

BLURRING THE LINES BETWEEN GENRES AND AUDIENCES: INTERACTION IN SCIENCE BLOGS¹

Maria Freddi

Abstract

Science blogs have been attracting the attention of linguists, rhetoricians and communications scholars alike as the discourse of science becomes more and more influenced by new digital media and more scientists engage in the practice of blogging for the purposes of knowledge dissemination and public engagement. The paper analyses writer-reader interaction in a corpus of blogs maintained by individual scientists, considering both posts and comments. The analysis is corpus-driven to the extent that it harnesses corpus linguistic tools for frequency observations to detect language patterns of interaction, but tries to interpret frequency in light of linguistic and rhetorical models of audience engagement in science popularization. The findings confirm a tendency of blogs to exploit all of the linguistic strategies of audience involvement already found in the literature, reader pronouns, questions and the conversational style typical of spoken science communication, testifying to the blurring of genres and audiences.

Keywords

popular and professional science, science blogs corpus, audience engagement, language patterns of interaction

1 Background and aim of present research

The paper is set in the context of linguistic and rhetorical analyses of knowledge dissemination and public engagement with science. It draws from recent literature on the linguistic and rhetorical features that make science blogging online an appealing genre of public communication of science and explores interaction and audience engagement as realised in a small corpus of blogs written by individual scientist-writers.

The paper builds on the argument that frequency-driven analyses are an especially apt way into the blogs dataset to address the question of how scientists interact with their web audiences and how interpersonal dynamics are construed through post-comment threads. It tries to show that the corpus-driven approach, by allowing to trawl the data for what is frequent and thus linguistically significant, might be conducive to communicative and rhetorical strategies typical of the genre, once frequency is analysed in context. The aim therefore is to exploit frequency observations to find traces of interaction and then interpret them in the broader context of online science communication and the changing rhetoric of science.

1.1 Immediacy, individuality and interactivity in science blogs

Both linguists and rhetoricians have been concerned with features of science communication and the new media (e.g. Twitter in Büchi 2017; blogs and wikis in Myers 2010; blogs in Luzón 2011, 2013a, 2013b, in Bondi 2018, and in Mehlenbacher 2019) and have been exploring the genesis and development of a successful genre of science writing – the science blog. Starting with the rhetorical approach to weblogs of Miller and Shepherd (2004, 2009) and Askehave and Nielsen (2005), recent research (e.g. Mehlenbacher 2019) has indicated how science communication on the internet has “blurred the boundaries between the scientific community and the public and between the genres intended for audiences with different degrees of expertise” (Luzón & Pérez-Llantada 2019: 9).

According to rhetoricians Miller and Shepherd (2004, 2009), the origin of this blurring of genres and audiences is to be found in a culture of self-disclosure typical of popular media of the late 1990s, such as reality television and talk-shows interviews, and intensified by the omnipresence of the internet. For the two scholars, blogs fulfil the dual communicative need for self-expression and relationship development, functioning as spaces for personal reflections while at the same time staying open to public scrutiny and critique. As blogs address everybody and no one at the same time, through “personal but anonymous contacts” (as in Myers 2010: 10), what emerges is the blurring of private and public sphere and of mediated and unmediated experience, changing the very idea of reality:

The “reality” offered by blogs is thus a thoroughly perspectival reality, anchored in the personality of the blogger. And although this reality may seem to be “immediate”, (that is, un-mediated), it is, of course, highly mediated. (Miller & Shepherd 2004: 8)

Myers (2010: 10) captures the same sense of individuality, freedom of expression and immediacy, when he writes that blogs are “outlets for the unbridled expression of opinion: comments, suggestions, reviews, outcries, rants”.

1.2 Changing rhetoric of science in blogs

As a special type of personal blogs, science blogs, sometimes also referred to as scholarly blogs (e.g. Puschmann & Mahrt 2012), academic blogs (as in Luzón 2011, 2013a, 2013b), or research blogs (“produced by active researchers who write about their own work” (Mauranen 2013: 9)), can be considered an example of the changing rhetoric of science, in that they have become preferred platforms of science dissemination showing features of argumentation not usually associated

with traditional science genres, e.g. freedom and nonconformity, immediacy and openness, interactivity and proximity to the audience (Beacco et al. 2002, Bell 2012, Puschmann & Mahrt 2012, Luzón 2011, Mauranen 2013, Angler 2017, Gallagher 2018, Mehlenbacher 2019). Indeed, an important communicative function of science blogging is that of mediation between academia and the wider public, facilitating public engagement with science (e.g. Kouper 2010, Mauranen 2013, Puschmann & Mahrt 2012, Luzón 2013a, Mehlenbacher 2019, Reid & Anson 2019), a feature which is characteristic of new digital media at large (Blanchard 2011). Both Puschmann and Mahrt's (2012) survey of scholarly bloggers active on a German platform and Mauranen's (2013) analysis of research blogs point to the academia's opening up to blogs as a new form of communication to educate and engage with a broader public and create a forum for discussion, thus as a form of promoting public understanding of science.

From the perspective of rhetoric of science and science popularization, blogs have facilitated community formation, allowing what Miller and Fahnestock (2013: 2) call the "accommodation of public audiences, given the increasing access by the public to primary scientific genres through online media" and realizing the increasing potential for interaction between scientific and public communities (ibid.: 3 and cf. also Fahnestock 1986 on the concept of accommodating science to different audiences). Miller and Fahnestock's (2013: 2) observation that "[i]t's possible that the Internet has made it easier for such discourse communities to form, since geographic proximity is less necessary" applies especially well to science blog writing that, while being characterised by marked personalization of contents, maintains an outward orientation and plays an important community building function. In the same vein, Mehlenbacher (2019: 135) has spoken of blogs "inhabiting an intermediary space between the internal sphere of science and the external sphere of science popularization". As all of these studies have shown, the rhetorical potential of blogs is manifest. In Myers' (2010: 11) words, "[b]loggers ... have developed some practices of persuasion that, when they work, allow people to keep on writing and interacting...".

1.3 Empirical studies of science blogs

A number of studies have focused on the linguistic realisations of immediacy, individuality and interactivity in science blogs, taking either a qualitative (Myers) or quantitative (Luzón, Bondi) empirical approach. With an interest in opinion making and enactment of interaction, Myers (2010) investigates the linguistic conventions of blogs, the ways bloggers state facts and express opinions and also how they define themselves in relation to other people. His results are relative to

the personal blog; it is therefore interesting to see how far they are confirmed by the dataset analysed here.

Through an analysis of social interaction in academic blogs, Luzón (2011, 2013b) has shown how blogs have a networking and knowledge sharing function more akin to spoken face-to-face discussions than written academic genres, for they “support interdisciplinary, informal and rapid sharing of ideas with scholars (and interested public) anywhere” (Luzón 2011: 536). However, she has also shown that blogs also display features of conversational conflict and disagreement engendered by the anonymity of the written medium. Drawing on Hyland’s (2010) discursive strategies of knowledge popularisation, Luzón (2013a) has also highlighted dialogic involvement and discourse informalisation as two prominent features of bloggers’ style of creating proximity with the readers. Bondi (2018) has studied how renown economists blog on economics and engage with their audiences employing a variety of dialogic strategies that point to a conversational style more typical of spoken interaction than written communication.

