DISCOURSE OF TWITTER AND SOCIAL MEDIA

HOW WE USE LANGUAGE TO CREATE AFFILIATION ON THE WEB

MICHELE ZAPPAVIGNA

is ava.

have some patience,

while the hand is at rest on the page.

each extra, with

See you at 6 jp?

about half past

for

BLEAH!

I WAS THINKING OF
GETTING A NEW ONE;

The endearing character from the

and we dan sit down and

with   my   own

Regarding every type of

Show You What to Do

sing_iug i

... when you want to

have you got a

as an subsequent complaints can be

" Ah,

SEND NO

educate

can organise everything and

know your opinion,

From: <auto@de

housands and thousands of

doing something about it.

It's high time

London

Copies

pleasure in

Nº 35

ood low cost accessory. It will help with

Duke

that we've cut are the prices.

make the most of what

Your

continuum

Discourse of Twitter and Social Media

CONTINUUM DISCOURSE SERIES

Series Editor:
Professor Ken Hyland

Discourse is one of the most significant concepts of contemporary thinking in the humanities and social sciences as it concerns the ways language mediates and shapes our interactions with each other and with the social, political and cultural formations of our society. The Continuum Discourse Series aims to capture the fast-developing interest in discourse to provide students, new and experienced teachers and researchers in applied linguistics, ELT and English language with an essential bookshelf. Each book deals with a core topic in discourse studies to give an in-depth, structured and readable introduction to an aspect of the way language in used in real life.

Other titles in the series:

The Discourse of Text Messaging, Caroline Tagg
Historical Discourse: The Language of Time, Cause and Evaluation,
    Caroline Coffin
Metadiscourse: Exploring Interaction in Writing, Ken Hyland
Professional Discourse, Britt-Louise Gunnarsson
Using Corpora in Discourse Analysis, Paul Baker
Workplace Discourse, Almut Koester

Discourse of Twitter and Social Media

MICHELE ZAPPAVIGNA

continuum

For my mother, Jill, and my grandmother, Elma,
who babysat tirelessly so that I could complete
this work during my maternity leave

CONTENTS

ACKNOWLEDGEMENTS

CONVENTIONS USED IN THIS BOOK

Systems of appraisal are shown in SMALL CAPS so as to differentiate them from their plain English equivalents. Codings of appraisal systems use the following conventions: AFFECT in **bold**, JUDGEMENT in **bold underline** and APPRECIATION in ***bold italics***. Linguistic features to be highlighted in general examples are shown in **bold**. All usernames in tweets, aside from those of celebrities and institutions, have been anonymised with the convention @user.

PREFACE

The history of linguistics is punctuated by sea changes of culture afforded by new technologies of communication – first writing, then audio recording and most recently Web 1.0 and Web 2.0. The invention of writing enabled the emergence of linguistics as a discipline and certainly encouraged the shift from a more rhetorical orientation to language focusing on speaking to a more logical orientation taking writing as a norm. It then took a couple of millennia before audio recording enabled linguists to effectively tackle spoken language again, as they set to work on phonetics and phonology and on dialectal and registerial variation with renewed vigour. Throughout the twentieth century, however, collecting and transcribing audio records remained costly, with advancing computer technology facilitating storage and retrieval but not offering a great deal of help with data gathering or analysis. Then comes Web 1.0, making huge quantities of essentially monologic data available for the first time, and then Web 2.0 and the launch of social networking. How has this positioned linguistics for the next decade and beyond?

In this volume Michele Zappavigna lays the foundation for a forthcoming generation of work in internet linguistics, drawing on her training in social semiotics, linguistics and information technology. This necessarily involves discussion of how to gather data from Web 2.0, how to use corpus linguistics to process it, how to use functional linguistics to interpret it and how to use social semiotics to make sense of what is going on. The most dramatic turn here, as far as linguistics is concerned, is her interpersonal focus on ambient sociality. This she explores in terms of the way in which tweeters affiliate through searchable talk, demonstrating for the first time in a large scale study how communities constitute themselves through shared values – where it's not just interaction that matters but shared meaning and where what is being shared is feelings about ideas (not just the ideas themselves). This axiological orientation, based as it is on appraisal theory and quantitative analysis, goes a long way to balancing the ideational bias which has for so long delimited linguistics as a theory of writing and holds great promise for the evolution of a more social sensitive and socially responsible discipline in the years to come.

