

‘I just found your blog’. The pragmatics of initiating comments on blog posts.

Ursula Lutzky, Vienna University of Economics and Business

Matt Gee, Birmingham City University

Abstract

In computer-mediated communication, the medium of blogs is typically viewed as consisting of posts composed by blog authors and comments which may be left by their readers. This study explores the relationship between these constituent parts of blogs and investigates the pragmatic ties that are established between blog posts and comments. The focus is on the preface position in comments, that is the very first position at the onset of a comment, to discover how they are generally introduced and which specific linguistic constructions are used to initiate them. The aim is to uncover how speaker changes – from blog author to commenter – are signalled linguistically, in addition to the blog specific metadata provided by the interface (e.g. the username or time stamp), and which pragmatic means are used to develop interpersonal relationships between users. Results show that the preface position of blog comments is fertile ground for the occurrence of expressive speech acts with commenters often initiating their comments by thanking or complimenting blog authors, which opens up further opportunity for the study of speech acts in large corpora.

Keywords: corpus pragmatics, prefatory expressions, speech acts, computer-mediated communication, blogs

1. Introduction

The widespread medium of blogs is generally regarded as comprising two main constituent parts: posts which are composed by blog authors and comments which may be left by the readers of these posts. While the former are an essential defining criterion of blogs, which are updated with new posts at regular intervals, comments are usually of an optional nature. Nevertheless, both posts and comments are today typically associated with blogs and it is therefore important to uncover the interplay between these different components and to explore the pragmatic ties that exist between blog posts and their associated reader comments, which have not been studied extensively from a linguistic perspective to date (but see e.g. Hoffmann 2012). This study investigates the opening sequences of words in comments on blog posts to gain further insights into the linguistic constructions used, the textual and interpersonal relations established and the interactive potential of the medium of blogs.

The data chosen for this study comes from the *Birmingham Blog Corpus* (BBC), a diachronically-structured collection of blogs covering the period 2000-2010 and totalling 600 million words. Our analysis is based on the Blogger and WordPress sub-corpus of the BBC (<http://www.webcorp.org.uk/blogs>), in which blog posts and their associated reader comments were identified with high accuracy, thus facilitating our study. The commenting feature of blogs provides a framework that allows for interaction between blog authors and readers or between individual readers, which makes this sub-corpus amenable to the analysis of pragmatic phenomena such as speech acts (cf. Searle 1976, 1979).

The aim of this study is to investigate the introductory signals of blog comments which linguistically mark a change in speaker, from blog author to commenter, and which are separate from the metadata usually provided by the blog interface, such as the username of the commenter or the timestamp indicating when a comment was published. In particular, our study focuses on the pragmatic means appearing in preface, or very first, position of comments to discover the interactional patterns that exist between blog users. The methodological approach used, which not only restricts our focus to a specific medium but further confines it to a specific position of attestation in the data, facilitates this analysis and has the potential to reveal patterns of speech acts, that is linguistic constructions with a specific illocutionary force, such as that of an apology or a request, and even “hidden manifestations” of speech acts (Kohnen 2007) in large corpora, as will be shown below. Thus, our study also contributes to the ongoing endeavour of enhancing ways of accurately identifying and studying speech acts in big data collections (see also Lutzky and Kehoe 2016, 2017a&b).

Our article begins by taking a closer look at the medium of blogs and its interactive nature in Section 2, including the development of blogs, the importance of comments to the medium, as well as previous insights into pragmatic differences between posts and comments. In Section 3, we discuss the preface position in blog comments and its crucial role in the context of blogs, where turns at talk may not always appear adjacent to each other. We introduce the distinction between speech external metadata that characterises blogs and speech internal perspective shifters (see Moore 2011) which signal the beginning of a new direct turn at talk and may establish cohesive and pragmatic ties to preceding turns. The analysis of our sub-corpus is presented in section 4 and moves from discussing single words appearing in initial position in blog comments (e.g. forms of address, greetings) to three- and five-word clusters. By expanding the window in this way, we are able to uncover longer interactional constructions functioning as different types of speech acts. In fact, in addition to rather general uses of speech

acts, we find several medium specific uses, which shows how speech acts may adapt to an online medium such as blogs (cf. Verschueren and Brisard 2002).

2. Blogs as an interactive medium

Blogs are a means of computer-mediated communication that has evolved significantly since the first use of the term weblog – later shortened to blog – to refer to a log or collection of links to noteworthy websites (Garden 2011: 485; see also Blood 2002). While the established definition of blogs today refers to them as online journals or diaries (see e.g. OED s.v. *blog n.*), research has shown that it is in fact already outdated and confirmed the increasingly diverse nature of blogs (see e.g. Kehoe and Gee 2012). That is to say that in addition to blogs which are regularly updated by single authors sharing personal reflections on their lives and interests, there are many further uses to which the blog format is put, including, for instance, large corporate blogs with multiple contributors and vast audiences (see e.g. Puschmann 2010, 2013). Given their variety of uses, blogs are today regarded as an online medium, rather than a genre (see e.g. Herring et al. 2005, boyd 2006, Kehoe and Gee 2012).

The facility for comments was introduced in 2002 and has since come to be typically associated with blogs. It allows readers “to actively engage in an exchange with the blogger and/or other readers by writing one or several comments” (Bolander 2012: 1609). Consequently, the comments feature turns blogs into “an online space where discussion and debate can take place” (Garden 2011: 485) and contributes to viewing the medium as being interactive and dialogic in nature. Commenters do, however, not have to react or respond to a post immediately as this asynchronous medium allows for a theoretically infinite time lag between post and comment, thus creating a degree of modulated interactivity or “interaction-at-one-remove” (Nardi et al. 2004: 46). Additionally, blog comments are generally of an optional nature: blog authors have the option of disabling the comment function, while commenters can choose whether they want to actively participate by contributing to an interaction or not. That is to say that blogs allow for interactivity but they do not inherently request it, unlike other online media such as email or chatroom conversations.

