%	2936	49.4%
<i>original addressee of a RT (user)</i>	<i>(304)</i>	<i>(3.7%)</i>	<i>(303)</i>	<i>(5.1%)</i>
<i>original addressee of a RT (other)</i>	<i>(58)</i>	<i>(0.7%)</i>	<i>(58)</i>	<i>(1.0%)</i>
reference	657	8.0%	622	10.5%
<i>reference to user in a RT</i>	<i>(51)</i>	<i>(0.6%)</i>	<i>(51)</i>	<i>(0.9%)</i>
<i>'via' reference in a RT</i>	<i>(47)</i>	<i>(0.6%)</i>	<i>(47)</i>	<i>(0.8%)</i>
<i>other references to user</i>	<i>(2)</i>	<i>(0.02%)</i>	<i>(2)</i>	<i>(0.3%)</i>
original author of a RT	2363	28.7%	2356	39.6%
<i>author of retweet is user</i>	<i>(49)</i>	<i>(0.6%)</i>	<i>(48)</i>	<i>(0.8%)</i>
other	43	0.5%	33	0.6%
<i>unclear instances</i>	<i>(19)</i>	<i>(0.2%)</i>	<i>(13)</i>	<i>(0.2%)</i>
grand total	8229	-	5947	-

NB. Figures in italics represent sub-totals

Appendix D. Most frequent hashtags

<i>frequency</i>	<i>hashtag</i>	<i>no. users</i>
38	#myboxrocks	1
20	#20peoplethinkarepretty	2
19	#oomf	8
16	#iraq	1
13	#greattobeaguy	1
12	#2yearsof1d	3
11	#syria	1
10	#sundaysupper	1
8	#thingsboysayafterrejection	4
7	#4wordsafterabreakup	3
6	#20peoplewhoithinkarehandsome	1
6	#np	4
6	#stlcards	1
6	#uk	1
5	#favoritelinesinclass	2
4	#emeraldgang	1
4	#firstworldproblems	1
4	#girl	1
4	#lrt	3
4	#mtpearlcitydays	1
4	#nfb	2
4	#rip	2
4	#summer	1
4	#us	1
4	#weird	4
4	#word	3
3	#awkward	3
3	#cute	1
3	#finally	3
3	#fun	1
3	#icecreamweek	1
3	#justsaying *	2
3	#loveit	3
3	#memories	2
3	#night	2
3	#nolie	2
3	#nw	2
3	#sorrynotsorry	3
3	#soundcloud	2
3	#thingsido	1
3	#tmlt	1
3	#truth	2
3	#usa	1

* including *#justsayin*, i.e. with an omission of the final 'g'