This turn is not of course without its challenges. The sheer scale of the enterprise makes it hard to see the forest for the trees, making the development of novel two- and three-dimensional animated visualisations

a priority. Alongside this are the trials of streaming data, as a microblog unfolds, as a blogger develops and as Web 2.0 evolves; the contingencies of time matter and cannot be theorized away. Finally, and perhaps most challengingly, Web. 2.0 is more than words, and ever more so; this demands not just a linguistics of words but a semiotics of multimodality, with all the implications for data gathering, analysis, interpretation and theorizing such entails. To her credit, Zappavigna dodges none of these issues and, with respect to the first two, shows us the way forward. We'll be hearing a lot more from her along these lines.

> As communication changes, so must its theorists. I for one welcome our new social semiotic overlords! :P

<div align="right">J. R. Martin</div>

CHAPTER ONE

Introduction

The social web and searchable talk

mum broke the blender. #fail

my prayers and thoughts are for those in #Christchurch #eqnz YOU GUYS WILL PULL THROUGH!!!!

The above microposts[1] were published online with a microblogging service; that is, an online platform for posting small messages to the internet in chronological sequence. The first post describes a mundane event in daily life. The hashtag #Fail aligns this personal expression with other potential instances of quotidian failure sharing the same tag. Hashtags are an emergent convention for labelling the topic of a micropost and a form of metadata incorporated into posts. The second post is a more public declaration of support for victims of an earthquake in the New Zealand city of Christchurch. It is similar to many other microposts produced during natural disasters and crises throughout the world. Like the first post, the hashtags #Christchurch and #eqnz, in this instance, seek parallel voices. The two posts differ greatly in the kind of connection they construe with a putative audience. Yet both actively invite connection with that audience by incorporating a hashtag to label the meanings they express. This kind of discourse tagging is the beginning of *searchable talk*, a change in social relations whereby we mark our discourse so that it can be found by others, in effect so that we can bond around particular values (Zappavigna 2011b).

Here we have the potential for users to commune within the aggregated gaze made possible with digital media, which I shall call ambient affiliation in this book. In other words, virtual groupings afforded by features of electronic text, such as metadata, create alignments between people who have not necessarily directly interacted online. Indeed these users may never

be able to grasp the extent of the emergent complexity in which they are involved due to the fast paced, organic quality of the connections generated. The social relationships made possible will emerge over time, generated and influenced by unfolding linguistic patterns.

The social web, or Web 2.0, are popularized terms used to signal a shift toward the internet as an interpersonal resource rather than solely an informational network. In other words, the social web is about using the internet to enact relationships rather than simply share information, although the two functions are clearly interconnected. Table 1.1 gives an overview of the ways in which the social web is said to differ from the first incarnation of the internet. At its centre is 'user-generated content'; that is, self-publication by users[2] of multimedia content such as blogs (websites displaying entries in reverse chronological order), vlogs (video blogs such as those posted regularly by millions of users on YouTube) and microblogs (streams of small character-constrained posts). This book will deal primarily with microblogging, sometimes seen as a cousin of more lengthy blogging, although likely quite a different meaning-making resource entirely.

The advent of social media, technology that aims to support ambient interpersonal connection, has placed new and interesting semiotic pressure on language. This book is concerned with the interpersonal dimension of making meaning using this new media. Social media is an umbrella term generally applied to web-based services that facilitate some form of social interaction or 'networking'. This includes websites where the design-principle behind the service is explicitly about allowing users to create and develop online relationships with 'friends' or 'followers'. The term also encompasses platforms where the focus is on generating and sharing content, but in a mode that allows comment and, potentially, collaboration.

Table 1.1 A general comparison between Web 1.0 and Web 2.0, taken from (Hsu and Park, p. 2)

	Web 1.0	**Web 2.0**
Mode of usage	Read	Write and contribute
Unit of content	Page	Record
State	Static	Dynamic
How content is viewed	Web browser	Browsers, RSS (Really Simple Syndication) readers, mobile devices, etc.
Creation of content	By website authors	By everyone
Domain of	Web designers and geeks	A new culture of public research?

Because of the rapid development of social media technologies and their constant change, they can be somewhat of a 'moving target' for scholars (Hogan and Quan-Haase 2010).