Nevertheless, previous research has shown that comments play a significant role in blogs. Kehoe and Gee (2012) studied reader comments as an aboutness indicator in blogs and demonstrated the importance of comments to the medium. When compiling the BBC, they created a 181 million word sub-corpus downloaded from the WordPress and Blogger hosting

sites¹ in which they accurately separated blog posts from comments. Of the 181 million words, comments constitute 48% (86m tokens) and thus account for a significant proportion of the textual content of blog sites, comprising almost as many tokens as posts, which make up the remaining 52% (95m tokens). Additionally, they found that 78% of posts have at least 1 comment, with a blog post receiving 10 comments on average, and that all the comments on a page, on average, contain more text than the post itself (498 tokens versus 427). On the basis of a large corpus comprising a variety of different types of blogs, Kehoe and Gee (2012) therefore showed that the comment facility takes up a prominent position in the blogosphere and provides access to interactional material.

Studying a comparatively smaller corpus sample, Hoffmann (2012), on the other hand, observed that the interactive potential of blogs may not always be fully exploited by its users. He found responsive exchanges between blog authors and commenters to “barely exist” (Hoffmann 2012: 199) in his Augsburg Blog Corpus, comprising 10 personal blogs and a total of 123,242 words. That is to say that one-to-one interaction does not frequently form part of his sample of blogs. Although generally rare, when commenters engage with a blog post, they commonly respond “in an acknowledging and appreciative manner”, which entails a higher incidence of expressive speech acts through which “users express their (usually positive) evaluation toward one of the focal topics or elements introduced in the related entry” (Hoffmann 2012: 200-201).

Concerning speech acts, comments have also been found to differ from blog posts in their use of specific pragmatic features. When investigating the speech act of apologising in blogs, Lutzky and Kehoe (2017a) noticed that the Illocutionary Force Indicating Device (IFID) *sorry* showed a significantly higher density of attestation in the comments of the BBC’s WordPress and Blogger sub-corpus (537.53 per million words) when compared to posts (141.15 per million words). At the same time, *sorry* has different collocates in these two blog sections, collocating with general terms of address such as *guys*, *folks*, or *everyone* in posts, but appearing frequently with the pronoun *you*, terms of endearment, such as *hon/hun* or *dude*, and with several intensifiers such as *so*, *sooo*, *very*, or *incredibly* in comments. These findings suggest that *sorry* serves different functions in blog comments, being attested mainly in expressions of sympathy rather than having the illocutionary force of an apology as in posts. In the following analysis (see Section 4), we will discuss which speech acts tend to introduce

¹ WordPress: <http://www.wordpress.com/>; Blogger: <http://www.blogspot.com/>

blog comments in the BBC sub-corpus and may thus be regarded as characteristic of this component part of blogs.

3. Preface position in blog comments

The present study of blog data and its interactional nature focuses on the preface position in blog comments. Prefatory expressions have been described as serving “the function of orienting the listener to the following utterance, especially in relation to what has preceded” (Biber et al. 1999: 1074). They thus establish a cohesive link between two turns at talk, with a turn being defined “as everything one speaker says before another speaker begins to speak” (Tsui 1994: 7). In this sense, blog posts and comments pertain to the turn-taking system developed by Sacks et al. (1974) with each of them constituting an individual turn that users take to contribute to the conversation. Each turn, i.e. the post as a longer monologic turn at talk and the comment, can be viewed as a holding of the conversational floor. As Bolander (2012: 1609) points out, “[b]y writing a comment, the active reader co-constructs an interaction consisting of two turns (post + comment), which may, but does not have to extend beyond the dyad of blogger-reader. In this way monologues become dialogues or polylogues”.

Contrary to other types of spoken or written interaction, in the asynchronous medium of blogs, turn adjacency and turn alternation do not follow traditional patterns as a single blog post may receive zero or multiple responses in the form of comments. In addition, a comment may be in response to the initiating turn (i.e. the post) or relate to one or more previously published comments. That is to say that turns may not always appear adjacent to each other and that a comment may be spatially separated from the preceding post or comment(s) to which it pertains, with one or several other comments intervening. By studying the linguistic constructions appearing in initial position in blog comments, we aim to discover which linguistic means are used by commenters to fulfill a textual function, linking back to the relevant preceding turn, but also an interpersonal function, pertaining to the relationship between the commenter and the blog/comment author of the preceding turn (cf. also Bolander 2012 on responsiveness in blog comments).

In the context of blogs, we need to distinguish between prefatory expressions or speech internal perspective shifters and speech external linguistic structures (Moore 2011: 43ff.). The latter relates to the metadata included in standard blog templates which appears outside of the actual blog post or comment and provides information about the users writing or commenting on a blog. This metadata usually marks off the beginning and end of a new post or comment

and it typically includes the blogger's and commenter's usernames as well as timestamps indicating when they published their post or comment. Figure 1 presents an example of a typical blog entry. Here the comments are clearly separated from the post by the '23 comments' section title and each comment includes text external metadata (i.e. username and timestamp) in addition to the text written by the commenter.

Figure 1. A typical blog entry

THE CAP'N AND ALI G

Big thanks to everyone for the wonderful comments yesterday! Wishing all a productive year ahead.

Today, I'm over on Joanna St James's blog, expounding on Cap'n Crunch, Ali G and more. Check it out if you have a second.

And there are lots of great contests going on these days! The lovely Jen Daiker's almost at 1000 -- if you're not already following, get over there and read all about her contest.

Everyone's favourite Alliterative Allomorph is also holding a 400 followers contest.

And over at Candyland, you can donate to help the women and children of Ghana.

Any more contests? Feel free to post about it in the comments! Have a fantastic weekend, everyone!

Posted by Talli Roland at [Friday, November 05, 2010](#)



Labels: [contests](#)

23 comments:



Heather 5 November 2010 at 16:23

Thanks for the excellent links. No contests that I know of but, Harlequin has opened submissions to include YA until Dec. 15th. Their YA line is usually closed to unsolicited (unagented) subs so this is a rare opportunity for YA writers to sub to them!