Drawing on all of the above discussion, in this paper I describe the typical forms of audience engagement and interaction as observed in a corpus of science blogs. Like in the research summarised above, the focus is on the linguistic realisation of self-expression and relationship development. Unlike the studies that use a coding scheme to analyse the data for some prior classification of functions or indicators of interaction (e.g. Luzón 2011, 2013a, 2013b), the approach of the present paper is data-driven, as no prior scheme is investigated in the data. It consists first, in designing and compiling a corpus of science blogs, and then, in running frequency queries and concordance analyses of patterns that might be revealing of the changing rhetoric of science. Section 2 describes the corpus compilation project in detail, the criteria adopted in the selection of blogs and the concordancing software used for the corpus queries. Section 3 presents the results of the analysis of interaction as observed in the individual blogs sub-corpus, and finally Section 4 discusses the findings further and draws some conclusions.

2 Data and methodology

2.1 The Science Blogs Corpus: Design and construction

The Science Blogs Corpus is comprised of two sub-corpora: one of weblogs written by individual scientists who, despite having an institutional and academic affiliation, maintain a personal webpage where they blog freely about scientific issues of their own interest; the other made of blogs that are representative either

of science magazines and newspapers such as *Scientific American*, *Discover* and *The Guardian*, or of educational institutions, research centres and scientific journals running their own blogging networks, such as *Public Library of Science (PLOS)* or *Nature*. The distinction between the two sub-corpora was inspired by the early discussions on weblogs as a new genre of science communication (cf. Section 1), and meant to highlight the difference between ‘individual, independent’ and ‘institutional’ blogs. Despite being identified by the specialised domain, individual blogs are not confined to a specific theme, rather they are “an outlet for expressing the personal experiences of the writer” (Mauranen 2013: 10), sometimes quite provokingly so, as is the case with Steven Novella’s blog *NeuroLogica*, sampled in the corpus. Instead, institutional blogs cannot be considered the expression of a single blogger, but rather a more collective platform or media outlet, often sponsored by some publishing institution, and therefore closer in kind to science journalism (Angler 2017). The distinction between individual and institutional blogs has, however, created issues of text classification and selection, but it was meant to be suitable for comparison not only across topic or domain, but also tenor, i.e. writer-reader roles, to study correlations between discourse features such as stance, hedging, evidentiality and the type of blog. Ultimately, the individual vs. institutional distinction might help revise the genre taxonomy to which different blogs are ascribed, thus deepening the understanding of the generic status of blogs (cf. on this Mehlenbacher’s 2019 genre analysis of PLOS). In line with this distinction, the individual blogs sub-corpus consists of four different blogs covering different areas within the hard sciences. These are, in alphabetical order, *Genomics*, *Medicine and Pseudoscience* by Steven Salzberg, *Mountain Beltway* by Callan Bentley, *NeuroLogicaBlog* by Steven Novella, and *Skulls in the Stars* by Gregory J. Gbur. The blogs chosen for the institutional blogs sub-corpus were *Discover Magazine*, *Physics Buzz*, *PLOS* and *Science News*.

In selecting the blogs, a number of factors were taken into account including range of topics/domains, activity status and circulation. The disciplinary domains covered belong in the Science, Technology, Engineering, Mathematics (STEM) grouping and comprise physics and astronomy, medicine and health, biology and life sciences, and earth and environmental sciences. The blogs all displayed a high degree of interactivity through the presence of threads of comments, as suggested by studies of online audience and comment functions by Gallagher (2018)² and Walsh (2015)³. One of the individual blogs even has a *Suggest a Topic* tab through which its writer invites his audience to make topic suggestions. As a further guarantee of circulation and interactivity, all of the blogs were active at the time of the texts’ selection, with the individual blogs averaging

two posts per month accompanied by a varying number of comments (reflected in the wordcount). All of the blogs chosen were either among the top-ranking science blogs in Feedspot's "Top 100 Science Blogs Websites & Newsletters to follow"⁴ list, or other reviewing service such as Ata Scientific's "17 science blogs everyone should follow"⁵, or else recommended by some research institution (e.g. the Institute of Physics⁶). The blogs also shared the formal characteristics of the genre described by Hoffmann (2012), combining words with images and videos, external links to online encyclopaedias, academic journal and newspaper articles, and other blogs or websites specializing in relevant topics. Four blogger-researchers were thus chosen of different seniority and institutional affiliations, all based in the US, not necessarily known through their being popular science writers, but responding to all of the criteria above. Most importantly, they were all taken to be representative of the independent expression of the practitioner's own voice and perspective on the topics dealt with.

The criteria for selecting the texts for each sub-corpus include time span and size. The target corpus size dictated the time frame of the posts: in order to reach the set number of words, only posts published between March 2019 and March 2014 were taken into consideration in the text collection, resulting in a five-year time span. This also allowed to choose recent posts that were as close as possible to the moment of analysis. Incidentally, the collection was concluded before the pandemic outbreak of late 2019 and early 2020, which unprecedentedly boosted the scale of web-based science writing.

In its current form, the Science Blogs Corpus altogether consists of approximately 800,000 words in total and includes 791 Posts and 393 Comment sections (each section gathers one or more comments). Each file was named to include the abbreviated blog name and the date of the post in the YY/MM/DD format, for chronological ordering. To distinguish posts from comments, a *Z*- was added in front of the blog name abbreviation, and the date of the post was maintained, so that, for example, *SKU_2014-04-17.txt* stands for post published on 17th April 2014 on Skulls in the Stars, and *Z-NEU_2019-01-18.txt* contains the comment section belonging to the post published in NeuroLogicaBlog on 18th January 2019. Tables 1 and 2 show the composition and size of both sub-corpora.

Individual Blogs	Word Count	
	Posts	Comments
<i>Genomics, Medicine and Pseudoscience</i>	93,068	7,481
<i>Mountain Beltway</i>	86,819	10,577
<i>NeuroLogicaBlog</i>	45,196	54,749
<i>Skulls in the Stars</i>	91,585	12,030
Subtotal	316,668	84,837
Total	401,505	

Table 1: Composition and size of the Individual Blogs sub-corpus

Institutional Blogs	Word Count	
	Posts	Comments
<i>Discover Magazine Blog</i>	83,714	17,554
<i>Physics Buzz</i>	94,858	6,982
<i>Public Library of Science (PLOS)</i>	93,826	6,578
<i>Science News</i>	74,585	25,034
Subtotal	346,983	56,148
Total	403,131	

Table 2: Composition and size of the Institutional Blogs sub-corpus

Since the institutional blogs sub-corpus is not used in this analysis, the themes and brief overview of each blog are given only for the individual blogs.

Genomics, Medicine and Pseudoscience deals with topics such as genomics and medicine as well as pseudoscience and unscientific medical practices. The blog was started in 2007 by Steven Salzberg, a Professor of Computational Biology and Genomics at Johns Hopkins University, who declares “Here’s where I can say what I really think about abuses and distortions of science, wherever I see them”. Besides an intense research activity focused on developing novel methods for the analysis of DNA and RNA sequences, Salzberg also writes for several international journals, like *PLOS Computational Biology*, *Nature* and *Science* and magazines such as *Forbes*, where he has a column dedicated to medical and biological discoveries (titled *Fighting Pseudoscience*).

Mountain Beltway is about geology and earth sciences. The author is Callan Bentley, Assistant Professor of geology at Northern Virginia Community College and Fellow of the Geological Society of America since 2017. His blog started in February 2010 and combines articles dealing with more general topics such as earthquakes, volcanoes and fossils with posts called *Friday Fold*, devoted to field trip reports and containing photos and videos of the excursions, and book reviews dealing with geology and other scientific issues. Despite being hosted

by the American Geophysical Union platform, the blog makes clear that “When this blog expresses opinions, those opinions are solely those of the author, and are not intended to be representative of Northern Virginia Community College, the Virginia Community College System, the Commonwealth of Virginia, or the American Geophysical Union”.