While I will comment on social media in general, my analysis will focus on patterns of meaning in microblogging. The method adopted in this book for exploring this patterning combines quantitative analysis of a large 100 million word corpus of microposts, HERMES, with a qualitative social semiotic approach to discourse analysis. These posts were taken from the microblogging service Twitter. Developed in 2006, Twitter allows users to post messages of 140 characters or less to the general internet or to a set of users who subscribe to a user's message 'stream',[3] known collectively as followers. These microposts are referred to as tweets and are presented to the user in reverse chronological order as an unfolding stream of content. This content is public and searchable unless the user actively makes his or her account private. Tweets may be accessed, sent and received via a variety of methods such as the web, email, SMS (Short Message Service) and third-party clients, often running on mobile devices. A tweet may also incorporate links to micromedia, small-scale multimedia and shortened aliases of longer hyperlinks (tiny URLs) intended to conserve characters within the constrained textual environment. In addition, Twitter collects supplementary metadata about a tweet, such as the time it was generated, the ID of the user to which the tweet was directed (if applicable) and information about the user's account, including the number of followers and the number of tweets the user has posted.

An example of a tweet is the following, which, as we will see in the third chapter, contains one of the most common patterns in microblogging, namely, a user thanking a follower for promoting him or her in the community:

@User1 thanks for the #FF

This post thanks User1 for the mention as part of Follow Friday, a collective practice where users are encouraged to endorse noteworthy tweeters. Tweets also contain metadata for managing interaction with others, for instance, @ indicating address (or reference) and # labelling topic. I will explore these features in more detail later in this book. While most of the metadata collected by Twitter is not presented directly to the general user, a notable exception is hashtags. These tags, a kind of in-text tagging visible within the body of a tweet, arose out of community use and were later incorporated into Twitter's search interface.

From a simplified technical perspective, Twitter consists of a farm of large databases that store the tweets that users have posted. The volume of this database of natural language is extremely large, with many millions of tweets being posted each day. Since its inception, Twitter has seen an extremely large increase in its user base. Prior to 2008–09 the service was

mainly used within the technology scene in the United States, after which it became more mainstream, with users posting status updates on everything from the most innocuous details of their personal lives to serious political opinion and reporting (Marwick 2010). Twitter's technical architecture has had to cope with a very high volume of traffic:

> Folks were tweeting 5,000 times a day in 2007. By 2008, that number was 300,000, and by 2009 it had grown to 2.5 million per day. Tweets grew 1,400% last year to 35 million per day. Today, we are seeing 50 million tweets per day – that's an average of 600 tweets per second. (Weil 2010)

From the viewpoint of users, Twitter consists of an interface that allows people to post new tweets, configure various settings, such as privacy, manage their list of followers and search historical tweets. Users may also interact with the service via a third-party application that presents the feeds of microposts in different ways, in some cases in novel visual forms. The extremely large volume of naturally occurring language is of great interest, as data, to linguists.

An important property of the social web is how it responds to time. Social media content is most often chronologically displayed. Indeed, many commentators describe the emergence of a 'real-time web'[4]; that is, a paradigm whereby web content is streamed to users via syndication. Tools, such as a feed reader, are used to aggregate many web feeds into a single view, meaning that users do not have to visit sources individually for current information. This type of convergent, real-time web experience combines with the social web to produce a semiotic world in which users have almost immediate access to what is being said in their social networks at any given moment. Users are able to subscribe to feeds of their associates' status updates and multimedia content (e.g. photos and video). Often these updates are shared via a mobile device at the time an event occurs or an observation is made.

An example of the real-time web in action might be the reactions to a major public event, such as the 2008 US presidential election (see Chapter 9), that are shared via microblogging. Even local, less significant events, such as weather, can trigger communal response. For instance, during a recent hailstorm in Sydney, a proliferation of comments about hail and photographs of hail rapidly appeared in my online social network via status updating on Facebook and Twitter. Many other kinds of events generate widespread social media response, including natural disasters, such as the 2008 Sichuan earthquakes (Li and Rao 2010), and celebrity deaths, a notable example being the collective outpouring of grief seen on Twitter after the death of Michael Jackson in 2009 (Sanderson and Hope Cheong 2010).