[Reply](#)



Candyland 5 November 2010 at 16:25

Thanks for the links! And for the mention! I read the interview. You rock:)

[Reply](#)



Talli Roland 5 November 2010 at 16:45

Heather, thanks for the heads up! Sounds like a fab opportunity for YA writers.

Candyland - thanks! :)

[Reply](#)

(<http://talliroland.blogspot.co.at/2010/11/capn-ali-g-and-me.html>, accessed 15/12/17)

This metadata fulfills a function similar to inquit clauses (e.g. *he said, she answered*) and quotation marks in traditional written texts, but it only appears once the post or comment has been published. That is to say that it is not visible to the blog author or commenter while composing their texts. In addition to the metadata, the blog template has a structuring function with blog comments usually separated from the post, for instance by appearing in a separate section labelled 'comments' as in Figure 1.

In contrast to speech external linguistic structures, prefatory expressions or speech internal perspective shifters occur within the text of a comment. They signal the onset of a new speaker turn or a speaker change and may, for instance, take the form of greetings (e.g. *hi*, *hello*) or interjections (e.g. *oh*, *wow*). Several of them have been said to be typical of interactive speech situations and to contribute to the creation of involvement (see Biber 1988; Biber and Finegan 1992, 1997). In addition to greetings and interjections, they include forms such as vocatives (e.g. *Andrew*, *dude*), deictic pronouns (*I*, *you*), spatio-temporal deictics (*here*, *now*), and other pragmatic markers (e.g. *well*). These forms are characteristic of direct speech quotations and have also been referred to as “auditory quotation marks” (Rühlemann 2013: 221) when appearing in initial position, as they may fulfil the same function as quotation marks in writing by signalling the onset of direct speech or a new turn at talk. In the following analysis, we investigate the lexical structures attested at the beginning of blog comments to find out how they are initiated in the presence of additional speech external means and which specific clusters are used to link comments to the relevant preceding turns. In particular, our focus is on the pragmatic means used in preface position and on their illocutionary force.

4. Data and analysis

This study is based on the 181 million word sub-corpus of the BBC which comprises blog posts and comments from the WordPress and Blogger hosting sites and covers the period 2000-2010 (see Kehoe and Gee 2012 for a detailed description of how the BBC was constructed). This sub-corpus includes 222,245 blog posts (95 million words) and their associated reader comments (86 million words) from the two hosting sites. All sub-corpora of the BBC are freely available and searchable through the WebCorp Linguist’s Search Engine (WebCorpLSE) software built by the Research and Development Unit for English Studies (RDUES) at <http://www.webcorp.org.uk/blogs>.

Our analysis begins by taking a closer look at single words appearing in initial position in blog comments, which we then extend to larger constructions by studying three and five word-clusters.² Table 1 provides the top 30 types appearing in preface position in the WordPress and Blogger sub-corpus of the BBC. This wordlist is sorted by frequency so that the very first item in the list, the first-person pronoun *I*, is the most frequent type appearing in this position with 283,732 attestations. Note that the lower-case form *i* takes up position 18;

² In this study, we only investigate the initial position in blog comments. Due to its size, the BBC is not annotated for commenter usernames, which means that we cannot address potential changes of addressee within a comment in our analysis but only study those prefatory expressions appearing in comment initial position.

while this does not represent the correct spelling of the first-person pronoun in Standard English, it nevertheless appears with a considerable frequency in our online data.

Table 1. Top 30 types appearing in preface position in the BBC sub-corpus

	TYPE	FREQUENCY		TYPE	FREQUENCY
1.	I	283,732	16.	My	17,621
2.	Oh	54,836	17.	Your	17,453
3.	Hi	51,881	18.	i	17,170
4.	This	48,340	19.	It	15,440
5.	Thanks	42,965	20.	Hey	15,412
6.	What	41,706	21.	I've	14,405
7.	I'm	38,675	22.	Beautiful	14,297
8.	You	36,668	23.	Well	13,694
9.	Wow	28,781	24.	Happy	13,658
10.	The	28,706	25.	A	12,704
11.	Great	28,236	26.	How	12,282
12.	Thank	24,289	27.	It's	11,993
13.	That	22,890	28.	Good	11,901
14.	Love	22,480	29.	Very	11,527
15.	So	17,910	30.	Congratulations	10,799

The use of the first-person pronoun *I/i*, but also the possessive variant *my*, and the abbreviated forms *I'm* and *I've* in preface position indicates that one common way of initiating a comment on a blog post is to shift the focus to the commenter and foreground the new voice that usually differs from the author of the post, as illustrated by the comments cited in lines 1 and 2 in Table 2. At the same time, the second person pronouns *you* and *your* appear among the top 30 types in initial position; they represent a form of address directed at the author of a blog post or a fellow commenter (see lines 3 and 4 of Table 2) and thus function as speech internal perspective shifters, indicating a change in perspective to the viewpoint of the commenter addressing the previous interlocutor. Greetings, such as *hi* or *hey*, have a similar function of starting an exchange, as the example in Table 2 illustrates. Additionally, the words most frequently introducing comments reveal specific speech acts, with thanking and congratulating (see line 6 of Table 2) being explicitly referred to in Table 1 (see *thanks*, *thank*, *congratulations*). Other words, such as several positively evaluative words (e.g. *great*, *beautiful*, *good*, *happy*, *love*), may also relate to the field of expressive speech acts, as the example in Table 2 shows, and we will discuss this further below. Finally, Table 1 comprises several attestations of pragmatic markers or interjections (e.g. *oh*, *wow*, *well*), which are generally associated with spoken

language and may provide written texts with a more immediate feel, as in the last comment quoted in Table 2.

