ORIGINAL ARTICLE

Association of Learned and Professional Society Publishers

(wileyonlinelibrary.com) doi: 10.1002/leap.1411

Received: 10 March 2021 | Accepted: 29 May 2021

Over-promotion and caution in abstracts of preprints during the COVID-19 crisis

Frederique Bordignon,[1]* Liana Ermakova,[2] and Marianne Noel[3]

[1]Ecole des Ponts ParisTech, Marne-La-Vallée, France

[2]Université de Bretagne Occidentale, HCTI, Brest, France

[3]LISIS, CNRS, INRAE, Université Gustave Eiffel, Marne-La-Vallée, France

ORCID:
F. Bordignon: 0000-0002-4918-9137
L. Ermakova: 0000-0002-7598-7474
M. Noel: 0000-0002-4965-2224

***Corresponding author:**
Frederique Bordignon, Ecole des Ponts ParisTech, Marne-La-Vallée, France.
E-mail: frederique.bordignon@enpc.fr

Abstract

The abstract is known to be a promotional genre where researchers tend to exaggerate the benefit of their research and use a promotional discourse to catch the reader's attention. The COVID-19 pandemic has prompted intensive research and has changed traditional publishing with the massive adoption of preprints by researchers. Our aim is to investigate whether the crisis and the ensuing scientific and economic competition have changed the lexical content of abstracts. We propose a comparative study of abstracts associated with preprints issued in response to the pandemic relative to abstracts produced during the closest pre-pandemic period. We show that with the increase (on average and in percentage) of positive words (especially *effective*) and the slight decrease of negative words, there is a strong increase in hedge words (the most frequent of which are the modal verbs *can* and *may*). Hedge words counterbalance the excessive use of positive words and thus invite the readers, who go probably beyond the 'usual' audience, to be cautious with the obtained results. The abstracts of preprints urgently produced in response to the COVID-19 crisis stand between uncertainty and over-promotion, illustrating the balance that authors have to achieve between promoting their results and appealing for caution.

Keywords: abstract, academic writing, COVID-19

INTRODUCTION

The COVID-19 pandemic (declared on 11 March 2020 by the World Health Organization [WHO]) has prompted intensive research through which researchers and clinicians are actively trying to understand the biology of the virus, the dynamics of its transmission, the development of symptoms, diagnostic tests, drugs, vaccines or the economic and social consequences of the pandemic. Be it with Ebola, H1N1, Zika and SRAS, researchers have already encountered health crisis situations (Xing et al., 2010; Zhang et al., 2020), but, in terms of scholarly communication, the coronavirus seems to have 'broken the mould' of traditional publishing (Kupferschmidt, 2020), with the massive adoption of preprints. Even if there is no common definition of what is a preprint (Chiarelli et al., 2019), we consider it as a potential article that has not (yet) gone through the peer-review process and is already freely accessible to the public.

Since the beginning of the pandemic, preprint servers have experienced a rush after the SARS-CoV-2 outbreak: for example, medRxiv saw an increase from around 200 posts per month in

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

January 2020 to around 2,000 posts in May 2020, and around 70% of all posted manuscripts from March to September 2020 related to the pandemic (Pool, 2020). Even some mainstream media mentioned the 'frenzy' of preprints (Morin, 2020) whilst preprint server managers were testifying that they were drowning in a 'tsunami' of preprints (Brennan, 2020). Looking beyond submissions, bioRxiv and medRxiv also experienced massive increases in attention (measured by downloads and citations) as stated by Hindle (2020), the bioRxiv Content Lead.

Many studies have shown that, under normal circumstances, more and more researchers tend to exaggerate the benefit of their research and use a promotional discourse in order to be published, to advance in their career or to obtain funding (Berkenkotter & Huckin, 1995; Cao et al., 2020; Caulfield & Condit, 2012; Vinkers et al., 2015). We explore whether the authors are increasingly resorting to positive words and persuasive linguistic devices during the COVID-19 crisis, in order to be visible or noticeable in the midst of this unprecedented volume of preprints. In an attempt to answer this question, we propose hereafter a lexical analysis of the abstracts of these preprints. Indeed, the abstract (as well as the title and keywords) is the main entry points into scientific literature. Besides providing a summary, it draws attention to the most important information of the document and should convey as much new information as possible. It is also a means of catching the reader's attention and inviting him to read further. Orasan (2001) states the abstract is not a 'mirror' of the document but is instead intended to draw attention to the most important information of the document it is supposed to summarize. The abstract is therefore considered to be a promotional genre, with many optimistic phrases intended to convince the reader of the importance of the work presented by the authors (e.g., 'it could be effective…', 'our findings suggest promising effects on…', 'this present study could provide a novel insight into…').

We propose a comparative study of the abstracts associated with preprints issued in response to the beginning of the COVID-19 pandemic (from 1 January to 12 April 2020) with those from the closest pre-pandemic period, produced in 2019 (i.e., from 1 January to 31 December 2019). The construction of a corpus of preprint abstracts enabling this comparison is described in the Methods and Data section and should also be considered as a contribution of the study, especially as it has been made available for further research (Bordignon et al., 2021). Based on this unique corpus, we can therefore propose a comparative analysis of the abstracts by evaluating the use of different types of words: positive, negative and hedge words.

Key points

- The COVID-19 crisis influences academic writing as we observe changes in authors' lexical strategies.
- Authors make a greater use of positive words (e.g., *effective*, *novel* or *promising*) in the abstracts of preprints during the COVID-19 crisis.
- There is a strong increase in hedge words (the most frequent of which are the modal verbs *can* and *may*).
- Hedge words counterbalance the excessive use of positive words and thus invite the 'usual' audience to be cautious with the results.
- The abstracts of preprints urgently produced in response to the COVID-19 crisis stand between uncertainty and over-promotion.

BACKGROUND

The abstract as a promotional genre

Bhatia (1993) suggests the abstract is meant to give the reader an exact and concise knowledge of the full article, whilst Martín (2003) goes further, stating that an abstract has a 'time-saving' function. But Berkenkotter and Huckin (1995) contend that the abstract is essentially a promotional genre where authors underline their most central claims with the aim of gaining readers' attention and persuading them to read on (Sun & Hyland, 2001). Cao et al. (2020) showed that positive words are more frequently used in abstracts than in full texts. They consider the abstract as the point at which the reader needs to be 'hooked'. This is what triggers the overuse of positive terms (Vinkers et al., 2015), and overstatements in abstract conclusions (Boutron, 2020; Shinohara et al., 2017). Moreover, both in Berkenkotter and Huckin's (1995) study of abstracts of articles published between 1944 and 1989 and in Vinkers' more recent study of PubMed abstracts published between 1974 and 2014, the tendency to over-promote results was seen to increase, leading the former authors to conclude that 'today's scientists seem to be promoting their work to a degree never seen before' and the latter that researchers 'assume that results and their implications have to be exaggerated and overstated in order to get published'.

Our present study was initially inspired by that of Vinkers et al. (2015) on the use of positive words; however, although this previous work was a diachronic study, it did not focus on periods of crisis.