An important function of social media is sharing experience of the everyday within this real-time paradigm. In microblogging this often involves bonding around collective complaint about life's little daily irritations (see Chapter 8). While social media can be used 'like Momus windows of Greek mythology, revealing one's innermost thoughts for all to see' (van Manen 2010), most users are conscious of not overexposing their followers to banalities, a practice known as over-sharing, or 'attention whoring' (Marwick 2010).

Perhaps the most commonly used form of social media is the social networking service (SNS). This technology, used by millions of people worldwide, generates a very large volume of multimedia text. At the time of writing in 2010, Facebook had over 500 million users, each with an average of 130 Facebook 'friends' (Facebook 2010), and Twitter users were generating 65 million Tweets a day (Schonfeld 2010). SNSs are services with which users create an online profile about themselves with the goal of connecting with other people and being 'findable'. Boyd and Heer (2006) suggest the role of online social network profiles in identity performance as an 'ongoing conversation in multiple modalities'. Indeed, interactions via social media are usually likened to some form conversation. Depending on the kind of relationship being construed, the 'dialogue' may be fairly limited, often involving two main avenues: making initial contact with a user and then maintaining occasional contact at important dates, such as birthdays (2010).

Most SNSs have in common a number of basic functions: profile creation, the ability to generate a list of affiliated users, privacy customization, and a mechanism for viewing the activities of affiliated users. These affiliated users are often referred to as 'friends' (e.g. Facebook friends) or 'followers' (e.g. Twitter followers). Boyd (2010, p. 39) categorizes SNSs as a genre of 'networked publics' involving an 'imagined collective' arising from particular permutations of users, their practices and the affordances of technology. Four affordances Boyd suggests are of particular significance:

- persistence (capture and archiving of content);
- replicability (duplication of content);
- scalability (broad visibility of content);
- searchability (access to content via search).

As we will see in the next chapter, these properties, particularly persistence and searchability, mean that SNSs afford an opportunity to collect and analyse many different aspects of online discourse. The large volume of data made publically available by these services offers a fascinating window on social life, though it also raises a range of ethical concerns about how this data is used (Parrish 2010).

Interpersonal search

Through the social web, talk is 'searchable' in a way and to an extent that has never been seen in history. The advent of social media means that the function of online talk has become increasingly focused on negotiating and maintaining relationships. From a semiotic perspective the searchability affords new forms of sociality. We can now search to see what people are saying about something at a given moment, not just to find information. This makes possible what I will refer to as *interpersonal search*; that is, the ability to use technology to find people so that you can bond around shared values (or clash over discordant ones!). The searchability is particularly useful for linguists collecting particular kinds of discourse, for example, online chatter about a particular topic or language occurring in a particular geographical region. The conversation-like interactions possible with social media can be tracked in ways not readily achievable with face-to-face interaction in the real world, where it would be invasive to monitor a person's private interactions.

In popular terms, it is becoming increasingly useful to search the 'hive mind': the stream of online conversation occurring across semiotic modes (e.g. blogs, online chat and social networking sites). The kind of real-time discourse search that Twitter affords has been described as a rival to a Google search, with commentators claiming that searching Twitter may soon be one of the most effective ways to gather useful information (Rocketboom 2009). Microblogging streams offer a way of finding out about dominant trends in what people are saying. For instance, consider the following anecdote by Boyd (2009), who applies an ethnographic perspective to her social media research:

> I have a funny habit. Every day, I login to search Twitter and search for common words. Admittedly, I primarily search in English because my language skills in other languages are poor. But sometimes, I entertain myself by looking in other languages just for fun. I search for words like 'the' or, better yet, 'teh' just to see what people will write. I look for common names or random words.

> Why on earth do I do this? I do this in order to habitually look at worlds that are different than my own. As a researcher and a scholar, this is an essential technique. I am familiar with Twitter and Facebook and MySpace as a participant, but to observe, I need to move beyond my narrow frame. Thank goodness for search and browse. I look into the lives of people in order to get a sense of the different cultural practices that are emerging. But you can also look into what people are doing.