Table 2. Examples of comments initiated by selected top 30 types

thecrazysheep lady said... 20/10/2010	I don't really think it's *not* pretty, but I do wish it wasn't so brown. I bet it would overdy e nice. That would be something fun to try.
Laughingwolf 10/07/2010, 1:41pm	i've become cynical enough to expect little, hence, am rarely disappointed. of course, one always HOPES for better
Sam said... 15/09/2010, 7:15am	You never fail to make me laugh! Gorgeous looking food as always.xx.
# posted by No Good Boyo 24/09/2008	Your insights into human nature, and Jack Nicholson movies, are humbling.
Anonymous said... 09/02/2007, 5:03am	Hey , nice site. I appreciate all the info here on home loan.
Hobo Divine 9:53pm	Thank you for sharing, you made my day. actually it's nighttime but to say you made my night sounds too weird.
Out on the prairie said... 11/10/2010, 9:17am	Beautiful shots of a gorgeous bird. I have been out over the winter quite a few times.
Sharon said... 07/02/2008, 12:11am	Oh my GAWD Dawn! That stupid AT&T automated thing is enough to make me want to rip someone's arm off and beat the electronic fake man with it.

The results in Table 1 and the examples in Table 2 show that the majority of the top types appearing in initial position in blog comments function as prefatory expressions or speech internal perspective shifters. That is to say that commenters refer to linguistic constructions that are generally associated with direct speech or spoken interaction to initiate their comments and thereby signal a shift in perspective from blogger to commenter. One might think that this is unnecessary given that the page structure and metadata make clear which contributions belong to whom and when they have occurred. However, the metadata (i.e. the username and timestamp) is not usually visible when writing a comment and is only displayed on publication of a comment, which may entail that commenters make use of additional means to mark the beginning of a new turn. In addition, commenters seem to associate the medium of blogs with a conversational style of interaction and may therefore draw on forms or patterns of use that

express involvement or indicate the onset of a new turn at talk in direct speech (see our discussion in section 3).

While studying single words appearing at the onset of blog comments already provides initial insights into the ways in which commenters engage in interactive exchanges with blog authors or other commenters, it is through extending the range to clusters of words that further details are revealed. Thus, we investigated the top 20 three-word clusters occurring in preface position in the comments (see Table 3) and found that several of the words we encountered in Table 1 also cluster in tri-grams. For example, the most frequent cluster *I love the* combines the word most frequently introducing comments *I* with the expressive verb *love*, which are ranked first and fourteenth respectively in Table 1. Variants of this cluster appear in the form of *I love your*, *I love this*, *I love that*, and *I would love* in Table 3.

Table 3. Top 20 three-word clusters in preface position in the BBC sub-corpus

THREE WORD CLUSTERS				FREQUENCY
1.	I	love	the	10,267
2.	This	post	has	9,318
3.	Thanks	for	the	7,594
4.	Thank	you	for	7,486
5.	I	love	your	5,673
6.	What	a	great	5,423
7.	I	love	this	4,908
8.	This	is	a	4,164
9.	I	am	so	3,910
10.	What	a	beautiful	3,653
11.	I	like	the	3,542
12.	What	a	wonderful	3,525
13.	Thank	you	so	3,219
14.	I	agree	with	2,968
15.	What	a	lovely	2,759
16.	I	love	that	2,753
17.	Thanks	so	much	2,588
18.	I	would	love	2,522
19.	Sounds	like	a	2,280
20.	I	have	a	2,275

Furthermore, Table 3 includes four variants of the cluster *what a* plus adjective (*great*, *beautiful*, *wonderful*, *lovely*). Similar to the clusters containing the verb *love*, these tri-grams convey a positive observation at the beginning of a blog comment, as the following sentences from our sub-corpus illustrate.

- (1) **What a wonderful** thing to have.
- (2) **What a lovely** blog featurette so many wonderful places to go visit!
- (3) **What a great** post!
- (4) **What a lovely** idea to display your perfect little piece on that perfect little book.
- (5) **What a beautiful** gift and the colours are very Scandinavian.
- (6) **What a great** place for the fair!
- (7) **What a beautiful** quilt that already looks like a family heirloom.
- (8) **What a wonderful**, magical Mum you are!

These examples show that the cluster *what a + positively evaluative adjective* may signal two types of speech acts. On the one hand, it is used to introduce the speech act of congratulating, as in the first example, which continues with the words *Congrats to you*. On the other hand,

this construction may indicate the attestation of a compliment, which Manes and Wolfson (1981: 116) define very broadly as “expressions of positive evaluation”. Our data shows that commenters may compliment a blog author on a post or feature(tte) (see examples (2) and (3)), a specific creation, such as a gift or quilt (see examples (5) and (7)), and even on their personality or social role, such as that of being a mother (see example (8)). Indeed, this cluster (i.e. *what (a) ADJ NP*) is one of the nine patterns of complimenting identified by Manes and Wolfson (1981) in their study of compliments in American English, which they based on the diary method and involved researchers observing and noting down compliments in everyday contexts.

Additionally, Table 3 comprises examples of the speech act of thanking. It lists the clusters *thanks for the*, *thank you for*, *thank you so* and *thanks so much* among the top 20 three-word clusters used to introduce blog comments in our sub-corpus of the BBC. Consequently, one may note that the preface position in comments is frequently occupied by examples of expressive speech acts (cf. Searle 1976, 1979), which establish a relation between speaker and addressee by, for instance, expressing gratitude or praise. This is in line with Bolander’s observation (2012: 1614) that ‘the paying of compliments’ and ‘the voicing of thanks’ are among the central acts performed in blogs, as her study of a corpus of personal blogs revealed.

Finally, the cluster *I agree with* is one of the top 20 tri-grams introducing blog comments (see Table 3). In fact, the verb *agree* is significantly more frequent than its antonym in our data. When searching the complete comment sub-section of our corpus, not restricting searches to initial position, we found that the cluster *I disagree* occurs 4.32 times per million words; on the other hand, when excluding negative constructions such as *not sure I agree*, *don’t think I agree*, *can’t say I agree*, *no way (do) I agree* etc., the form *I agree* occurs 131.25 times per million words. It is, thus, even more frequent than any single swearword in blog comments (see Lutzky and Kehoe 2016).