From promotion to over-promotion

Many researchers have used the marketing metaphor and a lexicon borrowed from this field to talk about 'hyping science' and promotion in academic discourse: 'niche', 'positioning' and 'boosterism' (Swales, 2004; Swales & Feak, 2009), 'to sell/selling' (Bhatia, 1993; Samar et al., 2014), 'marketization' (Nerlich, 2013), and 'quasi advertising discourse' (Lindeberg, 2004). Indeed, beyond promotion, there is the idea of glamorizing results or making one's research methodology or findings look 'sexier' and more ground breaking than they actually are (Millar et al., 2019), emphasizing the 'importance, uniqueness and novelty' of one's

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

research (Vinkers et al., 2015), or 'beautifying' the methods used (Boutron, 2020).

Dodds et al. (2015) showed that the words of natural human language possess a universal positivity bias and that people tend to use positive words more frequently. One of the advantages of exaggerating is that it engages readers and draws attention to the more important features of a study (Caulfield & Condit, 2012; Millar et al., 2019). Nerlich (2013) distinguishes between 'honest hype' and 'politicized hype', where moderation and modesty seem impossible, such as in the reporting of climate science results. Science hype is often seen as damaging in that it contributes to exaggerating the benefits of research and underplaying the costs and risks of science and its technological products (Caulfield & Condit, 2012). It inflates public expectations and results in a loss of trust in a given technology or research avenue if promises are not kept (Joly, 2015; Rinaldi, 2012), it biases the reader's interpretation and leads to a dissemination of misinterpreted results in press releases and news stories, with a possible effect at a larger scale on patients and the general public (Boutron, 2020).

Many studies have identified linguistic and rhetorical devices used by authors of scientific texts (whatever the academic genre) that lead to an exaggerated promotion of their results or methods, that is:

- misleading use of causal language (a clause or phrase where one event, state, action or entity is explicitly presented as influencing another) in publications when no statistical evidence or clinical knowledge exist to establish any causal relationship (Rubin & Parrish, 2007; Thapa et al., 2020);
- self-citation as a rhetorical strategy for emphasizing a writer's contribution (Hyland, 2001);
- heavy use of *I* and *we* (Gragson & Selzer, 1990; Harwood, 2005);
- use of adjectives imposing subjective or positive value like *important*, *critical*, *original* (V. Fraser & Martin, 2009; Vinkers et al., 2015);
- use of 'drama' words (that is adverbs such as *importantly*, *surprisingly*, *dramatically*, *strongly* or *markedly*) (Wheatley, 2014);
- use of promotional elements and 'appeals' to claim centrality and therefore persuade the readers of the importance, worthiness and significance of a research study (Lindeberg, 2004; Swales, 2004; Wang & Yang, 2015);
- spinning, that is 'a specific reporting that fails to faithfully reflect the nature and range of findings and that could affect the impression that the results produce in readers, a way to distort science reporting without actually lying', (Boutron, 2020).

Publishers have tried to curb what is considered as a form of misconduct by attempting to establish guidelines for formatting the presentation of results (e.g., the CONSORT statement for Randomized Controlled Trials (Hopewell et al., 2008), the PRISMA statement (Beller et al., 2013; Liberati et al., 2009) for abstracts of systematic reviews, and Ethical Principles for Medical Research Involving Human Subjects [World Medical Association, 2001]), or even to prohibit the inappropriate use of exaggerating words (e.g., *ACS Catalysis* journal [Scott & Jones, 2017]).

Caulfield and Condit (2012) demonstrate that most of the major entities involved in public communication about science (genomics in this case study), including researchers, commercial forces, research institutions, the media and general public, participate in hyping, which is accelerated via pressure to publish, commercialize and translate work.

Some studies (Berkenkotter & Huckin, 1995; Lindeberg, 2004; Shehzad, 2010) have focused on over-promotion of results in specific research article sections. They show that the most promotional parts are titles, introductions and discussions, but also the abstracts, which has led to specific studies. None of them considered that promotion could be more important in times of crisis. This is one of the points we want to demonstrate in our study.

Hedge words in academic writing

In contrast to over promotion, hedging is a procedure of argumentation acting as a 'face-saving' strategy when claiming the centrality of findings (Pérez-Llantada, 2008). Indeed, according to Hyland (1996b), hedges allow writers to tone down their statements and thereby strengthen arguments by admitting limitations or uncertainties. Hedging even led him to raise the question of whether authors are 'writing without conviction'.

Hedge words (like *possible*, *quite*, *might*, *perhaps*, *will*, *may*, *suggest*) represent a weakening of a claim, show doubt and indicate that information is presented as opinion rather than accredited fact (Hyland, 1998). They aim to persuade readers (Hewings, 2001), and give them the right to participate in the ratification of knowledge (Hyland, 1996a), by soliciting acceptance (Myers, 1989). These words, deliberately denoting vagueness, imprecision (Hewings, 2001) or fuzziness (Lakoff, 1973), also belong to common standards of permissible imprecision (Dubois, 1987; Myers, 1996). Lewin's survey (2001) shows that authors consider uncertainty to be a better representation of the truth and a more honest representation of data.

Hyland (1996b) provides a list of the different lexical and strategic devices that are used to express hedging:

- modal verbs (*would*, *may*, *could*), which are the most used uncertainty makers (Bongelli et al., 2019);
- epistemic lexical verbs (*indicate*, *suggest*, *appear*);
- epistemic adjectives (*unlikely*, *likely*, *possible*, *most*);
- epistemic adverbs (*apparently*, *probably*, *relatively*, *quite*, *almost*, *usually*).

Many studies have reviewed the use and function of hedge words in the scientific literature: for example, Lindeberg (2004) addressed them in comparison with other sections of the paper and Gillaerts and Van de Velde (2010) proposed a historical approach. Based on a quantitative study of corpora of abstracts in applied linguistics from 1982 to 2007, they show that longer,

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

and more recent research article abstracts contain more 'factual' material, not more metadiscourse. The results suggest that even in highly condensed genres such as research abstracts, interpersonal considerations are pervasive and exhibit striking diachronic dynamics. Other studies have shown differences across languages (Hu & Cao, 2011; Martin-Martin & Burgess, 2004; Vassileva, 2001), between native speakers and non-native speakers (Hinkel, 2005; Lee, 2007), between students and expert writers (Aull et al., 2017; Doró, 2014), and across disciplines (Hyland, 1998). Milanović and Milanović (2010) carried out a study on softening the words of the economic crisis, although this was focused on journalistic discourse published on the internet rather than scientific literature. As far as we know, no study has yet focused on the use of hedge words in the scientific literature produced in times of crisis (or identified as such).

METHODS AND DATA

Figure 1 shows the three sequences of the study workflow:

- data collection:
  - constitution of a corpus of preprints issued in response to the COVID-19 pandemic (COVID-19 corpus),
  - delineation of a corpus of preprints produced in 2019, the closest pre-pandemic period (Pre-pandemic corpus),
- automatic identification of positive, negative and hedge words with CorTexT Manager (CorTexT Manager Documentation, 2020),
- data analysis in Excel (to create the source spreadsheets) and Tableau (a data visualization software we used to query the spreadsheets, make our calculations and generate the graphs presented below, see www.tableau.com).