This notion of interpersonal search is a linguistic perspective on the concept of social search used within information science. The term 'social search'

refers to a mode of searching that leverages a user's social networks, for example, by asking a question on a social networking site. This mode of search is deemed complementary to an informational search with a search engine. Evans and Chi (2008) provide the following definition:

> 'Social search' is an umbrella term used to describe search acts that make use of social interactions with others. These interactions may be explicit or implicit, co-located or remote, synchronous or asynchronous. (Evans and Chi 2008, p. 485)

Meredith Ringel and colleagues surveyed users of social networking services, such as Twitter and Facebook, about the nature and motivation of the questions they asked using these media. Recommendation and opinion questions were the most frequent; a recommendation question was defined as asking 'for a rating of a specific item' and an opinion question as 'open-ended request for suggestions' (2010, p. 1742). The most common motivation expressed by respondents for engaging in this form of social search was a higher perceived level of trust in their friends to provide replies to questions deemed too subjective to be effectively answered by a search engine. Indeed, one study found differences in the kinds of queries issued to the Twitter search engine compared with those issued to a web search engine and that 'Twitter results included more social content and events, while Web results contained more facts and navigation' (Teevan et al. 2011, p. 44). These studies suggest that interpersonal meaning is at the heart of social search. Social bonds have becoming increasingly important in the way information is located and consumed with the internet. The significance of interpersonal search forms part of an explanation for the proliferation of social recommendation sites, social bookmarking, crowdsourcing[5] and related practices that make use of the social opinions extractable from online networks.

Approaches to online social networks

Social media affords a lens on types of social interaction previously not easily viewed. The streams of online social contact produced by users leave permanent traces that can be captured and modelled by researchers trying to understand the properties of the social networks arising with these social media feeds. Emerging areas of interest range across a large number of disciplines: marketing and opinion mining, information systems, computational linguistics and psychology. An example of the last is a study of the relationship of personality to social media which found extroversion to be positively associated with social media use and emotional stability to be negatively associated (Correa et al. 2010). The scope of networking practices seen with social media is immense. This book deals primarily

with microblogging networks, which, due to the particularities of the channels within which they are formed and evolve, may be distinct from practices developing in other kinds of social media. As a consequence this section will primarily review research into social networks formed in microblogging.

Most studies of social media communities, particularly computationally oriented studies of microblogging communities, use some derivative of social network analysis (SNA) (Wasserman and Faust 1994). SNA typically models and visualizes links between users based on a range of criteria such as frequency of contact and topic of message. It generally focuses on the topology, structure and evolution of these networks (Lin et al. 2009; Ahn et al. 2007; Wilson and Nicholas 2008; Holme et al. 2004). For example, Kumar and colleagues (2010, p. 612) note the prevalence of star structures in online social networks, where this structure represents 'a single charismatic individual (in the online sense) linked to a varying number of other users who have very few other connections'. According to the study these networks are made up of 'singletons' (non-participatory users), star-structured 'isolated communities', and 'a giant' component (a core region of the network unaffected by removal of star networks). Some studies look at how one variable (for instance, restricting the number of a user's friends) affects the topological properties of a network (Ghosh et al. 2010).

Computationally, interest generally lies in trying to understand the properties of online social networks. In general this kind of work uses techniques from computational linguistics to answer questions such as 'Given a social network and known preferences or behaviors of individuals in the network, how can we employ the connectivity to determine the preferences or behaviors of others in the network?' (Tang et al. 2010, p. 1). For example, we might build models using supervised[6] machine learning to predict which users will post tweets containing which URLs (Galuba et al. 2010). Alternatively the modelling may use unsupervised machine learning. For example, topic models can be used in predicting network structure. One study suggested that topic models can outperform some supervised alternatives derived from traditional social network analysis for determining network structure (Puniyani et al. 2010, p. 19). Another approach would be to track a particular social variable with no a priori definition of the social network's shape.

Alongside considering the shape of a social network is interest in how information travels through the network. The concept of 'influence' (Weng et al. 2010; Cha et al. 2010) refers to which users have the most impact on 'information diffusion', where diffusion is defined as the patterns by which information spreads 'virally' through a network (Liere 2010). Work has also been undertaken on the authority of a microblogger's posts by considering the link structure of explicit connections made between users

and also by supplementing this link structure by factoring in messages that are rebroadcast by other users (Yamaguchi et al. 2010).