When extending the window to five-word clusters (see Table 4), a similar picture emerges as for three-word clusters, in particular with regard to the speech acts attested. However, in contrast to the examples discussed above, five-word clusters also reveal several medium-specific uses. The most frequent of these is the automated message *This post has been removed*, which occurs in our data whenever the comment author or the blog administrator deleted the original comment. As it does not reveal any insights into language use in the preface position of blog comments, we will not discuss it further but focus on other medium-specific attestations in the following analysis.

Table 4. Top 20 five-word clusters in preface position in the BBC sub-corpus

FIVE WORD CLUSTERS						FREQUENCY
1.	This	post	has	been	removed	9,260
2.	Thank	you	so	much	for	1,762
3.	I	know	what	you	mean	513
4.	I	follow	you	on	twitter	478
5.	I	just	found	your	blog	438
6.	I	would	love	to	win	390
7.	Thanks	so	much	for	the	387
8.	I	follow	you	on	Twitter	387
9.	Thanks	for	stopping	by	my	364
10.	I	love	the	idea	of	346
11.	Thank	you	for	sharing	this	310
12.	I	can't	wait	to	see	310
13.	Sounds	like	you	had	a	308
14.	I	love	the	way	you	298
15.	Thank	you	for	sharing	your	294
16.	I	am	so	glad	you	286
17.	This	is	a	great	post	275
18.	Thank	you	all	for	your	270
19.	All	I	can	say	is	264
20.	I	can't	wait	to	read	262

As Table 4 shows, we find several attestations of the speech act of thanking among the top 20 five-word clusters in comment initial position: *Thank you so much for*, *Thanks so much for the*, *Thanks for stopping by my*, *Thank you all for your*, *Thank you for sharing this*, and *Thank you for sharing your*. While the first two examples, which are intensified by *so much*, are more general in nature, the remaining examples exhibit medium-specific features. Thus, blog authors, for instance, thank their readers for stopping by their blog; in fact, 85% of all attestations of *Thanks for stopping by my* have *blog* appear in first position to the right of the cluster. Other examples comprise the words *site*, *page*, *place*, or *world* in this position or include a modification of the noun *blog* in the form of an adjective such as *wee*, *insane* or *new* (e.g. *Thanks for stopping by my wee blog*) or another noun such as *art*, *garden*, *photo*, or *recipe* (e.g. *Thanks for stopping by my photo blog*). These five-word clusters can be clearly attributed to the blog authors, who are actively engaging with their audience by thanking their readers for visiting their blogs, as in examples (9) to (11).

- (9) **Thanks for visiting my blog** ... yours is gorgeous and really intensively useful, nice work!

- (10) **Thanks for visiting my blog** and following! I'm now following you ... and hoping to get some design ideas for my blog.
- (11) **Thanks for visiting my blog** and commenting. It's so nice to have visitors. I hope you will return often. I certainly will return to yours after that wonderful story. What an amazing adventure! You are a good writer. Can't wait to read more.

As example (9) illustrates, the speech act of thanking is used by blog authors, for instance, to introduce a compliment on their readers' own blogs. At the same time, it may be combined with a reference to the medium-specific feature of following a blog (see example 10), which is an indication that readers are interested in receiving notifications of new blog posts in the future and thus intend to engage further with the blog and its materials. Finally, blog authors may thank their readers not only for visiting but also commenting on their blog, as example (11) shows, where the blog author initiates a longer turn at talk, indicating that they would want the reader to return and complimenting them on their own blog (cf. the construction *can't wait to* discussed further below).

On the other hand, the cluster *Thank you all for your* mainly refers to the comments left on a blog post. Thus, one finds the word *comment* in first position to the right of the cluster or variations of this form that have a similar meaning, such as *words, wishes, support, encouragement* or *thoughts*. These concepts also collocate with different evaluative modifiers, such as *lovely* or *kind*, as examples (12) to (15) from our corpus illustrate.

- (12) **Thank you all for your** comments.
- (13) **Thank you all for your** lovely comments.
- (14) **Thank you all for your** encouraging and kind comments.
- (15) **Thank you all for your** kind, understanding, and encouraging comments.

While the above clusters pertain primarily to the author of a blog who thanks their readers for visiting their page or leaving a comment, the remaining two medium-specific clusters refer back to the blog post or a previous comment. That is to say that the clusters *Thank you for sharing this* or *Thank you for sharing your* are typically written by the commenters rather than the author of a blog post and stress one of the main functions of blogs: the sharing of information. This can take different forms, as the randomly extracted examples (16) to (24) show.

- (16) **Thank you for sharing your** experiences!
- (17) **Thank you for sharing your** thoughts here.

- (18) **Thank you for sharing your** life, I love reading about your family and precious Noah.
- (19) **Thank you for sharing your** thoughts about it and I learn a lot from you.
- (20) **Thank you for sharing your** space with me Sherry
- (21) **Thank you for sharing your** talent with us!
- (22) **Thank you for sharing your** insights, I always enjoy visiting your blog!
- (23) **Thank you for sharing your** wonderful treasures with us (and have fun!! xo).
- (24) **Thank you for sharing your** corner of heaven ... luxuriate in it.
- (25) **Thank you for sharing your** finding and thoughts on this.

While the above examples of five-word clusters illustrating the speech act of thanking indicate that the blog participants interact with each other, one of the limitations of the current large-scale corpus study is that the specific linguistic connections between the constituent parts of blogs (i.e. between a post and a comment or a comment and a reply to a comment) cannot be discussed in detail. This does, however, not minimise the importance of these ties, which are established, for instance, through the use of terms of address (see e.g. the example in Figure 1) or other means of lexical cohesion, and calls for further qualitative studies of smaller data samples, such as the study of cohesive ties in Facebook comments by Frobenius and Harper (2015).

As with the three-word clusters discussed above, Table 4 also comprises several clusters that signal examples of the speech act of complimenting, such as *I love the idea of, I love the way you* or *This is a great post*. At the same time, five-word clusters reveal insights into two further categories of compliments, which we will refer to as ‘compliments for future action’ and ‘compliments upon discovery’. The first category is illustrated by the clusters *I can’t wait to see* and *I can’t wait to read*. Both of them point to a time in the future when a blog author will have accomplished a specific task and are illustrated by examples (26) and (27).