NeuroLogicaBlog, with subtitle *Your daily fix of Neuroscience, Skepticism and Critical Thinking*, is written by Doctor Steven Novella, Assistant professor of clinical neurology at Yale University School of Medicine and a co-founder of the New England Skeptical Society. The blog, which he started in January 2007, covers not only news and issues in neuroscience, but also general science, scientific skepticism, and the intersection of science and society. Since its launch, it has been widely appreciated by both the public and the critics. It was featured in OpenLab 2007 (*The Open Laboratory: The Best Science Writing on Blogs 2007*, edited by Reed Cartwright and Bora Zivkovic) and in 2018 it was mentioned among the “17 science blogs everyone should follow” by Ata Scientific. Steven Novella’s blog is openly interactive in that it has a *Topic Suggestions* tab where he is inviting readers’ suggestions for new topics (“Post and discuss your suggestions for new topics here”).

Skulls in the Stars, with subtitle *The intersection of physics, optics, history and pulp fiction*, was launched in August 2007 by Professor Gregory J. Gbur, Associate Professor of physics at UNC Charlotte and specializing in optical science. The blog covers a wide range of topics such as physics, astrophysics, optics and mathematics, but also literary fiction. Like the other individual blogs so far, this too has a disclaimer whereby “The views presented on this blog are the author’s own and do not represent the views of the university”. The blog was 2010 Research blogging awards finalist⁷ and mentioned in OpenLab 2010 and 2012. Gregory J. Gbur is also the author of a number of popular science books.

2.2 Methodology

The methodology used for the analysis of bloggers-audience interaction in Section 3 is limited to the individual blogs component of the Science Blogs Corpus but aims to lay the groundwork for further comparative investigations. It consists in extracting information about frequency from the corpus and in deriving patterns of writer-audience interactions through the reading of concordances and by interpreting them in the context of strategies of science communication and audience engagement. The software used for the extraction of frequencies and for concordancing is #LancsBox v.5.1 (Brezina et al. 2020) a software package developed at the University of Lancaster that has many useful features. It allows to work with one’s own dataset, it automatically annotates data

for parts-of-speech and, in its latest version, has an integrated tool (the Wizard) that combines in one search the tools usually found in concordancing software, namely it computes frequency and dispersion measures for both word and n-gram types, lemmas and POS tags, it compares frequency distributions using the keywords tool and finds collocates using different association measures for collocation extraction, including the MI (Mutual Information) statistic used here (cf. Brezina et al. 2015).

Because of the range of statistical tests one can choose from, the software was found especially suitable for a frequency-driven methodology and was therefore used to run different types of queries on the individual blogs sub-corpus and generate multiple information on frequency distributions, namely lemma lists of both words and n-grams (sequences of n words), keywords (words that are typical of a text/corpus when compared to another text/corpus) and keyword collocates. As far as n-grams are concerned, 4-grams were deemed to be the most appropriate length for an investigation into formulas of interaction for two reasons: first, 4-grams have been shown to reduce the amount of overlap between n-grams compared to shorter sequences (Biber et al. 1999), second they made the data comparable to Biber et al.'s (ibid.) 4-word bundles in spoken conversation vs. written academic prose.

Expanded concordances were then used to be able to view the search in the wider textual context and identify the audience engagement strategy. Because the ultimate aim of this research is the analysis of writer-reader interaction, the data on word frequency distributions were kept separate for the posts and the comments on the assumption that they together form a dialogic unit where posts are addressed one-to-many, while comments sections many-to-one⁸. As a further methodological note, because both lemma lists and word lists were used in the analysis, the small caps font was used to identify lemmas as opposed to simple text identifying word forms.

3 Results: Interaction in individual science blogs

3.1 Comparing word frequency distributions in Posts and Comments

One of the first queries run on the individual blogs sub-corpus to identify language patterns of interaction was to generate frequency lists of the top lemmas in the Posts and Comments separately. The most frequent items in the lemma lists overlap. In fact, if we take the top ten frequency lists (Table 3), both have nine items in common, namely *THE*, *BE*, *OF* with identical ranking, *TO*, *A*, *AND*, *IN*, *THAT*, *IT* with a slight difference in the ranking but comparable normalised frequencies. Interestingly, however, the first-person pronoun *I*, including all its

inflectional variants, appears in the top ten list of the Comments and not of the Posts, which instead display the forms of the proximal demonstrative *THIS*, both pronoun and determiner.

Individual Blogs	Posts		Comments	
Rank	Lemma	Norm. frequency PTT (Raw freq.)	Lemma	Norm. frequency PTT (Raw freq.)
1	THE	605.334 (19,169)	THE	501.550 (4,255)
2	BE	384.946 (12,190)	BE	440.374 (3,736)
3	OF	337.483 (10,687)	OF	279.831 (2,374)
4	A	258.820 (8,196)	TO	253.309 (2,149)
5	TO	251.936 (7,978)	A	233.742 (1,983)
6	AND	235.767 (7,466)	AND	225.373 (1,912)
7	IN	196.041 (6,208)	THAT	176.220 (1,495)
8	THAT	165.378 (5,237)	IN	166.673 (1,414)
9	IT	98.368 (3,115)	I	136.261 (1,156)
10	THIS	91.958 (2,912)	IT	117.519 (997)

Table 3: Top 10 lemmas in Individual Posts and Comments

This difference can be further explored by using the Keywords application to compare the two frequency distributions from the Posts and Comments. Keywords are a way of zooming in on distinct word types that are typical of posts and comments respectively, thus complementing the more general picture of vocabulary provided by the lemma list in Table 3. As shown in Table 4, the top ten keywords in the Posts, i.e. words whose frequency is statistically significant when compared to that of the Comments, contain exclusively content words, almost entirely nouns, which are either general scientific terms such as *heat*, *plane*, *motion*, *wave*, and *measurement*, or references to the geopolitical context in which the bloggers work, for example, *US* and *Trump*. *Friday*, the second most significant keyword in the posts might be due to one blog skewing the results because of the section *Friday Fold* in *Mountain Beltway*.

Rank	Posts keywords	Comments keywords
1	heat	thank_n
2	Friday	l be
3	U.S.	thank_v
4	plane	helmet
5	motion	aspartame
6	million	Novella

Rank	Posts keywords	Comments keywords
7	Trump	can't
8	wave	your
9	Mayer	soil
10	measurement	don't

Table 4: Keywords comparison of Individual Posts and Comments

On the contrary, among the Comments' keywords, one notices pragmatic routines such as *thank* (as both verb and noun), the negative operators *can't* and *don't*, and two deictic pronouns, namely the first-person singular pronoun *I* followed by various forms of the verb *BE* and the *possessive your*. The only lexical words appearing among the top ten keywords in the Comments are *helmet*, *aspartame*, *soil* and the name of one of the bloggers, *Novella*. Therefore, while the keywords in the Posts are all lexical words and mainly nouns, the Comments are characterised by a majority of grammatical words including verbal and deictic items. The difference observed reflects the different communicative nature of posts and comments, the posts informing about and prompting discussion on science-related topics, whereas the comments having an interpersonal function of responding utterance, as shown by the presence of first- and second-person pronouns and interactional formulas such as thanks and other markers of politeness, which are part of the necessary etiquette regulating the disputation of the post-comment new type of social media adjacency pair.

As revealed by the concordance lines, the two negatives *can't* and *don't* point to broader meanings of subjectivity, modality and interactivity as dominant categories in comments. A quick look at the concordances of *can't* shows that the form often follows a Subject pronoun (in 26 of the total 49 instances of *can't*), *I* and *you* being the two most frequent ones (respectively occurring 10 and 7 times), as can be seen in the filtered KWIC concordances in Figure 1.