These notions of influence and authority give rise to theorizing how communal attention operates in social media. In terms of users' interaction with content, microblogging seems to involve a kind of 'information snacking' (Brooks and Churchill, p. 4), where users pay attention to content that interests them at the time they happen to be accessing the service, dropping in and out of the discourse over time. Users, however, are still able to leverage this form of sporadic attention for their own social purposes. Marwick (2010, p. 230) suggests that the particular type of attention that social media encourages commodifies online persona, with many users aiming to achieve a form of 'micro-celebrity'. In other words they adopt 'a mindset and set of practices in which one's online contacts are constructed as an audience or fan base, popularity is maintained through ongoing fan management, and self-presentation is carefully assembled to be consumed by others'. Users' streams of updates become a means for establishing their 'brand' and social status, measurable by the level of attention the brand attracts, as manifest in the number of followers the user accumulates.

Attention is also significant when thinking about social media in terms of its relationship to existing news media. For example, while microblogging has a broader, more varied body of content producers, 'attention remains highly concentrated, where roughly 0.05% of the population accounts for almost half of all attention' (Wu et al. 2011). Wu and colleagues also note that networks of elite users have a highly homophilous structure in terms of attention, 'with celebrities following celebrities, media following media, and bloggers following bloggers'.

Practical applications of research into online social networks include 'opinion mining' in marketing, which is concerned with determining consumer opinion about brands (Jansen et al. 2009). Another less dominant area is the development of health-related services. Examples include the detection of influenza outbreaks with analysis of large quantities of microblogging posts (Culotta 2010) or monitoring of public concern about health issues, such as the H1N1 outbreak in 2009 (Chew and Eysenbach 2010). The principal area of application aimed at capitalizing financially on social media is viral, 'word-of-mouth' marketing. Viral marketing relies on brand awareness generated by content, often a video or image, that spreads rapidly throughout a social media network.

Language and affiliation in social media

Since forming and maintaining social bonds is one of the main functions of language, it seems reasonable to conclude that linguistic analysis may offer

a useful lens on how networks of interpersonal relationships are formed and maintained. A linguistic perspective on virtual community aims to describe explicitly how people use language to construe social bonds by creating interpersonal meaning. Historically, interpersonal meaning has been marginalized as an area of inquiry in linguistics (Poynton 1990). The reasons are both practical and political. Interpersonal meaning is difficult to study because it is prosodic in nature – in other words, it is not reducible to constituent parts (Martin 1996; Halliday 1979). Since a constituent structure is a prerequisite for annotation and statistical analysis, prosodic meanings may evade quantification. In addition Poynton (1990, p. 8) suggests that 'because of the hegemony of Chomskian linguistics from the 1960s, . . . when linguists wanted to get back to the social (as increasingly they have from the 1970s), there is a built-in hierarchy which gives priority to the cognitive and individual over the interpersonal and social'.

While studies of online discourse from a linguistic perspective are relatively established (Crystal 2006; Baron 2008; Herring 1996), whether analysis of linguistic function and structure can serve as evidence for defining online communities is an emergent area of inquiry. Approaches to affiliation within sociolinguistics and discourse analysis have used concepts such as speech communities (Labov 1972; Bolinger 1975), discourse communities (Swales 1990; Nystrand 1982) and communities of practice (Wenger 1998; Meyerhoff 2008) to explore linguistic patterning and social structure. These concepts have resounded with the interest of CMC (Computer-mediated Communication) studies in online community formation (for an overview, see Androutsopoulos [2006], who has summarized some of the main tendencies in CMC in this respect).

Using linguistics as a lens on community means that we are using semiotic evidence to group instances of meaning-making, rather than simply using contextual speaker variables, such as age or geography. Evidence of this kind is important to ensuring that the classification is more than a folk categorization. As Herring (2004) notes, much research into CMC tends towards generalization rather than applying a particular strategy for empirically analysing the evidence about the online behaviour studied:

> 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';[1] the language of the Internet is a single style or 'genre'[2]. Notions such as 'community' and 'genre' are familiar and evocative, yet notoriously slippery, and unhelpful (or worse) if applied indiscriminately. An important challenge facing Internet researchers is thus how to identify and describe online phenomena in culturally meaningful terms, while at the same time grounding their distinctions in empirically observable behavior. (Herring 2004, p. 338)

The basic notion of online community was popularized in Rheingold's work on 'virtual community', where he suggests that virtual communities 'emerge from the Net when enough people carry on . . . public discussions long enough, with sufficient human feeling, to form webs of personal relationships in cyberspace' (1993, p. 5). Since the emergence of this definition, often criticized for its vagueness, there has been a debate surrounding which criteria establish the bounds of an online community and the structure of such community and how communities are built or emerge (see, for example, Wellman 2001; Hagel and Armstrong 1997; Jones 1997; Burnett 2000; Herring 1996, 2004, 2008). No stable definition of online community has prevailed. The metadata collected by most SNSs may assist with these problems and be used in quantitative studies of SNS usage, as well as to complement discourse analysis.