- (26) **I can’t wait to see...** all your new vlogs
 more of your photography/work, more photos/pictures...
 the finished book/product/room/work...
 your book/design/new hairstyle/new look/painting...
- (27) **I can’t wait to read...** more, more of your stories, more posts from you...
 this/your book, this/your novel, book 2, part three...
 your answers/list/review/updates...

Thus, commenters use the phrase *I can't wait to see/read* to refer to a future creation (such as books, photos, posts, updates) and positively evaluate it by expressing their eager anticipation of its arrival or completion. As examples (26) and (27) indicate, this evaluation is based on readers' perception of bloggers' previous posts/stories/photography/etc. However, it should be added that not all attestations of these two clusters initiate compliments. They may also introduce general statements of praise referring, for instance, to a new book that may have been reviewed in a literature blog but was not published by the blog author themselves. These examples still fall within the realm of positive evaluation but they do not have a specific recipient, i.e. they are not "addressed to any one, implying or involving praise" (OED: s.v. compliment, *n.*), and can therefore not be classed as compliments.

The second category of compliments that we identified among five-word clusters are 'compliments upon discovery'. These compliments are linked to having come across a blog for the first time and are introduced by the cluster *I just found your blog*, which is attested 438 times in our comment data. While the cluster cannot be classed as a compliment in its own right, studying concordance lines of the examples in question shows that it is frequently followed by a compliment, as illustrated by the randomly extracted examples (28) to (37).

- (28) **I just found your blog** today, what a wonderful site!!
- (29) **I just found your blog** today, and then this post on Errol LeCain.
- (30) **I just found your blog** tonight and am so encouraged by your gift of words!
- (31) **I just found your blog** and was blown away ...
- (32) **I just found your blog**, your feelings are totally normal!!
- (33) **I just found your blog** and I really love your blog. I'm a follower now.
- (34) **I just found your blog**. It's fabulous!
- (35) **I just found your blog** and I'm excited.
- (36) **I just found your blog** and love many things about it.
- (37) **I just found your blog** and I'm so glad because we have some similar struggles.

Of these ten examples, seven may be classed as compliments expressing praise for the blog(ger) by referring to the wonderful site (28), the blogger's gift of words (30), or using positively evaluative expressions to describe one's impression of the blog (see (31) and (33) to (36)). It is a by-product of the medium studied that this is a "speaker-centred view of speech acts", where not each turn at talk will necessarily receive a reply (De Felice 2013: 168). Contrary to Válková (2013: 49), we would therefore argue that compliments should not necessarily be

treated as speech act sets as they may not entail a reaction from the addressee, depending on the context and medium of occurrence.

In our data, the cluster *I just found your blog* in combination with the comment initial position functions as a signal uncovering the otherwise “hidden manifestations” (Kohnen 2007) of the speech act of complimenting. According to Kohnen (2007: 140-141), one of the main problems in studying speech acts is that we usually do not have access to a full inventory of the forms in which a speech act may be attested and it is therefore particularly difficult to search for them in a corpus. While a corpus-based speech act study may focus on typical examples, such as *sorry* in the case of apologies, it will not be able to provide a complete picture of the speech act as it will naturally miss any insights into its attestations not using these routinized forms. In order to tackle this problem and gain access to further manifestations of a specific language function, Kohnen (2007: 141) suggests that analyses are to be based “on comparable text types or genres”. We would argue that, in addition to that, focusing on a particular position of attestation, such as the initial position in blog comments, offers new possibilities of uncovering forms functioning as a particular speech act. In our data, *I just found your blog* is a medium-specific means of initiating compliments which, together with the remaining clusters we have identified for this speech act, allows for the study of complimenting in a large corpus and renders a speech act that is otherwise difficult to search for more easily accessible.

Of the three remaining examples, (37) – while not conveying praise – shows appreciation for the blog and like (32) it addresses shared experiences (of feelings or struggles). A more neutral example introduced by *I just found your blog* is (29); it is only in the next sentence that the commenter also refers to common ground that they share with the author of the post: *I too had this book as a child and find myself to be completely obsessed with him as an adult*. Similar references to mutual experiences are conveyed by the five-word clusters *I know what you mean* (rank 3 in Table 4) and *I am so glad you* (rank 16). Thus, as discussed with regard to several examples above, the initial clusters in blog comments foreground one of the main functions of blogs: sharing. While this may be first and foremost the sharing of information, the comment function of blogs furthermore allows the expression and recognition of shared experiences and the establishment of rapport on the basis of common ground.

Two further clusters in Table 4 resemble the cluster *I just found your blog*. They express that a commenter is following a blogger on the micro-blogging platform Twitter, which is spelt either with a capital or lower case *t* (*I follow you on T/twitter*), and together they make up 865 attestations in comment initial position in the BBC sub-corpus. While one may be tempted to

assume that these clusters have a similar function to the one discussed above, a close analysis of their occurrence in context shows that they are primarily associated with a specific activity on blogs: the giveaway. That is to say that they mainly relate back to a blog post in which a giveaway was announced and blog authors encouraged their readers to follow them on Twitter; if they subsequently also left a comment on their post, they would have the chance to win a prize, such as the copy of a book or some artwork. This type of prize draw allows bloggers to increase the number of their followers and results in an increased frequency of comments such as *I follow you on Twitter as happysnappy16*. At the same time, in the context of giveaways, one finds individual attestations where *I follow you on T/twitter* is used to introduce a compliment and thus serves a function similar to *I just found your blog*. Examples of these rather few occurrences are given in (38) and (39). Both examples appear in the context of a current or upcoming giveaway and therefore imply that the reader would love to win the prize, which in (39) is a rabbit tarot deck and several other rabbit themed rewards.

- (38) **I follow you on Twitter** and Ive just come over to read your blog. Love your new posts so I'm going to go back and read some of your earlier posts x x.
- (39) **I follow you on Twitter** and love your bunnies.