Z-NEU_2019-02-28.txt	for the parents or whatever, I see	100%	fully agree to see it as a
Z-NEU_2019-02-28.txt	three drugs for decades, and I certainly	100%	be good. Continue and learn used to
Z-NEU_2015-12-22.txt	"news". Since he never measures "environmental", he	100%	allocate degrees of freedom to the environment,
Z-GEN_2016-08-01.txt	not hundreds—of advisory boards, to be	100%	possibly pay attention to most of them.
Z-NEU_2016-02-28.txt	what would I tell my child? I	100%	imagine any "typical" person feeling cheated when
Z-NEU_2016-01-29.txt	image of the fly brain. Disclaimer: I	100%	lower that I have got this completely
Z-GEN_2016-01-29.txt	have become extremely prevalent and inescapable, I	100%	help but be scared whatever I read?
Z-GEN_2016-01-29.txt	in class rather than a lecture. I	100%	say that the flipped classroom seems that
Z-NEU_2018-05-23.txt	only one that works for me. I	100%	even describe to the psychosocial WHAT it
Z-GEN_2017-09-18.txt	is to be drowned out by proponents, I	100%	believe I'm the minority in any group?
Z-GEN_2016-05-23.txt	to put my mind at rest, I	100%	help but worry otherwise. Anonymous, I don't
Z-GEN_2016-10-15.txt	not available in the U.S. Steven: I	100%	thank you enough for your articles, and
Z-NEU_2018-05-23.txt	Adherent, Britain and another one that I	100%	find a reference to, 25 years ago.
Z-NEU_2019-01-18.txt	"Steven? What is this article about? It	100%	be about the safety of GMO foods,
Z-NEU_2019-01-18.txt	the absolute safety of GMO foods. It	100%	be done, simply because GMO foods will
Z-NEU_2019-02-08.txt	approach does result in phantom herbality. It	100%	not result in phantom herbality. We assumes
Z-NEU_2019-02-08.txt	can imagine another possible problem. You probably	100%	will if it misfires, if the chance
Z-NEU_2019-02-08.txt	like we saw it because they	100%	write their own software, maybe??? But like
Z-NEU_2015-12-22.txt	of public and economic life because they	100%	compete. To be clear, I'm not against
Z-GEN_2014-04-07.txt	vaccines and Pasteurization got rid of. They	100%	imagine why these things are necessary given
Z-NEU_2013-12-22.txt	cells can't work together "to save", they	100%	moderate missing (which is an emergent property
Z-NEU_2017-05-16.txt	of courtroom drama? This is why we	100%	get people to listen—they hear the court?
Z-NEU_2016-12-11.txt	schedules would require more teachers, because we	100%	got expert visiting teachers to add their
Z-NEU_2017-12-09.txt	enter news gear about legal cost. You	100%	communicate. Period. I never saw anything rational
Z-GEN_2014-10-29.txt	when others do it. I mean, we	100%	they it anyway—so better be prepared.
Z-NEU_2016-05-18.txt	everything. The same with social status. You	100%	subtract these from science until robot scientists
Z-NEU_2017-12-09.txt	yourself into thinking you know things you	100%	possibly know it a major cause of
Z-NEU_2018-10-30.txt	a mistake, at least with vaccines? You	100%	even question their quality? You can believe

Figure 1: Concordances of Subject pronouns + *can't* in Comments

The type of modality expressed by the sequence Subject pronoun + *can't* is predominantly of the epistemic type (1) and sometimes fixed into a pragmatic routine (2). When combined with *you* the modal can also express capacity/ability as in Example (3), where it is accompanied by the modal adjunct *probably* with a hedging function (the file name in brackets specifies blog name and date of the post):

- (1) *Steven? What is this article about? It can't be about the safety of GMO foods, it seems to be about marketing.* (Z-NEU_2019-01-18.txt)
- (2) *but those are not available in the U.S. Steven: I can't thank you enough for your articles and support of vaccination.* (Z-GEN_2016-10-15.txt)
- (3) *I can imagine another possible problem. You probably can't tell if it misfires. If the chance of it misfiring is non-negligible, I would think that patients would have to test themselves to make sure it worked. Perhaps it will work well enough when the stomach is empty...* (Z-NEU_2019-02-08.txt)

A quick check for concordances of *don't* sorted to the left reveals a similar preference for Subject pronouns in the immediate left co-text, with the sequence *I don't*, followed by *think* and *know* among the most frequent collocates to the right, outnumbering the other pronominal sequences (61 instances of the total 153), e.g. *we don't* and *you don't* (16 instances each) and third-person plural reference *they don't* (occurring 13 times). Fronted negative imperatives

with a strong interpersonal function (e.g. *don't ignore or deny them, don't get me wrong, don't even think about it, don't fucking dare call this a problem that doesn't exist, please don't stigmatize psychiatrists*), together with two instances of focusing strategies, as in the pseudo-clefts *what I don't understand is...*, *what I don't get is...*, are the other two noticeable patterns that explain the keyness of *don't* in Comments when compared to Posts.

The observation that first- and second-person pronouns are prominent among the top ten Comments' keywords takes us one step further to explore the distribution of personal and possessive pronouns in more detail. Going beyond the top ten keywords and considering the top fifty keywords, one finds that while Comments also rank *I WILL, you, you WILL, you BE* (the small caps signalling the lemmas), Posts only display one form of personal reference, namely *let's* occupying the 14th position in the list. This distribution recalls Hyland's (2010: 125) "reader pronouns", which, together with questions, are the most direct way of bringing readers into the discourse in popular science texts, as also shown by Luzón's (2013a: 449) analysis of academic blog posts. It also reminds of Bondi's (2018: 54-55) finding that first-person imperative *let's* is a preferred form of reader engagement in the posts besides explicit mentions of *readers*.

These frequency-driven observations therefore call for a more extensive analysis of pronouns and other markers of interaction in the individual blogs sub-corpus, as presented in the following subsection.

3.2 Bloggers enacting interaction with their audiences: *you* and *I*

Although first- and second-person pronouns were found to occur in both the Posts and Comments components of the individual blogs sub-corpus, a statistically significant difference between the two was found with regard to the linguistic variables *I* and *you*, which prevail in the Comments (calculated with a t-test, as in Brezina 2018). This, for example, is visible through a simple search for *you* in the Posts, which yielded 1,112 instances (35.12 PTT) in 277 of the total 411 texts, signifying that not all posts contain an instance of *you*, although all four blogs do. Conversely, a search for dispersion of *I* and *you* in the Comments returned instances spread over 155 and 124 comments respectively of the total 197 texts and *I* was found in the comments to all four blogs.

A useful way of looking at the usage of *you* and *I* is by considering collocates. Following Sinclair (1991), collocates up to four words on either side of the target term were computed, to compare the two pronouns collocates across Posts and Comments.

In the Posts *I* often co-occurs with verbs encoding mental processes such as *tire, regret, doubt, guess, appreciate* lexicalising the blogger's viewpoint,

a couple of material/verbal processes such as *finish* and *blogged* referring to the blogger's actions, and some negative finite operators *haven't*, *hadn't*, *wasn't*, *couldn't* (Table 5). A few mental processes of affection are also present among the top collocates, e.g. *welcome*, *delight (in)*. In the Comments, the list of top collocates of *I* is no less rich in the expression of subjectivity, displaying verbs of cognition expressing an opinion or attitude, such as *notice*, *wish*, *suspect*, *guess*, *feel*, *wonder*, but also *agree* and *recall* pointing to the inherent link between post and comment and the function of rebuttal to a claim. The adverb *personally* similarly reinforces subjectivity and stance (Table 6).