Some language-based research into social media communities has been begun by computational linguists in the area of sentiment analysis – and natural language processing more generally. This work attempts to automate the detection of language patterns for categorizing groups of users into social networks (Haythornthwaite and Gruzd 2007; Gruzd 2009). Nguyen and colleagues (2010, p. 23) propose 'grouping patterns of communities purely from their sentiment'. This study used both mood tags, a form of metadata applied to blog posts by their authors, and 'emotion-bearing words' (Nguyen et al. 2010, p. 26), extracted from the content of posts, to classify sentiment. Chapter 4, on appraisal in microblogging, explores how a rich linguistic model of evaluative language can assist in understanding how sentiment is construed in social media. Following on from this, in Chapter 5, I provide a model of a type of ambient affiliation seen around evaluative targets marked by Twitter hashtags.

A social semiotic perspective

The perspective on ambient bonding applied in this book relies upon a social semiotic approach to language as it is used in its functional contexts (Halliday 1978). The discourse analysis undertaken is informed by systemic functional linguistics (SFL). SFL acknowledges that language is itself 'full of resources for negotiating community – both across metafunctions (ideational resources such as technical and specialized lexis, interpersonal resources such as naming and vocatives) and across strata (accents in phonology, grammatical variation and discourse semantic style)' (Martin 2010, p. 24). The theory refers to the domain of meaning making associated with forging social bonds as interpersonal meaning (for a more detailed overview of SFL, see the next chapter). It is interpersonal meaning that builds and sustains online social networks. The language used in social media, particularly microblogging, is under significant interpersonal

pressure. On the one hand, it is deployed in a modality where interpersonal meanings that might otherwise be expressed paralinguistically must be expressed via other means.[7] On the other hand, it is bound by the need for linguistic economies arising from the character constraints imposed on microposts. Page (2011) has noted how the resources used to signal interpersonal bonding in social media vary according to the nature of the web genre in question.

In adopting a social semiotic approach, I look both qualitatively and quantitatively at the ways interpersonal meaning co-patterns with ideational meaning in texts. The theoretical basis of this approach is the concept of 'coupling' introduced by Martin (2000a) and taken up by Zhao (2010, 2011) and Zappavigna and colleagues (2008) for looking at textual relations:

> **Coupling** concerns the temporal relation of *'with'*: variable *x* comes with variable *y*. To put it another way, it is the relation formed between two semiotic elements at one given point in time within the logogenetic timeframe. Coupling can be formed between metafunctional variables (e.g. ideational and interpersonal), between different semiotic resources (e.g. image and verbiage) and across strata (e.g. semantics and phonology). (Zhao 2011, p. 144)

This perspective is adopted in tandem with concepts drawn from corpus linguistics of collocation and clustering (in particular n-grams; that is, word-level sequences) as a means for exploring syntagmatic relationships in texts. Coupling is also central to the view of community and affiliation proposed in this book. Zhao (2011) argues that cultures incline toward stable coupling patterns when viewed from the perspective of a particular time frame (e.g. a historical perspective) and that a set of stable coupling patterns can theoretically be described, noting that coupling is a process and that cultures change over time. This modelling perspective is akin to Bakhtin's (1986) work on speech genres as stable patterns of utterance. To this perspective I add the notion developed within complexity theory that cultures are complex adaptive systems[8] and that, as self-organizing ecosocial systems, they look different depending upon the timescale from which they are viewed (Lemke 2000). Adopting these two perspectives, we can theorize that the overlapping bounds of communities of microbloggers will incline towards stable coupling patterns. These coupling patterns will have high dimensionality, and therefore the social semiotician has two options: using some form of expert knowledge of the culture to isolate particular couplings to use as markers for particular communities, or adopting a text of visualization strategy that can reveal patterns which manual analysis and the human gaze of the analyst cannot perceive directly (Zappavigna 2011a).