5. Conclusion

In this study, we investigated the preface position in blog comments in a 181 million word sub-corpus of the BBC to gain further insights into the interactive nature of blogs. By focusing on the linguistic constructions used in initial position in comments, we were able to discover how they are pragmatically tied to the preceding blog post or comment(s). Our results show that the top 30 types introducing comments comprise many speech internal perspective shifters, which are forms that indicate a change in speaker and the beginning of a new turn. That is to say that in addition to the speech external metadata that typically characterises blog templates, users draw on specific linguistic means that allow them to establish textual links as well as interpersonal relations. This is done through the use of greetings such as *hi* (see rank 3 in Table 1) or interjections such as *oh* (see rank 2 in Table 1). Additionally, we found pronouns (e.g. *I*, *you*, *your*), positively evaluative words (e.g. *great*, *beautiful*, *happy*) and illocutionary force indicating devices (e.g. *thanks*, *congratulations*) to appear among the most frequently used prefatory expressions in our data.

When extending the window to three- and five-word clusters, we discovered that the preface position attracts the attestation of expressive speech acts. In particular, our study revealed that the speech acts of complimenting and thanking have a tendency to introduce blog comments and readers thus tend to initiate their comments by establishing rapport with fellow bloggers. While some of the clusters identified are more general in nature (e.g. *What a + positively evaluative adjective*, *Thanks for the*, *Thank you for...*), we were able to uncover several medium-specific uses, especially among the most frequently occurring five-word clusters. They show how blog authors actively engage with their audience and reply to comments that were left on a post by thanking readers for visiting their page or leaving a comment (e.g. *Thanks for stopping by my*, *Thank you all for your*). At the same time, readers of posts express their appreciation for the information shared (e.g. *Thank you for sharing this / your...*), thus referring to one of the main functions of blogs.

In the category of compliments, we found the cluster *I just found your blog* to signal a ‘compliment upon discovery’. This example is the fifth most frequent five-word cluster in comment initial position and typically introduces a compliment directed at the author of a blog post. It is a medium-specific means of initiating compliments in blogs and in Kohonen’s terms a ‘hidden manifestation’ (2007) that our analysis helped to uncover. In the medium of blogs, compliments will not always receive a reply from their recipients as the interactive potential of blogs is not necessarily exploited by both blog authors and commenters. This compares to other

asynchronous forms of digital discourse, such as Facebook, where “the absence of a response ... is not considered either rude or a communication failure but rather a natural effect of the asynchronous character of the channel” (Maíz-Arévalo 2013: 63). Consequently, in contrast to previous studies of synchronous media, such as face-to-face conversations, when studying compliments in blogs, we decided to focus on their illocutionary force rather than their perlocutionary effect and to treat them as a speech act and not a speech act set (cf. Válková 2013).

As previous studies in the field of corpus pragmatics have shown, “[s]peech acts are not readily amenable to corpus-linguistic investigations” (Jucker et al. 2008). This is mainly due to the fact that they are defined by their function rather than their form and consequently difficult to search for using corpus linguistic software. Adopting a corpus-driven perspective, we discovered the speech acts of thanking and complimenting to repeatedly initiate blog comments; both of these speech acts are regarded as more formulaic in nature and are associated with “a very small number of lexical items and syntactic patterns” (Holmes 1988: 452; see also Manes and Wolfson 1981). Despite their formulaic nature, however, searching for them in corpora is problematic and the solutions discussed to date (e.g. manual searches or pragmatically tagged corpora, see e.g. Jucker et al. 2008) are only feasible to a certain extent and mainly suitable for smaller corpora. Our study has introduced a new methodological approach that helps streamline the study of speech acts in large corpora. This involved restricting our focus not only to a particular medium (cf. Kohnen 2007, 2008 and his genre-based approach), but also to a specific structural point in the medium of blogs, the preface position of comments. Instead of searching for constructions that are typically associated with a specific speech act, we thereby allowed the data to reveal speech act patterns. This has the advantage of uncovering medium-specific uses of speech acts, and in our case led to further understanding of the things bloggers do with words.

While this article has opened a new methodological window on the study of speech acts in large corpora and revealed innovative insights into the use of speech acts in blog comments, further research into speech acts in digital discourse is needed. In a study of cohesive relations in a significantly smaller blog corpus, Hoffmann (2012: 214) noted that “the form and function of speech act patterns between blog entries and comments was highly constrained” and limited mainly to assertives and expressives. The results of our study partly support this claim as we found expressive speech acts to occur with considerable frequencies in preface position in blog comments. Future studies will need to expand on our findings in two main ways. Firstly, more

fine-grained analyses of smaller blog samples may account for the specific pragmatic ties between the individual component parts of blogs and the interplay between users adopting different roles in the interaction. Secondly, similar large-scale experiments, in which the position under analysis is fixed, could reveal yet more 'hidden' manifestations of speech acts across a range of computer-mediated communication technologies and thereby uncover the adaptability of speech acts to different online media.