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
1	R	tire_v	7.500	4	6
2	R	regret_v	7.179	8	8
3	R	haven't_n	7.179	5	5
4	R	hadn't_n	7.009	8	9
5	L	zero_n	6.956	6	7
6	R	finish_v	6.792	13	17
7	R	doubt_v	6.638	11	16
8	R	guess_v	6.594	22	33
9	R	blogged_v	6.500	5	8
10	R	wasn't_v	6.487	13	21
11	R	appreciate_v	6.442	9	15
12	R	welcome_v	6.442	6	10
13	R	delight_v	6.390	11	19
14	R	couldn't_n	6.371	8	14
15	R	wish_v	6.342	14	25

Table 5: Top collocates of *I* in the Posts

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
1	R	notice_v	6.036	13	14
2	R	wish_v	6.017	11	12
3	R	personally_adv	5.991	9	10
4	R	suspect_v	5.973	8	9
5	R	don't_n	5.950	7	8
6	R	guess_v	5.931	19	22
7	L	skip_v	5.880	5	6
8	R	agree_v	5.799	26	33
9	R	breakfast_n	5.728	15	20
10	R	recall_v	5.728	6	8
11	L	sorry_adv	5.657	5	7

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
12	L	damien_n	5.657	5	7
13	R	feel_v	5.652	37	52
14	R	wonder_v	5.640	24	34
15	R	love_v	5.581	21	31

Table 6: Top collocates of *I* in the Comments

As far as collocations of *you* in the Posts are concerned, some of the same verbs co-occurring with *I* in the Comments appear among the strongest collocates, namely *recall* and *notice*, reflecting the switch in speaker deixis between posts and comments. Moreover, material processes such as *order*, *buy* and *eat* appear in the list, pointing to generic situations the reader might easily relate to. The strongest collocate to the right of *you* is the negative *haven't* and interestingly the conditional conjunctions *unless* and *if* appear in the left co-text, together forming the pattern *if/unless you haven't* introducing an imagined situation or suggestion (Table 7).

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
1	R	haven't_v	7.497	7	11
2	L	unless_con	7.231	9	17
3	R	smarter_adv	7.149	6	12
4	R	order_v	7.033	6	13
5	R	recall_v	6.901	8	19
6	R	can't_v	6.770	15	39
7	R	won't_v	6.756	8	21
8	R	notice_v	6.689	16	44
9	M	shot_n	6.646	6	17
10	R	fly_v	6.383	5	17
11	R	physically_adv	6.383	5	17
12	R	buy_v	6.319	9	32
13	R	eat_v	6.314	30	107
14	L	fast_adv	6.312	7	25
15	L	if_con	6.163	207	820

Table 7: Top collocates of *you* in the Posts

One of the most frequent collocates of *you* in the Posts (the most frequent if the search span is restricted to three words to the right) occurs in the phrase *you might be wondering* which seems peculiar to one blogger, Steven Salzberg of *Genomics, Medicine and Pseudoscience*, who uses it as a parenthetical comment inside a question as a strong marker of interaction, as in *How (you*

might be wondering) *this can be sold...*, or jocular remark with the same function, as in *You might be wondering what my score was*. The two bloggers of *Mountain Beltway* and *NeuroLogica* also make use of a similar phrase, when they write, respectively *Perhaps, you are wondering what's going on here.*, also said jokingly, and *At this point you might be wondering why we allow people...* which is prefaced by signposting language at this point which is also typical of interactive discourse. Both instances function as direct address and at the same time anticipate a possible question or objection that might be raised by a reader and function as explicit markers of a carefully constructed interaction.

Similarly, the collocate pair *you haven't* belongs in *if*-clauses addressing the readers directly and often containing a suggestive as in: ... *A Short History of Nearly Everything. If you haven't read that one yet, it really...*, *So has conservative columnist David Frum (if you haven't read his piece in the Atlantic..., So, if you haven't given any thought to AI, I..., Here's his "make good art" speech: If you haven't seen it, you should., If you haven't read The Broken Earth trilogy, it...* The straight conversational style of this phrase, sometimes going together with advice or suggestions, is used almost exclusively by *Mountain Beltway* except for *The Broken Earth trilogy* instance, also concerned with what readers might or might not have read, which comes from *Skulls in the Stars*. Direct reference to readings and other sources of information, the normal context in which these *you*-clauses occur, seem to point to another discourse function identified by Myers (2010) that responds to the need to build a space of shared knowledge and common evaluations between writer and audience.

Generic *you*, replacing the impersonal subject pronoun *one*, is also present in the Posts, recalling the use of generic *you* meaning 'anyone', which Myers (2010: 10) says occurs when talking about norms of behaviour or common practices (as in *You can lounge around on Northern Line upholstered seats and sip...* In the Posts there are instances of generic *you* when the blogger gives an explanation of some scientific law, as for example in the expanded concordances below, in Example (4) from *Skulls in the Stars* and (5) and (6) from *Mountain Beltway*:

- (4) *The conservation of energy proves the non-existence of perpetual motion machines: in order for a machine to provide unending motion, it must have an inexhaustible source of energy to power it. Or, in other words: **you** can't get more energy out of a machine than **you** put into it.* (SKU_2018-12-10.txt)

- (5) *Another thing we may notice is that there's a beautiful aplite dike cutting semi-horizontally through the prominent knoblet of Alabama Hills granite in the right foreground. These fine-grained felsic dikes are common in the plutons of the Sierra Nevada batholith, and thanks to their clear cross-cutting relationship with the host granites, reveal themselves to be younger than the rocks into which they have intruded. It has a closely-spaced set of dike-orthogonal fractures cutting across it, giving it a blocky weathering pattern. These joints formed in the shortest possible direction to cut across the dike from one side to the other, and these fractures of course must post-date the dike. **You** can't break something that doesn't yet exist. (MB_2018-02-08.txt)*
- (6) *The bedding was laid down sometime around 392 to 385 million years ago. The cleavage developed as these rocks were squeezed almost 100 million years later, around 300-250 million years ago. **You** can't deform something unless it already exists. (MB_2017-02-27.txt)*

What is absent from the Posts are *you* + modifier sequences of the kind discussed by Myers (2010: 80), such as *you guys*, *you traditionalists* referring to all the readers of the blog at once, or to a small group within the readership, with general indefinite reference becoming more specific.

In the Comments, however, the list of collocates of *you* is quite different from that of the Posts, as visible in Table 8. The second strongest collocate of *you* is *thank* to the left of the target pronoun, belonging to the routine expression *thank you*, once again pointing to the different interpersonal dimension of Comments compared to Posts. Then three verbs *realize*, *comment* and *expect* also accompany *you* sometimes with a generic meaning, and elsewhere *you comment* referring to the blogger or another commenter. More verbs encoding mental processes appear when going down the list, e.g. *wonder*, *want*, *demonstrate*. These verbs accompanying the subject pronouns reveal what Bondi (2018: 54) has referred to as the “initiative and reactive function” of blog posts and comments.