Corpus delineation

The starting point for our method was N. Fraser's corpus (2020), which is a corpus of preprints focused on the COVID-19. Since our aim was to identify possible differences between the writings produced during the crisis compared with those produced previously, the delineation of the control corpus was as critical as it was challenging. Furthermore, we also needed to be able to technically retrieve abstracts. As a consequence, although OSF preprints appears as an important server in N. Fraser's corpus (7th in terms of volume), we had to give up collecting abstracts from this server as they are poorly indexed in both Dimensions (www.dimensions.ai) and Lens (www.lens.org). After several tests and data retrieval attempts, we were able to build a corpus from the following seven servers: SSRN, arXiv, medRxiv, bioRxiv, Research Square, Preprints.org, and ChemRxiv. We expanded the choice of servers beyond the biomedical community to make comparative analyses between servers, and try to draw conclusions regarding disciplines and research communities.

Our final corpus was based on 23,957 preprints and their metadata.

COVID-19 corpus

To locate preprints about COVID-19, N. Fraser (2020) used the following query hereinafter referred to as *COVID query*: 'coronavirus' OR 'COVID-19' OR 'sars-cov' OR 'ncov-2019' OR '2019-ncov'. On 12 April 2020, we downloaded N. Fraser's dataset to retrieve the abstracts and other metadata of 3,341 preprints deposited since 1 January 2020 on the seven different preprint servers mentioned above.

The pre-pandemic corpus as the control corpus

In order to be sure to base our study on comparable corpora, the control corpus also contained preprints taken from the same preprint servers as the abstracts and for which we tried to delineate the same perimeter. Indeed, if we wanted to determine a shift in writing practices with the advent of the pandemic in the papers released since 1 January 2020, we needed to find papers from 2019 that dealt with similar subjects and which we could be almost sure were written by the same communities. We therefore developed a strategy for each server depending on the bibliographic databases that index them and the quality of the data available:

• SSRN

We ran the COVID query in Dimensions over the year 2020 and identified the two most important sub-disciplines (Research categories): 1,103 Clinical Sciences and 1,117 Public Health and Health Services. We applied this limitation to all preprints indexed in Dimensions for the year 2019.

• arXiv

We ran the COVID query in Dimensions over the year 2020 and identified the main sub-discipline, 1,117 Public Health and Health Services. We applied this limitation to all preprints indexed in Dimensions for the year 2019.

• medRxiv

Since medRxiv is a disciplinary server, we considered all preprints indexed in Dimensions for 2019 to be within the same scope.

• bioRxiv

We ran the COVID query in Dimensions over the year 2020 and identified the two main sub-disciplines (1,108 Medical Microbiology and 0601 Biochemistry and Cell Biology). We applied this limitation to all preprints indexed in Dimensions for the year 2019.

• Research Square

We ran the COVID query in Dimensions over the year 2020 and identified the main sub-discipline: 1103 Clinical Sciences. We applied this limitation to all preprints indexed in Dimensions for the year 2019.

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

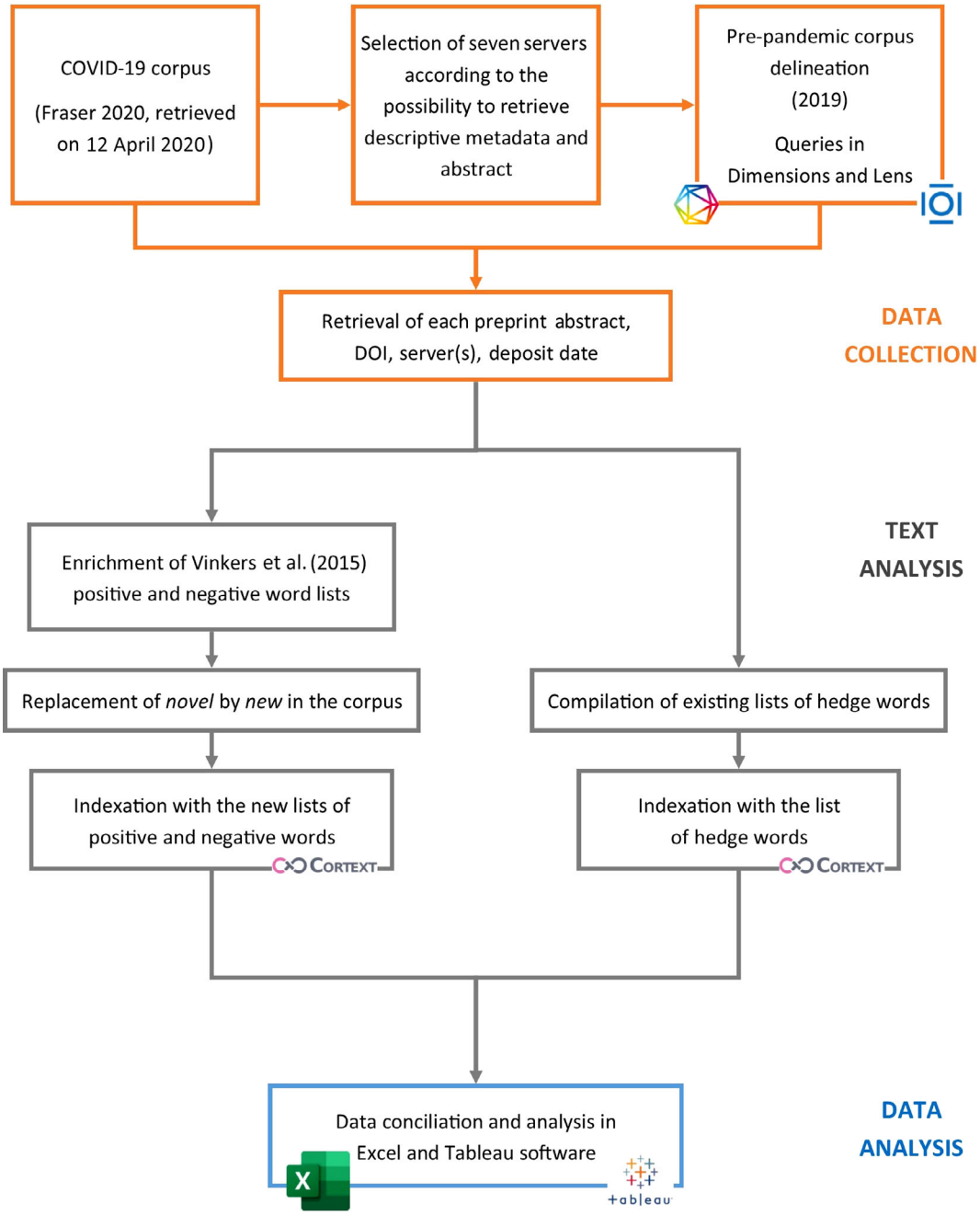

FIGURE 1 Study workflow.