References

- BBC = *Birmingham Blog Corpus*. 2010. Compiled by the Research and Development Unit for English Studies at Birmingham City University. URL: <http://www.webcorp.org.uk/blogs>
- Biber, Douglas, 1988. *Variation Across Speech and Writing*. Cambridge University Press, Cambridge.
- Biber, Douglas, Finegan, Edward, 1992. The linguistic evolution of five written and speech-based English genres from the 17th to the 20th centuries. In: Rissanen, M., Ihalainen, O., Nevalainen, T., Taavitsainen, I. (Eds.), *History of Englishes: New Methods and Interpretations in Historical Linguistics*. De Gruyter, Berlin, pp. 688-704.
- Biber, Douglas, Finegan, Edward, 1997. Diachronic relations among speech-based and written registers in English. In: Nevalainen, T., Kahlas-Tarkka, L. (Eds.), *To Explain the Present. Studies in the Changing English Language in Honour of Matti Rissanen*. Soci t  N ophilologique, Helsinki, pp. 253-275.
- Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, Finegan, Edward, 1999. *Longman Grammar of Spoken and Written English*. Longman, Harlow.
- Blood, Rebecca, 2002. Weblogs: a history and perspective. In: Rodzvilla, J. (Ed.), *We've Got Blog. How Weblogs are Changing our Culture*. Perseus Publishing, New York, pp. 7-16.
- Bolander, Brook, 2012. Disagreements and agreements in personal/diary blogs: A closer look at responsiveness. *Journal of Pragmatics* 44, 1607-1622.
- boyd, danah, 2006. A blogger's blog: exploring the definition of a medium. *Reconstruction* 6(4). Retrieved from <http://www.danah.org/papers/ABloggersBlog.pdf> (last accessed June 2017).
- De Felice, Rachele, 2013. A corpus-based classification of commitments in Business English. In: Romero-Trillo, J. (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013. New Domains and Methodologies*. Springer, Dordrecht, pp. 153-171.
- Garden, Mary, 2011. Defining blog: a fool's errand or a necessary undertaking. *Journalism* 13 (4), 483-499.
- Frobenius, Maximiliane, Harper, Richard, 2015. Tying in comment sections: The production of meaning and sense on Facebook. The self-explicative organization of communication acts on and through Facebook. *Semiotica* 204, 121-143.
- Herring, Susan C., Scheidt, Lois A., Wright, Elijah, Bonus, Sabrina, 2005. Weblogs as a bridging genre. *Information Technology and People* 18 (2), 142-171.
- Hoffmann, Christian R., 2012. *Cohesive Profiling: Meaning and Interaction in Personal Weblogs*. John Benjamins, Amsterdam.
- Holmes, Janet, 1988. Paying compliments: a sex preferential politeness strategy. *Journal of Pragmatics* 12 (4), 445-465.
- Jucker, Andreas H., Schneider, Gerold, Taavitsainen, Irma, Breustedt, Barb, 2008. Fishing for compliments. Precision and recall in corpus-linguistic compliment research. In: Jucker, A. H., Taavitsainen, I. (Eds.), *Speech Acts in the History of English*. John Benjamins, Amsterdam, pp. 273-294.
- Kehoe, Andrew, Gee, Matt, 2012. Reader comments as an aboutness indicator in online texts: Introducing the Birmingham Blog Corpus. In: Oksefjell Ebeling, S., Ebeling, J., Hasselg rd,

- H. (Eds.), *Aspects of Corpus Linguistics: Compilation, Annotation, Analysis*. VARIENG, Helsinki. http://www.helsinki.fi/varieng/series/volumes/12/kehoe_gee/
- Kohnen, Thomas, 2007. Text types and the methodology of diachronic speech act analysis. In: Fitzmaurice, S., Taavitsainen, I. (Eds.), *Methods in Historical Pragmatics*. Mouton de Gruyter, Berlin, pp. 139-166.
- Kohnen, Thomas, 2008. Tracing directives through text and time. Towards a methodology of a corpus-based diachronic speech-act analysis. In: Jucker, A. H., Taavitsainen, I. (Eds.), *Speech Acts in the History of English*. John Benjamins, Amsterdam, pp. 295-310.
- Lutzky, Ursula, Kehoe, Andrew, 2016. *Your blog is (the) shit*. A corpus linguistic approach to the identification of swearing in computer mediated communication. *International Journal of Corpus Linguistics* 21 (2), 165-191.
- Lutzky, Ursula, Kehoe, Andrew, 2017a. "I apologise for my poor blogging": Searching for apologies in the *Birmingham Blog Corpus*. *Journal of Corpus Pragmatics* 1 (1), 37-56.
- Lutzky, Ursula, Kehoe, Andrew, 2017b. "Oops, I didn't mean to be so flippant". A corpus pragmatic analysis of apologies in blog data. *Journal of Pragmatics* 116, 27-36.
- Manes, Joan, Wolfson, Nessa, 1981. The compliment formula. In: Coulmas, F. (Ed.), *Conversational Routine. Explorations in Standardized Communication Situations and Prepatterned Speech*. Mouton, The Hague, pp. 115-132.
- Maíz-Arévalo, Carmen, 2013. "Just click 'Like'": Computer-mediated responses to Spanish compliments. *Journal of Pragmatics* 51, 47-67.
- Moore, Colette, 2011. *Quoting Speech in Early English*. Cambridge University Press, Cambridge.
- Nardi, Bonnie A., Schiano, Diane J., Gumbrecht, Michelle, Swartz, Luke, 2004. Why we blog. *Communications of the ACM* 47 (12), 41-46.
- OED = The Oxford English Dictionary. Online edition. Oxford University Press, Oxford. www.oed.com
- Puschmann, Cornelius, 2010. *The Corporate Blog as an Emerging Genre of Computer-Mediated Communication: Features, Constraints, Discourse Situation*. Universitätsverlag Göttingen, Göttingen.
- Puschmann, Cornelius, 2013. Blogging. In: Herring, S. C., Stein, D., Virtanen, T. (Eds.), *Pragmatics of Computer-Mediated Communication*. De Gruyter, Berlin, pp. 83-108.
- Rühlemann, Christoph, 2013. *Narrative in English Conversation. A Corpus Analysis of Storytelling*. Cambridge University Press, Cambridge.
- Sacks, Harvey, Schegloff, Emanuel A., Jefferson, Gail, 1974. A simplest systematics for the organisation of turn-taking for conversation. *Language* 50 (4), 696-735.
- Searle, John, 1976. A classification of illocutionary acts. *Language in Society* 5 (1), 1-23.
- Searle, John, 1979. *Expression and Meaning. Studies in the Theory of Speech Acts*. Cambridge University Press, Cambridge.
- Tsui, Amy, 1994. *English Conversation. Describing English Language*. Oxford University Press, Oxford.
- Válková, Silvie, 2013. Speech acts or speech act sets: apologies and compliments. *Linguistica Pragensia* 23 (2), 44-57.

Verschueren, Jef, Brisard, Frank, 2002. Adaptability. In: Verschueren, J., Oestman, J.-O., Blommaert, J., Bulcaen, C. (Eds.), Handbook of Pragmatics Online. John Benjamins, Amsterdam.