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
1	R	worldview_n	6.713	10	10
2	L	thank_v	6.683	47	48
3	R	realize_v	6.076	9	14
4	R	rational_adj	5.865	5	9
5	R	comment_v	5.713	5	10
6	L	ask_v	5.491	9	21
7	R	yourself_pron	5.228	5	14
8	L	picture_n	5.211	6	17
9	L	steven_n	5.211	6	17

ID	Position	Collocate	Stat (MI)	Freq. coll.	Freq. corpus
10	R	expect_v	5.128	7	21
11	L	wonder_v	5.085	11	34
12	R	wrong_adj	5.076	9	28
13	R	want_v	5.073	17	53
14	M	hope_v	5.050	6	19
15	R	demonstrate_v	5.035	5	16

Table 8: Top collocates of *you* in the Comments

Among the top left-collocates of *you* in the Comments is the word *Steven*, which can be explained as a marker of direct address, i.e. the explicit mention of the blogger in the text (Myers 2010, Bondi 2018). According to Myers (2010: 78), the use of named addressees is rare in the actual posts and more common in the comments on posts, usually when the commenter wants to single out the original post or a particular subset of the audience and therefore uses a direct form of address or a vocative. In her comparison of posts and comments, Bondi (2018: 55-56) similarly finds the blogger’s name among the comments keywords to signal more direct involvement, often with a challenging or evaluative function. This function is indeed present in the comments of the sub-corpus considered here, as can be seen in the following expanded concordances in (7), (8) and (9):

- (7) *I beg to differ on one point, **Steven**. I’m sure that if most of us were to eat “like a gorilla” (i.e. plant stems, bamboo shoots, fruit, termites) we would probably lose weight.* (Z-GEN_2015-12-07)
- (8) ***Anonymous**: I don’t comment on YouTube videos or website claims, which don’t serve as evidence of anything. Any idiot (pardon my bluntness) can go on YouTube and make wacky claims, and many people have. ...* (Z-GEN_2015-01-20.txt)
- (9) ***BB**- you seem to be inferring that I prefer so-called “natural” methods. You are preaching to the choir.* (Z-NEU_2017-05-16.txt)

3.3 The conversational style of blogs: Questions in Posts and Comments

In both Posts and Comments, personal deixis often occurs as part of wider chunks that are instances of what Myers (2010: 84-86) calls “a conversational style” of blogs including conversational inserts and asides (*I know, I know... What’s that you say?* Myers’ examples), and other markers of orality such as question tags (*eh?*), discourse markers and interjections.

Of the four blogs, Steven Salzberg’s seems to especially adopt this blogging style. As can be seen in the concordances in Examples (10) and (11), the discourse

marker *okay*, the clause separated by a comma *The problem is*, the dialogical *Sorry, but no: you can't...* and the afterthought segment *At least not like...* help the writer make his point in a forceful way albeit concealed by the conversational and, in Example (11), humorous tone:

- (10) 5. *Antibiotics. Okay, these are real medicine, and you can't buy them over the counter at pharmacy. But Americans take them in huge quantities to treat the common cold. The problem is, antibiotics don't work for colds.* (GEN_2014-11-17.txt)
- (11) *Sorry, but no: you can't eat like a gorilla and lose weight. At least not like a healthy gorilla.* (GEN_2015-12-07.txt)

Myers (2010: 85) finds that some of these inserts and asides anticipate readers' comments often in Free Indirect Speech (FIS), freely enacting the speech attributed to the audience, creating a sort of "conversational self-interruption". Argumentatively, he explains, this corresponds to replying to an imagined objection, thus turning one's own text into a polyphonic one by making it open for criticism or mockery and often serving a self-mocking function (see also Mauranen (2013: 28-29) highlighting bloggers' use of FIS to create spontaneity and immediacy).

Interjections and similar markers of spontaneous conversation appear with the 4-word bundle *it turns out that*, which is also a sign of a spoken style as in (12), (cf. Subsection 3.4 below for an analysis of bundles):

- (12) *Uh oh. It turns out that these products are little more than very, very expensive bottles of sterile water. For \$9.99 you get 10 ml of water that contains several extracts—in vanishingly small amounts—for which there is no evidence whatsoever that they have any effect on allergies.* (GEN_2016-05-22.txt)

The next concordance in (13) well captures Salzberg's colloquial style full of markers of interactive discourse and is illustrative of another two common structures bloggers use to engage with their audience, namely questions and directives (see Hyland 2001, 2010, Myers 2010, Luzón 2013a, Bondi 2018). The conditional clause with *you* starts the conversation with a dialogic orientation signalled by the reiteration of the second person pronoun: *if you're... even if you don't... how would you...?* Then the two rhetorical questions and the final directive, also prefaced by the discourse marker *well*, close the exchange with a call for rethinking and the same building-shared-knowledge function observed in the previous examples:

- (13) *If you're reading this, you care deeply about net neutrality- even if you don't know what it means. **How would you like to pay 2, 3, or maybe 10 times as much to access all the websites you frequent? Think it can't happen? Well, think again.*** (GEN_2014-06-30.txt)

Questions are another example of the conversational style of blogs. Several studies of popular science texts have analysed the use of questions as an engagement device. Drawing on Hyland (2010), Luzón (2013a: 449-450) has shown that questions are a frequent strategy in blog posts, occurring in more than 60% of the posts analysed and fulfilling the interpersonal function of creating a dialogue with readers by presenting them with a hypothesis or scenario to which they can immediately relate. Similarly, Myers (2010: 82-84) stresses how both questions, calling for a response, and directives, a call for action, are used by bloggers to elicit information or comments. In particular, he argues that rhetorical questions always convey the bloggers' stance by aligning writer and readers through the idea that the answer is known and shared (ibid.: 109).

In the Individual Blogs sub-corpus an automated search for questions was attempted by querying the question mark. In the Posts this returned 747 instances (23.59 PTT) distributed across all four blogs. The interrogatives retrieved have very different functions. Some are a way to incorporate a possible question or objection in the post, to have a chance to address it as in (14) and (15) both from Steven Salzberg's blog. Others are meant to catch the reader's attention, as in (16), also from the same blog. All of these questions are a sign of interactivity:

- (14) *But wait, **you might ask, don't chiropractors provide pain relief? And don't they have medical degrees?** Well, on the second question, the answer is that they have special Doctor of Chiropractic (D.C.) degrees, which are given out by just 15 special chiropractic colleges in the U.S. The entire field was invented out of thin air by D.D. Palmer in 1895, and later popularized by his son.* (GEN_2014-04-20.txt)
- (15) *Right about now **you might be asking: is anyone actually doing this, and if so, what on earth are they thinking?*** (GEN_2014-10-20.txt)
- (16) ***See how that works?** Doshi uses this slight-of-hand to suggest that* (GEN_2014-11-03.txt)

The questions in the Comments amount to 388 (45.73 PTT), almost twice as much as in the Posts, and some of the functions identified are different from those in the Posts. For example, sometimes commenters address the blogger with a question they then provide an answer for to carry on with their argument (17). In

general, questions in Comments seem to have an explicitly challenging function prefacing a rebuttal, as exemplified by the comment in (18) to Steven Novella's blog:

- (17) *Hey Steve. What impact do you think AI will have on the field of radiology? I doubt radiologists will be replaced but I think there will be a reduction in workload* (Z-NEU_2019-03-12.txt)
- (18) *Steven? What is this article about? It cant be about the safety of GMO foods, it seems to be about marketing. You won't ever convince anyone about the absolutely safety of GMO foods* (Z-NEU_2016-02-26.txt)

3.4 Spokenness and interaction: Lexical bundles in Posts and Comments

The next step of the analysis was to check the frequency and distribution of 4-word lexical bundles across the whole sub-corpus, and then compare the findings to conversation and academic prose as in Biber et al. (1999: 996-997, Table 13.1 and Table 13.2), i.e. the proportional distribution of 4-word lexical bundles across the major linguistic patterns in each register. This would help place blogs somewhere along the cline between spontaneous conversation at one end and academic prose at the other end. The same query was then run on the Posts and Comments separately to highlight the specificities of each component.

This kind of analysis relies on Biber et al.'s (1999) finding that lexical bundles, i.e. sequences of words of varying length, are frequent and widely used across texts and therefore can be regarded as lexical building blocks that are independent of the topic, but rather are specific to the register. The large-scale comparison of conversation and academic writing made by Biber et al. (ibid.: 992) yielded the finding that "most lexical bundles in conversation are building blocks for verbal and clausal structural units, while in academic prose for extended NPs or PPs". More in detail, we read, "In conversation, the large majority of lexical bundles are clause segments: either declarative structures with a subject pronoun followed by an extended verb phrase, or interrogative structures" (ibid.: 1000).