• Preprints.org

As Dimensions does not index Preprints.org, we used Lens and narrowed our search to the 2019 results by limiting these to the Fields of Study of Virology, Medicine, Biology and Coronavirus.

• ChemRxiv

Dimensions poorly indexed ChemRxiv preprints, particularly 2019 preprints. Lens indexes them but does not calculate the Field of Study. In consequence, we selected all the 2019 preprints of ChemRxiv.

All the resulting queries are available online with our data (Bordignon et al., 2021). We retrieved the abstracts semi-manually from Dimensions or via the Lens database when available. Some papers had been deposited on several servers; they were counted each time in order to provide a more accurate view per server and, since as they are few in number (1.06% only), this does not modify the main trends in any way.

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

TABLE 1 Distribution of preprints by server for the pre-pandemic period and during the pandemic.

Servers	Pre-pandemic corpus	COVID-19 corpus	Total
arXiv	1,408	436	1,844
bioRxiv	7,784	374	8,158
ChemRxiv	4,578	95	4,673
medRxiv	897	1,236	2,133
Preprints.org	371	234	605
Research Square	3,485	336	3,821
SSRN	2,259	723	2,982
All	20,616	3,341	23,957

Table 1 shows the distribution of the preprints by server for the COVID-19 corpus and the control (pre-pandemic) corpus.

By using the method described above, we ensured that we were comparing similar linguistic materials. It did not, however, allow us to control volumes and thus to have a uniform number of preprints per server. The weight of each server before and during the pandemic is completely different, which is certainly an interesting piece of information in its own right, but out of the scope of our research. Therefore, we will concentrate on the results per server rather than overall results, which could be biased.

Identification of positive/negative adjectives and hedge words

Since Vinkers et al.'s study (2015) was partly our inspiration for this work, we first reviewed the lists they had proposed. In order to start without preconceived ideas and to be sure to include as many positive words as possible, even those that we could not have thought of, since their use is unusual in the scientific literature, we used a dictionary of synonyms to increase the lexical range. With the addition of synonyms, we were able to broaden the range of tonality of these adjectives and to integrate very strongly positive words such as *astounding*, *bewildering*, *breathtaking*, *impressive*, *incredible*, *striking* and *stunning* to check for their occurrence in the corpus. With the same method, we added *inefficient* to Vinkers et al.'s negative word list. We thus obtained the two following lists:

Positive word list

amazing, assuring, astonishing, astounding, bewildering, breathtaking, bright, creative, effective, efficient, encouraging, enormous, excellent, extraordinary, fantastic, favourable, favourable, groundbreaking, hopeful, impressive, incredible, ingenious, innovative, inspiring, inventive, marvellous, miraculous, novel, outstanding, phenomenal, prodigious, prominent, promising, reassuring, remarkable, robust, satisfying, sensational, spectacular, staggering, startling, striking, stunning, stupefying, stupendous, substantial, successful, supportive, unique, unparalleled, unprecedented, wonderful, wondrous.

Negative word list

detrimental, *disappointing*, *disconcerting*, *discouraging*, *disheartening*, *disturbing*, *frustrating*, *futile*, *hopeless*, *impossible*, *inadequate*, *ineffective*, *insignificant*, *insufficient*, *irrelevant*, *mediocre*, *pessimistic*, *substandard*, *unacceptable*, *unpromising*, *unsatisfactory*, *unsatisfying*, *useless*, *weak*, *worrisome*, *inefficient*.

The Vinkers et al.'s list of negative words contains the word *insignificant* whilst *significant* surprisingly does not appear in the list of positive words. However, we do think that announcing in the abstract that the results presented in the article are significant is a way of promoting one's work. Perhaps Vinkers et al. realized that including *significant* in the list of positive words would be difficult due to the polysemy of the word and its use in a negation. To overcome this problem, we applied a specific procedure to the adjective significant. Indeed, significant does not always express something positive or promising, especially about a result. It can express a simple order of magnitude or importance, without any particular tonality, as in the following examples

(1) *Significant research has shown that UV-C exposure is an effective disinfectant for a range of bacteria and viruses, including coronaviruses*

(2) *Targeted immunization or attacks of large-scale networks has attracted significant attention by the scientific community*

(3) *One significant symptom that occurs in the COVID-19 is Tachypnea*

(4) *Significant gaps persist in our knowledge of COVID-19 epidemiology*.

Additionally, in a negative sentence, *significant* can be used to express non-significant results.

Therefore, in order to mitigate this bias and maximize the likelihood that this adjective will be used as a positive word, we identified particular sentence patterns that may indicate whether the results are significant. Additionally, rather than individually indexing the word *significant* in abstracts, we indexed the following phrases:

were significantly improved, was significantly improved, was statistically significant, were statistically significant, were statistically significant, was significant, were significant, significant advantage, has significant promise, have significant promise, are significantly improved, is significantly improved, is statistically significant, are statistically significant, are statistically significant, is significant, are significant, had significant promise, had significant promise.

Symmetrically, we designed the following list to enrich the list of negative words:

was not statistically significant, were not statistically significant, was not significant, were not significant, were not significantly improved, was not significantly improved, is not statistically significant, are not statistically significant, is

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

not significant, are not significant, are not significantly improved, is not significantly improved.

From our first tests, we rapidly noticed the occurrence of unwanted expressions containing the adjective *novel*, such as *novel coronavirus* or *novel COVID*. Since *novel* is a typical positive word whose improper use reveals a risk of over reporting results, we modified these expressions in the corpus before starting the indexing process. We replaced:

- *novel coronavirus* and *novel corona virus* by *new coronavirus*
- *novel bat coronavirus* by *new bat coronavirus*.
- *novel COVID* by *new COVID*
- *this novel virus* by *this new virus*

The aim was not to judge the tonality of the whole abstract, just to identify words and phrases with positive and negative meaning.

To identify hedge words in the abstracts, as there is no single comprehensive list on which to draw, we built our own on the basis of other lists, examples and definitions commonly used in related literature. We assumed that in scientific texts, hedge words are words that introduce vagueness or imprecision whereas accredited and accurate facts are expected. This is why we have included, for example, words expressing approximation (*about*, *often*) or *will* which can sometimes be associated with a hypothetical condition. We are aware that it may be considered imperfect, but it should be remembered that our aim here is to produce a comparative analysis and that the crucial point is that the same list is used for the two corpora being compared. Our list consists then of the following 111 terms and their possible inflections and variants:

a bit, *about*, *actually*, *allege*, *almost*, *always*, *and so forth*, *apparently*, *appear*, *approximately*, *around*, *assume*, *at least*, *basically*, *be sure*, *believe*, *bunch*, *can*, *certainly*, *clear*, *conceivably*, *consider*, *consistent with*, *could*, *couple*, *definite*, *diagnostic*, *don't know*, *doubt*, *effectively*, *estimate*, *evidently*, *fairly*, *few*, *find*, *frequently*, *generally*, *guess*, *hopefully*, *improbable*, *in general*, *in my mind*, *in my opinion*, *in my understanding*, *in my view*, *inconclusive*, *indicate*, *kind of*, *largely*, *like*, *likely*, *little*, *look like*, *mainly*, *many*, *may*, *maybe*, *might*, *more or less*, *most*, *mostly*, *much*, *must*, *my impression*, *my thinking is*, *my understanding is*, *necessarily*, *occasionally*, *often*, *overall*, *partially*, *perhaps*, *possibility*, *practically*, *presumable*, *pretty*, *probability*, *quite*, *rarely*, *rather*, *read*, *really*, *roughly*, *say*, *seem*, *seldom*, *several*, *should*, *so far*, *some*, *somebody*, *somehow*, *someone*, *something*, *sometimes*, *somewhat*, *somewhere*, *sort of*, *speculate*, *suggest*, *suppose*, *surely*, *tend*, *think*, *understand*, *unlikely*, *unsure*, *usually*, *virtually*, *will*, *would*.