In this book I focus on couplings of ideational and interpersonal meanings in memes, slang, humour and political discourse in microblogging in order to explore how these textual relations realize social relations, particularly the shift toward searchable talk as a bonding strategy. In so doing, I draw on an additional, complementary perspective that also uses the concept of coupling as its basis: Knight's (2010 a,b,c) model of offline affiliation. This theory, developed through analysis of conversational humour, describes 'communal identity' as discursively negotiated in text (Knight 2010c, p. 43). It applies Martin and White's (2005) framework of evaluative language to understand how communities form as people rally around, defer or reject different values construed in language (Knight 2008). According to Knight, social bonds are negotiated via couplings of ideational and interpersonal meaning in discourse. She comments on these couplings as follows:

> Specifically, couplings realize bonds of value with experience linguistically, as bonds are on a higher order of abstraction in the socio-cultural context. We discursively negotiate our communal identities through bonds that we can share, and these bonds make up the value sets of our communities and culture, but they are not stable and fixed. (Knight 2010c, p. 43)

This book will employ coupling theory to think about processes of affiliation with searchable talk as ongoing negotiation of couplings in text. From a theoretical perspective focusing on textual patterns to explore community is in line with the principle of 'emergence' from systems theory, namely, that complex patterns arise from interactions of simple elements (such as coupling) (Lewes 1875). The aim is to avoid imposing a predetermined structure or hierarchy of organization on the patterns that become visible.

Due to the online channel of microblogging, different practices of negotiation are in play to those seen in face-to-face casual conversation of the kind Knight explores. For example, users may be 'familiar strangers' who share common interests and attributes but do not necessarily know each other (Agarwal et al. 2009), and they may not even engage in direct exchanges. Instead users may be involved in communal performances, such as hashtagging, mentioned earlier in this chapter. For instance, applying Knight's perspective to the tweet introduced at the beginning of this book illuminates the hashtag as the ideational target of the values expressed. In other words, expressing support for the victims of the Christchurch earthquake involves a coupling of positive attitude with the ideational topic realized in the hashtag (i.e. eqnz). Chapter 5 will show how these type of tags are used to label the ideational meanings which users axiologize around, so forming ambient communities of value.

The structure of this book

The chapter which follows will explain why linguists should be interested in social media such as microblogging services as natural language corpora. It details HERMES, a 100 million word Twitter corpus that forms the major dataset used in the book and explains the methodology used: a corpus-based discourse analysis that makes use of both quantitative corpus analysis and close discourse analysis informed by SFL. The chapter provides an overview of the major issues facing researchers when using and managing internet language data.

Chapter 2 gives an overview of the essential features of microblogging and an introduction to the dominant kinds of research in this area. The chapter also begins to explore some of the dominant linguistic patterns in HERMES. Following on from this, Chapter 3 looks in detail at the kinds of evaluative language used in microblogging. Evaluative patterns in HERMES are explored by drawing upon appraisal theory (Martin and White 2005) and corpus-based theories of stance and emotion (Bednarek 2006). The aim is to show that people use Twitter and other microblogging services to share their experiences and enact relationships rather than to simply narrate the mundane details of their activities, as has been claimed in the popular press.

The middle chapters of the book deal with a number of specialized corpora of microposts employed in following case studies:

- Ambient affiliation – An examination of hashtags (e.g. #Obama) used to mark the topic of tweets in a process of 'ambient affiliation', whereby people sharing associated values bond around these user-defined topics.

- Social media memes – An investigation of the function of phrasal template memes used for affiliation (e.g. im in ur [noun] [present infinitive verb] ur [noun]).

- Social media slang – An exploration of how slang (e.g. noob) functions in social media to invoke solidarity.

- Social media humour and fail – A study of how the fail meme (e.g. epic fail, full of fail, bucket of fail.) is used for bonding via internet humour.

- Political discourse – A study of evaluative language in microposts related to politics (e.g. a corpus of tweets containing the name Obama, collected in the 24 hours following the 2008 US presidential election).

The theme tying these studies together is 'ambient affiliation': how social bonds are realized in language, in particular, the emerging searchable talk of microblogging.