Using #LancsBox n-grams function, it was possible to process the blogs for 4-word lexical bundles. The type and distribution of 4-grams is slightly different in the Posts from the Comments, as can be seen in Tables 9 and 10 containing the top ten 4-gram types. The list is cut at the 10th rank as frequencies become much smaller in the higher ranks.

Rank	Type	Raw freq. (Norm. freq.)	Dispersion (Coefficient of Variation)
1	it turns out that	35 (1.110)	4.174
2	one of the most	35 (1.110)	4.342
3	is one of the	32 (1.014)	5.156
4	the end of the	32 (1.014)	5.559
5	the speed of light	25 (0.793)	9.961
6	if you want to	19 (0.602)	4.959
7	than the vacuum speed	18 (0.571)	12.634
8	the vacuum speed of	18 (0.571)	11.832
9	vacuum speed of light	18 (0.571)	11.832
10	the rest of the	17 (0.539)	5.415

Table 9: Top ten n-gram types in Posts

Rank	Type	Raw freq. (Norm. freq.)	Dispersion (Coefficient of Variation)
1	at the same time	10 (1.187)	8.360
2	the rest of the	9 (1.068)	8.220
3	that there is a	7 (0.831)	6.894
4	there is evidence that	7 (0.831)	13.286
5	a and b are	6 (0.712)	14.000
6	for which there is	6 (0.712)	8.702
7	I don't know if	6 (0.712)	8.063
8	is one of the	6 (0.712)	6.936
9	an increased risk of	5 (0.593)	12.042
10	as a result of	5 (0.593)	6.300

Table 10: Top ten n-gram types in Comments

Interestingly, both Posts and Comments rank a sequence containing a personal pronoun, respectively *if you want to* in the Posts and *I don't know if* in the Comments, both showing a lower Coefficient of Variation (CV) as an indication of their being distributed across all blogs.

The most frequent 4-gram in Posts, and one with a low CV, is the colloquial sequence *it turns out that*, which points to the spoken nature of blog posts. Some overlap is also observable between *one of the most* and *is one of the* (as apparent from comparing the concordances) and more significantly no. 5, 7, 8 and 9 in the Posts having to do with the 6-gram *than the vacuum speed of light*. This specific n-gram appears in four different posts by the same blogger, namely Gregory Bur of *Skulls in the Stars* and is repeated 18 times, thus skewing the significance in terms of dispersion (as shown by the high CV).

The sequence with the various forms of the lemma *BE* followed by *one of the most*, which is spread across all four blogs, is completed by a variety of evaluative adjectives, expressing mostly positive evaluations (*amazing places I've ever been to, beautiful and amazing fields of mathematics, extensive collections of animal specimens, important unifying principles of physics, striking geological features, etc.*) and a few negative (*dangerous viruses, infectious human viruses*) or neutral ones (as in the adjectives on a scale of usuality in *common cancers in the U.S. and Europe, unusual properties of many fractal*). *NeuroLogicaBlog* shows one instance only of *BE one of the most* and a slightly different pattern with the past participle of the verb 'study', *Aspartame is one of the most studied food additives that exists*. That this pattern serves the purpose of conveying the blogger's evaluation and opinion is confirmed by the co-occurrence with expressions of stance such as *I think, [f]rom my perspective, for me*, reinforcing the opinion-making function of blogs (cf. the discussion in Section 4). Especially in two of the four blogs in the corpus, *Genomics* and *Skulls in the Stars*, *it turns out that* often introduces a conclusive claim or fact, having an argumentative function. The other two bundles *the rest of the* and *the end of the* are used with a framing, deictic or temporal, function (e.g. *see the rest of the list, described in detail at the end of the book, by the end of the century*).

Compared to Biber et al.'s (1999) findings, one can see a mixture of clausal (e.g. *it turns out that, if you want to*) and nominal (e.g. *the vacuum speed of, the wave properties of, an increased risk of*) units among the most frequent bundles, which would place blogs in between more typically spoken registers, such as conversation, and more typically written registers like academic prose. The number and variety of prepositional bundles is also quite high in the individual blogs, when considering the sequences (*at the end of the, (in) the middle of the, at the speed of, (for) the rest of* in the Posts and *at the same time, as a result of* in the Comments). These bundles, which contain nominal elements, bring blogs closer in style to written prose. Through analysis of 4-grams, blogs are thus located in between spoken conversational genres and more formal writing, sharing frequent lexical material from both.

4 Discussion and conclusions

Despite the written medium, the blogs analysed display many features of spoken genres, as already pointed out by Myers (2010), Mauranen (2013: 9) and Bondi (2018), who have studied the effects and affordances of the digital medium on writing. This seems to corroborate rhetoricians' finding that one of the reasons for the success of blogs as a genre of science communication lies in

the sense of immediacy and spontaneity that is more typical of spoken forms of communication than written academic prose.

The paper has explored interaction in science blogs and found that it is in the comments where interaction is more directly observable. The analysis has been driven by frequency observations based on a corpus of blogs maintained by four American scientists, taken to be representative of individual science blogging at large, i.e. blogging that is independent of research institutions or popular science publications. The corpus thus compiled has offered a reliable basis for frequency-driven explorations of the kind of discourse characterising blogs as well as a quantitative testbed for previous findings in the literature about enactment of interaction and audience engagement in blogs and other digital genres. The frequency observations through the lemma lists and keywords comparisons have revealed that deictic reference (through pronouns *I* and *you*) is a prominent marker of interaction in blogs, and an especially salient feature of comments when compared to posts. This is revealing of the double conversational and rhetorical role of comments, both as essential parts of the post-comment adjacency pair and rebuttals in the discussion prompted by blog posts.

The frequency observations have been expanded to include collocation and lexical bundles (n-grams). Collocation has highlighted the verbs associated with the Subject pronoun *I* and *you* in both posts and comments, revealing the cognitive and affective processes of bloggers and commenters. Lexical bundles (n-grams) have helped to define the blog genre in terms of spokenness and place it along a continuum with spontaneous conversation at one end and academic writing at the other end. The concordance analysis has then showed that Salzberg's style stands out for being especially dense with markers of a conversational style that aims at engaging the audience.

Finally, previous findings on questions as engagement strategy in blogs and other digital genres of science communication have prompted a search for interrogatives in the corpus, which have been found to serve the same rhetorical purpose of building a responsive and engaged audience as that already observed in the literature by Hyland (2010), Myers (2010) and Luzón (2013a), with a preference, both in frequency and type, for questions with a challenging function in the comments. In all, the frequency-driven analysis together with the reading of the concordances have allowed to distinguish the linguistic patterning of comments from that of posts and have helped in the definition of blogs as a genre whose features are shared and perpetuated by the community of practice of science bloggers and science readers, actively interacting to “promote science with and through the public” (Mehlenbacher 2019: 136).

Following Myers (2010), engagement strategies are only one side of the process of creating and maintaining interaction, the other side being the language writers employ for expressing opinions and stance, which deserves future study.