CorTexT Manager is a text-mining tool that enables term extraction and lexical analyses: the three lists of words (positive, negative and hedge) were indexed in the abstracts of the corpus. NLP tools that are developed in CorTexT Manager (CorTexT Manager Documentation, 2020) allow to identify not only single terms but also multi-terms (*n*-grams).

We identified for each preprint which of these words appear in the abstract. The same abstract may of course contain the three types of words. We then exported the data into Excel and Tableau and calculated the percentage of abstracts that contained at least one of these words (for each type of words, separately), and also the number of times each word was used on average per abstract. We also present relevant examples of their use to illustrate the significance of the results.

RESULTS

Overview of results

Figure 2 gives a comprehensive synthesis of our results: it shows both (1) general trends that contribute to a better qualification of the writing of scientific results in times of crisis, and (2) differences amongst the servers that reveal writing patterns that are specific to scientific communities. One hundred percent corresponds to the total number of abstracts on each server. An abstract can contain positive, negative and hedge words at the same time. In this case, it is included for each type of words.

There is an increase in the number of preprints using positive and hedge words, whilst the number of papers with negative words decreases very slightly. At the same time, the average number of positive and hedge words per abstract also increases, whilst the average of negative words also decreases very slightly.

(Over-)promoted results

As indicated above, both the number of abstracts containing at least one positive adjective and the average number of positive adjectives per abstract tend to increase between the two periods. Figure 3 shows the 10 terms that occur at least once within the three most frequent terms per server and over each of the two time periods.

At the early stage of the pandemic, the adjective *effective* appeared in many more abstracts than in similar scientific output produced just a few weeks/months earlier. Amongst all positive adjectives, *effective* showed the highest increases on all servers in terms of the proportion of abstracts containing this word at least once: *effective* appeared in 4.5% (arXiv) to 12.6% (Research Square) of the abstracts before the pandemic, and then in 14.7% (arXiv) to 22.1% (ChemRxiv) of the abstracts in the COVID-19 corpus. The following sentences, (5), (6) and (7) are examples of uses of *effective*:

(5) *The resulting control measure could be an* <u>*effective*</u> *strategy to control local and re-emerging outbreaks of COVID-19*.
(6) *These results corroborate that mass social isolation is a highly* <u>*effective*</u> *measure against the dissemination of SARS-CoV-2, as previously suggested*.
(7) *Initial in vitro pharmacology with EA.hy926 and HUVEC endothelial cells indicated that these bivalent ligands are* <u>*effective*</u> *binders of* $\alpha V\beta 3$ *and potent agonists of PAR2*.

The adjective *novel* was also frequently used, and its usage increased between the periods on all servers except medRxiv.

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

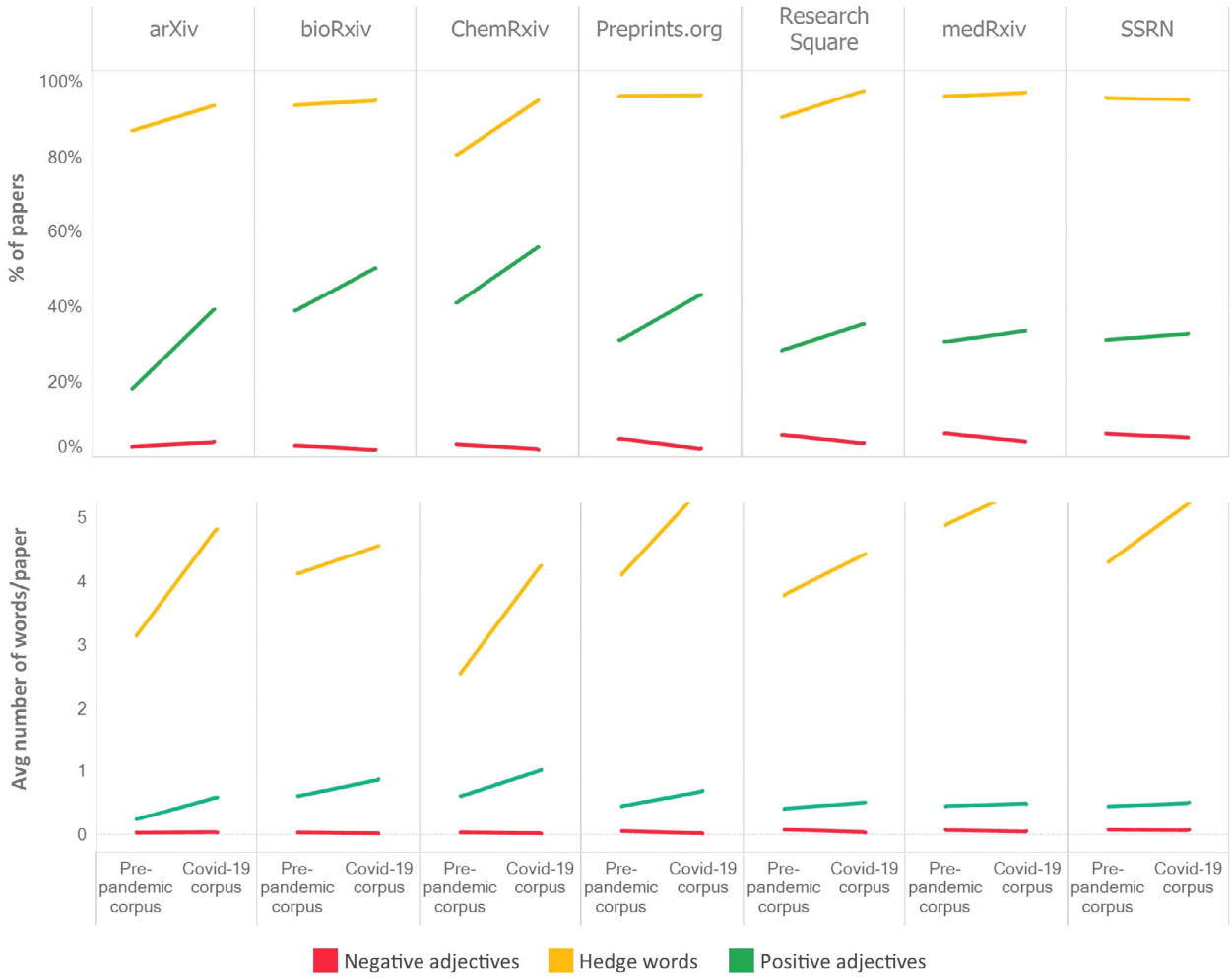

FIGURE 2 Results synthesis per preprint server: percentage of paper abstracts with positive/negative adjectives and hedge words, average number of positive/negative adjectives and hedge words per abstract. Comparison before and during the COVID-19 crisis.

Vinkers et al. (2015) showed that the use of *novel* in Pubmed abstracts increased dramatically between 1974 and 2014, and even predicted that it would appear in all abstracts by 2,123. We should remember that in order to have the best chance to identify *novel* as a positive word and not as the announcement of a new event, we took care to remove expressions like *novel coronavirus* from the corpus. This approach enabled us to count uses in which *novel* related to the positive aspect of certain advances, as in the following examples:

(8) *Here we use a novel approach to analyse all currently available aminoacyl-tRNA synthetase structures, which cover the recognition of all proteinogenic amino acids across all kingdoms of life*
(9) *This present study could provide a novel insight into the molecular basis of overgrowth Syndrome*
(10) *Therefore, targeting TORC1 signalling and endoplasmic reticulum stress pathways may be useful in developing novel targets for antifungal drugs.*

Vinkers et al. (2015) do not seem to have taken this precaution, which surely led them to overestimate their assessment and projection, although this likely had no implications for the overall upward trend of this word.

In many contexts, when *novel* is used as an epithet, it is almost interchangeable with *effective*, or at least the reader infers the effectiveness of the novelty proposed by the authors; see the following examples:

(11) *Gene mania network helps to design the novel drugs and diagnosis*
(12) *We believe that the results of this study could offer a novel avenue for drug development against CoVID-19*
(13) *Based on this information, we propose guidelines to develop novel N protein-based antiviral agents that target CoVs*
(14) *We have developed a novel two-stage simulation model to simulate the spatiotemporal changes in the number of COVID-19 cases and estimate the future worldwide risk.*

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

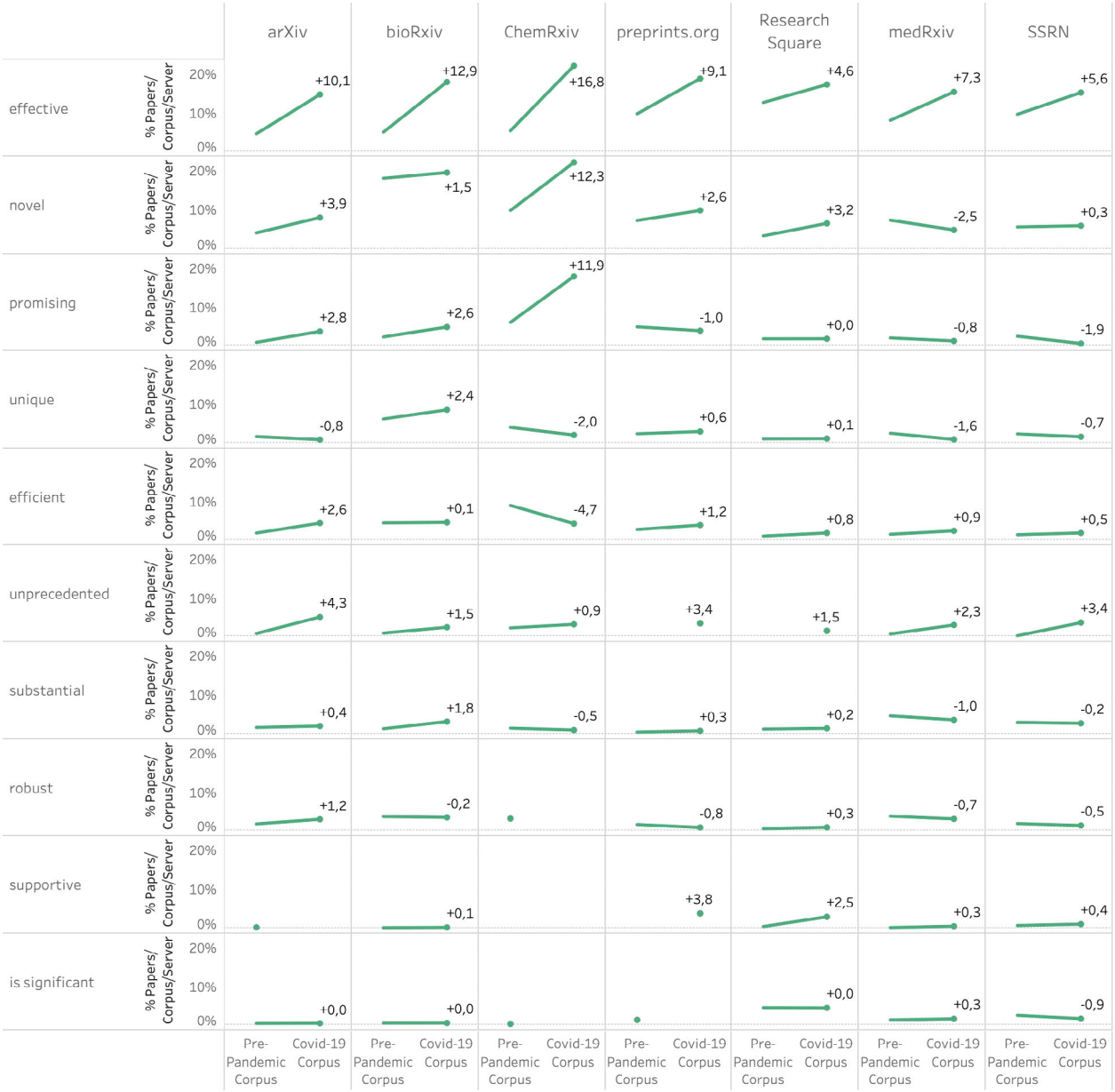

FIGURE 3 Evolution of the use of positive words (Top 10): Percentage of papers per period and per server.

Goodman (1993), who had already detected an increase in its frequency in the titles of scientific papers, wrote that *novel* was in danger of becoming no more than the synonym of *new*. Whilst *new* can be used for an original idea, *novel* evokes a sparkling and ingenious one. The study we are conducting might throw light on the weak signal that *effective* will soon supplant *novel* to qualify the latest medical advances. The use of this word reflects a stronger focus on proof of concept (whose purpose is to verify that some concept or theory has the potential of being used), rather than on the priority of the announced results.

As far as *promising* is concerned, ChemRxiv is the server on which the increase in usage was the greatest (+11.9 points). It was more moderate on arXiv (+2.8 points) and bioRxiv (+2.6 points) and the variations were not very significant on the other servers.