Notes

- ¹ This research has been carried out within the framework of PRIN 2015TJ8ZAS (Italian Ministry of University and Research), a national research project on “*Knowledge Dissemination Across Media in English: Continuity and Change in Discourse Strategies, Ideologies, and Epistemologies*”.
- ² Gallagher (2018) analyses how blog writers readdress their audience by quoting and drawing from comments, and how they rewrite and recontextualise their arguments when reading comments.
- ³ Walsh (2015) considers the content and quasi-deliberative function of discussions among commenters on science blogs that led to the decision by popsci.com editors to shut off public comments.
- ⁴ blog.feedspot.com/science_blogs/
- ⁵ <https://www.atascientific.com.au/17-science-blogs-everyone-should-follow/> Ata Scientific is an Australian company specialised in producing scientific instruments, who also contributes to the scientific community by sponsoring meetings and by reviewing articles and blogs that deal with science.
- ⁶ <https://www.iop.org>
- ⁷ <http://researchblogging.org/static/index/page/awards>
- ⁸ Both Luzón (2013b) and Bondi (2018) have shown that commenters however address other comments as well.

References

- Angler, M. W. (2017) *Science Journalism: An Introduction*. London: Routledge.
- Askehave, I. and Nielsen, A. E. (2005) ‘Digital genres: A challenge to traditional genre theory.’ *Information Technology & People* 18(2), 120-141.
- Beacco, J.-C., Claudel, C., Doury, M., Petit, G., and Reboul-Touré, S. (2002) ‘Science in media and social discourse: New channels of communication, new linguistic forms.’ *Discourse Studies* 4(3), 277-300.
- Bell, A. (2012) ‘Has blogging changed science writing?’ *Journal of Science Communication* 11(1), 1-5.
- Biber, D. Johansson, S., Leech, G., Conrad, S. and Finegan, E. (1999) *Longman Grammar of Spoken and Written English*. London: Longman.
- Blanchard, A. (2011) ‘Science blogs in research and popularization of science: Why, how and for whom?’ In: Cockell, M., Billotte, J., Darbellay, F., Waldvogel, F. (eds) *Common Knowledge: The Challenge of Transdisciplinarity*. Lausanne: EPFL Press. 219-232.
- Bondi, M. (2018) ‘Blogs as interwoven polylogues. The dialogic action game.’ *Language and Dialogue* 8(1), 43-65.
- Brezina, V. (2018) *Statistics in Corpus Linguistics: A Practical Guide*. Cambridge: Cambridge University Press.
- Brezina, V., McEnery, T. and Wattam, S. (2015) ‘Collocations in context: A new perspective on collocation networks.’ *International Journal of Corpus Linguistics* 20(2), 139-173.
- Brezina, V., Weill-Tessier, P., and McEnery, T. (2020). #LancsBox v.5.1 [software]. Available at: <http://corpora.lancs.ac.uk/lancsbox>.

- Büchi, M. (2017) 'Microblogging as an extension of science reporting.' *Public Understanding of Science* 26(8), 953-968.
- Fahnestock, J. (1986) 'Accommodating science. The rhetorical life of scientific facts.' *Written Communication* 3(3), 275-296.
- Gallagher, J. R. (2018) 'Considering the comments: Theorizing online audiences as emergent processes.' *Computer and Compositions* 48, 34-48.
- Hoffmann, C. R. (2012) *Cohesive Profiling: Meaning and Interaction in Personal Weblogs*. Amsterdam and Philadelphia: John Benjamins.
- Hyland, K. (2001) 'Bringing in the reader: Addressee features in academic articles.' *Written Communication* 18(4), 549-574.
- Hyland, K. (2010) 'Constructing proximity: Relating to readers in popular and professional science.' *Journal of English for Academic Purposes* 9(2), 116-127.
- Kouper, I. (2010) 'Science blogs and public engagement with science: Practices, challenges, and opportunities.' *Journal of Science Communication* 9(1), Special Issue on Peer-to-peer and User-led Science. Online document. Retrieved on 9 December 2015 Available at <http://jcom.sissa.it/>. <<https://doi.org/10.22323/2.09010202>>.
- Luzón, M. J. (2011) 'Interesting post, but I disagree: social presence and antisocial behavior in academic weblogs.' *Applied Linguistics* 32(5), 517-540.
- Luzón, M. J. (2013a) 'Public communication of science in blogs: Recontextualizing scientific discourse for a diversified audience.' *Written Communication* 30(4), 428-457.
- Luzón, M. J. (2013b) "'This is an erroneous argument': Conflict in academic blog discussions.' *Discourse, Context and Media* 2(2), 111-119.
- Luzón, M. J. and Pérez-Llantada, C. (2019) 'Connecting traditional and new genres: Trends and emerging themes.' In: Luzón, M. J. and Pérez-Llantada, C. (eds) *Science Communication on the Internet. Old Genres Meet New Genres*. Amsterdam and Philadelphia: John Benjamins. 1-18.
- Mauranen, A. (2013) 'Hybridism, edutainment, and doubt: Science blogging finding its feet.' *Nordic Journal of English Studies* 12(1), 7-36.
- Mehlenbacher, A. R. (2019) *Science Communication Online. Engaging Experts and Publics on the Internet*. Columbus: The Ohio State University Press.
- Miller, C. R. and Fahnestock, J. (2013) 'Genres in scientific and technical rhetoric.' *Poroi. An Interdisciplinary Journal of Rhetorical Analysis and Invention* 9(1), 1-4.
- Miller, C. R. and Shepherd, D. (2004) 'Blogging as social action: A genre analysis of the weblog.' In: Gurak, L., Antonijevic, S., Johnson, L., Ratliff, C. and Reyman, J. (eds) *Into the Blogosphere: Rhetoric, Community, and Culture of Weblogs*. Minneapolis: University of Minnesota Press. 1-21.
- Miller, C. R. and Shepherd, D. (2009) 'Questions for genre theory from the blogosphere.' In: Giltrow, J. and Stein, D. (eds) *Genres in the Internet. Issues in the Theory of Genre*. Amsterdam and Philadelphia: John Benjamins. 263-291.
- Myers, G. (2010) *The Discourse of Blogs and Wikis*. London and New York: Continuum.
- Puschmann, C. and Mahrt, M. (2012) 'Scholarly blogging: A new form of publishing or science journalism 2.0?' In: Tokar, A., Beurskens, M., Keuneke, S., Mahrt, M., Peters, I., Puschmann, C., van Treek, T. and Weller, K. (eds) *Science and the Internet*. Düsseldorf: Düsseldorf University Press. 171-181.
- Reid, G. and Anson, C. M. (2019) 'Public- and expert-facing communication: A case study of polycontextuality and context collapse in Internet-mediated citizen science.' In: Luzón, M. J. and Pérez-Llantada, C. (eds) *Science Communication on the Internet*.

Old Genres Meet New Genres. Amsterdam and Philadelphia: John Benjamins. 239-241.

Sinclair, J. (1991) *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Walsh, L. (2015) 'The double-edged sword of popularization: The role of science communication research in the Popsci.com comment shutoff.' *Science Communication* 37(5), 658-669.

Sources

Genomics, Medicine and Pseudoscience available at <http://genome.fieldofscience.com>

Mountain Beltway available at <https://blogs.agu.org/mountainbeltway/>

NeuroLogica Blog available at <https://theness.com/neurologicablog/>

Skulls in the Stars available at <https://skullsinthestars.com/>

Discover Magazine available at <https://www.discovermagazine.com>

Physics Buzz available at <http://physicsbuzz.physicscentral.com>

PLOS available at <https://theplosblog.plos.org>

Science News available at <https://www.sciencenews.org>

Maria Freddi is Associate Professor of English Language and Linguistics at Pavia University, Italy. She specializes in English for Academic and Specific Purposes, focusing on scientific and technical discourse, and is currently involved in the 3-year EU-funded Becoming a Digital Global Engineer (BADGE) research project, addressing the language needs of future engineers working in the increasingly challenging international market.

Address: Maria Freddi, Department of Humanities, Theoretical and Applied Linguistics Section, Strada Nuova 65, 27100 Pavia, Italy. [e-mail: maria.freddi@unipv.it]