The word *promising* was used in 0.9% (arXiv) to 6% (ChemRxiv) of abstracts in the pre-pandemic corpus, and in 0.6% (SSRN) to 17.9% (ChemRxiv) of abstracts in the COVID-19 corpus. Here are a few examples:

(15) *In conclusion, LXRβ, by restoring the differentiation of HSCs, may be a promising therapeutic target for liver fibrosis without the adverse side-effects of LXRα activation*

(16) *Overall, our findings suggest the promising effects of the QFPD decoction for COVID-19 treatment*

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

(17) *Support vector machine demonstrated a promising detection accuracy after 32 features were detected to be significantly associated with the COVID-19 severeness.*

Figure 2 shows a general upward trend in the use of positive adjectives, but Fig. 3 shows even sharper increases for *novel*, *effective* and *promising* in the abstracts of ChemRxiv preprints.

Despite the fact that preprints have the advantage of rapid publication, chemists have been reluctant to produce them (ACS Publications, 2020; Warr, 2003). In the COVID-19 pandemic, however, the preprints from ChemRxiv that we analysed here were deposited in the very early stages of the crisis. Chemists undoubtedly tried to adapt to the crisis and enter the global competition by reacting quickly with a concentration of work responding to the health crisis, but they also increased the use of positive adjectives to promote their results.

This study shows that authors sought to be more impactful than usual (i.e., than the pre-pandemic period) in presenting their results by using positive adjectives more often, especially *effective*. Nevertheless, they did not massively resort to using the words *groundbreaking*, *miraculous* or *stunning*, which more clearly imply exaggeration and overstatement, as the following examples demonstrate:

(18) *The results are based on high quality cryo-TEM images, cutting edge synchrotron solution X-ray scattering measurements and state-of-the-art data analysis, using our home developed groundbreaking analysis software, D+*
(19) *We firmly believe that Nature itself provides a simple solution for any complicated problem created in it which motivated us to carry out In Silico investigations on some bioactive natural compounds reportedly found in the fruits and leaves of Anthocephalus cadamba which is a miraculous plant found on the earth aiming to predict the potential inhibitors against aforesaid virus*
(20) *Moreover, as the case history shows, a stunning reduction in deaths from the diseases was accomplished largely through accretive advances—without developing a vaccine, an unambiguous test, or a complete cure.*

Toning down (over-)promoted results

Negative adjectives were already very infrequent on all servers before the pandemic: from less than 3% of the abstracts on arXiv to about 6% on medRxiv, SSRN and Research Square. It barely exceeded 5% (SSRN) and dropped below 2% (bioRxiv) with the arrival of the pandemic. Only arXiv recorded a slight increase, of +1.1%.

As an indication, here is the list of negative adjectives that are in the top three most frequent adjectives per server, and over each period studied: *insufficient*, *inadequate*, *weak*, *ineffective*, *pessimistic*, *irrelevant*, *is not significant*, *impossible*, *detrimental*. The inclination to use positive words during the crisis was, therefore, accompanied by the decrease in the frequency of negative words.

Negative words (such as *inadequate*) rarely describe the findings like in (21), but rather are used at the beginning of the abstract to point out a research gap and to prepare the reader for a new claim by the author as (22) shows:

(21) *False positives occurred due to incomplete exams and inadequate bowel preparation.*
(22) *However, social distancing in the wake of COVID-19 has been frustratingly slow and inadequate.*

Nevertheless, negative words are so infrequent that it is difficult to draw definitive conclusions about their use and contribution to toning-down the results, as opposed to hedge words. Figure 2 shows that with the increase (on average and in percentage) of positive adjectives and the relatively small decrease of negative adjectives, there is a very strong increase in hedge words. It also shows that hedge words are very frequent (in at least 80% of abstracts); here some examples:

(23) *Regulating Temperature and Humidity level can provide drastic results to stop and arrest the outbreak.*
(24) *Enterocytes may act as a conserved cell reservoir for coronaviruses, which may be partially explained by the Red Queen hypothesis*
(25) *These data raise the possibility that the effects of serotonin on satiety may have their origins in feedback, homeostatic metabolic responses from the periphery*
(26) *We present a comprehensive analysis of somatic mutations in miRNA genes and show that some of these genes are mutational hotspots, suggesting their potential role in cancer.*

Figure 4 shows that the modal verbs *can* and *may* are the most frequent hedge words and increase between the periods. In addition to an increase in the proportion of abstracts containing these words, the average frequency per abstract rises sharply for the most frequent hedge words.

DISCUSSION

The increase in the number of hedge words per abstract contributes to fuzziness in the reader's mind. However, like Salager-Meyer (2017), we believe they reveal the writers' true state of knowledge. This is also what Lewin (2001) showed in her survey, which reveals that, in an authors' view, the use of modals is a reflection of 'the truth'. The COVID-19 corpus gathers in fact the very first published abstracts of this crisis. The research they present, thus, consists of the first attempts to provide scientific answers. This is one of the major differences with the pre-pandemic corpus which grew out apart from any urgency generated by a health crisis.

Even though medical journals have managed to greatly accelerate their publication processes to make them nearly twice as fast for coronavirus-related articles (Horbach, 2020), researchers

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

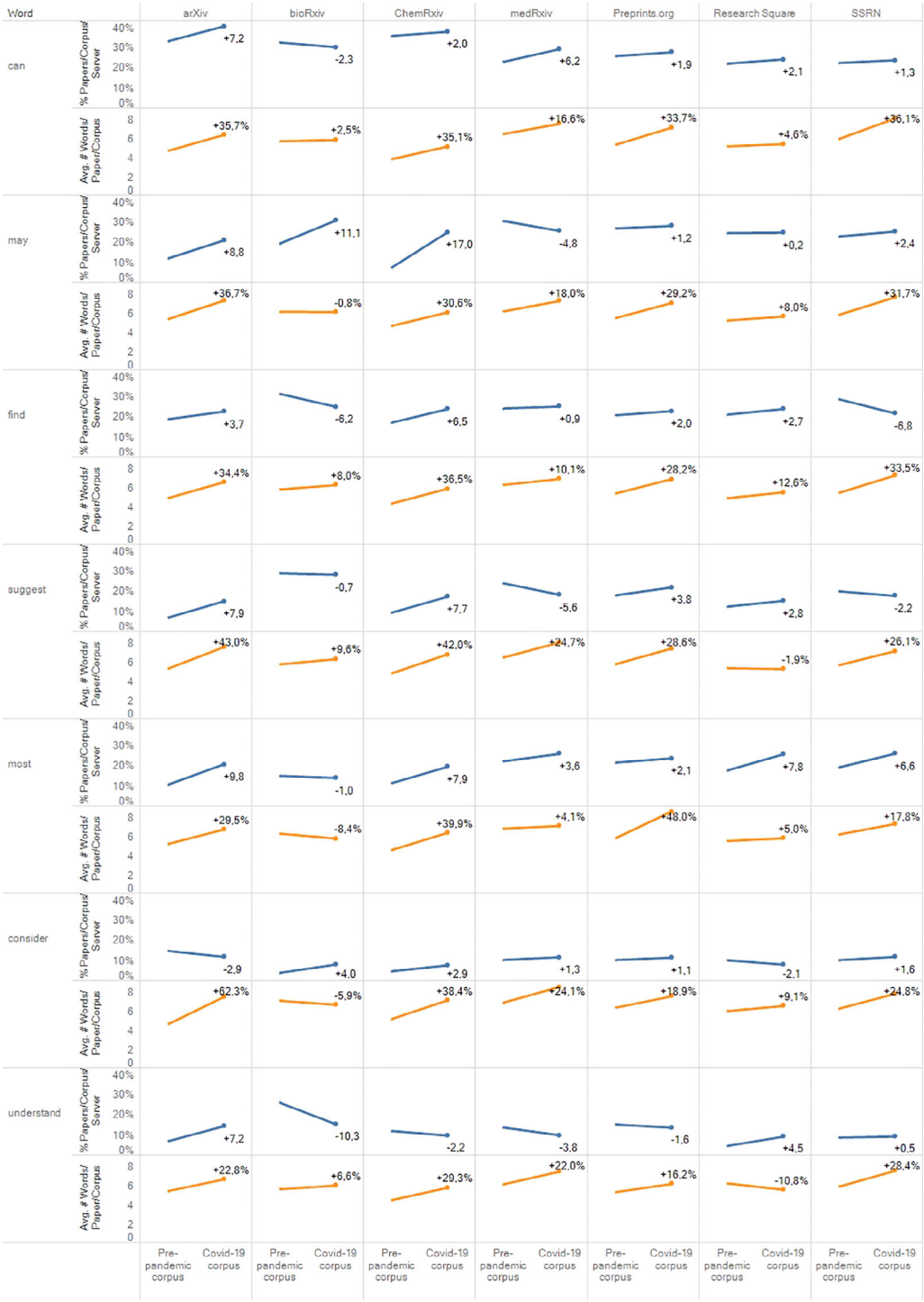

FIGURE 4 Evolution of the use of the most frequent hedge words (Top 7): Percentage of papers and average number of hedge words per period and per server.

have turned overwhelmingly to preprint servers to share their research. After many unsuccessful attempts, the fast and recent development of preprint infrastructures (for some equipped with downloading statistics or altmetrics) has encouraged researchers to deposit their manuscripts on these servers. We believe that researchers clearly perceived from the very beginning of the crisis that the results they were reporting would be in the spotlight and, therefore, the way in which they were presented would be crucial. Choosing to present one's results on COVID-19 by depositing a paper on a preprint server means exposing oneself

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

to a circle of readers far beyond one's peers, whilst only a few months earlier, the likelihood of being read by a journalist, a policy maker or any citizen was minimal. After all, the raison d'être of many researchers is to have their work read and discussed, and to be useful, especially in a pandemic of this magnitude.

Hedge words counterbalance the excessive use of positive words and thus invite the readers, including peers, politicians, journalists and citizens, to be cautious. It is, therefore, rather reassuring to find them in large numbers if one considers scepticism as a fundamental principle of the scientific process. Moreover, in the context of a health crisis, it is even courageous of the authors to put them if one considers, like MacGregor et al. (2020), that uncertainty creates particular discomfort amongst public health professionals in discussions of 'risk reduction' messages directed at the general public and the media.

As for the increase in the frequency of positive words (enhanced by the decrease in the use of negative words), this may be the consequence of two things: on the one hand, the human inclination to raise hope (an inclination accentuated by the seriousness and extent of the crisis), and, on the other hand, the need for researchers to be visible in the mass by contributing encouraging results. From the beginning of the crisis, authors have more frequently used words indicating the validity of their results (*effective*, *efficient*, *robust*) and their potential impact in the fight against the virus (*promising*) and they have also attempted to show the innovative nature of their work to highlighting novelty (*novel*, *unique*) in this competitive context.

It is plausible that the aim would be to attract the attention of funders as invoked by Vinkers et al. (2015), of journals that could host the article version of these preprints, of the media that would contribute to the reputation of the authors, and/or of the politicians, who could choose one way rather than another to inform their decisions.

The study of this 'crisis literature' reveals that (over-)promotion is mitigated by hedging. Our corpus is teeming with such examples:

(27) *This imaging modality <u>could</u> be <u>effective</u> for the diagnosis of enterocolitis associated with COVID-19*.

This combination of promise and uncertainty conveyed by a large number of preprints made available to all may have contributed to the 'infodemic' (i.e., 'over-abundance of information—some accurate and some not—occurring during an epidemic') denounced by the WHO (2020) and to hyping the research. Whilst media hype in times of health crisis can have the virtue of engaging the public with complex scientific issues as suggested by Highfield (2021) about the COVID-19 crisis, the use of exaggeration in preprint abstracts can have detrimental consequences, especially in times of crisis. Abstracts are often the only part of the work that will be read, and over-promotion will raise false expectations that will not be addressed by the reading of the full-text.

Our final point is directed to scientists who wrote preprints: the excessive use of positive words which they try to balance with many hedge words is just a band-aid solution and does not work in favour of an informative abstract. Our recommendation is that authors (including ourselves) should take the time to carefully write their abstracts, and ponder on the use of words so as to avoid the risk of being inaccurate, discrediting the long work that constitutes the scientific article. Even if they are far from perfect, the lists that have been compiled here are a first basis to think about the presence of positive, negative or hedge words in the genre the research article abstract (and preprint abstract in particular) has become.

CONCLUSION AND PERSPECTIVES

A corpus specially created for a contrastive analysis was built with the aim of making a comparison between preprint abstracts produced before and in the early phases of the pandemic. Results suggest that the COVID-19 crisis influences academic writing as we observe changes in authors' lexical strategies. We show that with the increase (on average and in percentage) of positive adjectives (especially *effective*) and the slight decrease of negative adjectives, there is also a very strong increase in hedge words (the most frequent of which are the modal verbs *can* and *may*). Our conclusion is that the scientific literature urgently produced in response to the COVID-19 crisis is an illustration of the balance that authors have to achieve between promoting their results and appealing for caution.

The need to speed up access to results of this research using preprint servers, to be found amongst hundreds of other works on the same topic and to establish anteriority all weigh on researchers and have a strong influence on the way they present their work in the abstracts they make available to the widest number of readers.

On the one hand, the competition, observed through an unprecedented volume of preprints, prompted them to make greater use of positive words to stand out from the crowd and get their work noticed. On the other hand, being aware of the uncertainty surrounding their results, they have nevertheless also increased the use of hedge words. By toning down their discourse, they try to gain community acceptance and keep critics away, but they also invite their readers to be cautious.

This work also shows that the results are different from one server to another. This is important for two reasons: it shows (1) that there are community-centered writing patterns, and (2) it also presents preprints and, more precisely, preprint servers, as possible bases for other types of analyses that would examine communities or even disciplinary distinctions.

The strength of our study is that we considered the peak in production in the very early months of 2020, but also a limitation that calls for work to continue over a longer period of time. Finally, this work could also be supplemented by an analysis of the conversion of preprints into articles to see whether the varying use of positive/negative adjectives and hedge words affects

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

the speed of publication of a preprint article. Based on the examples reviewed, we also suggest a series of interviews with the authors (as did Milard and Tanguy (2018) for in-text citations) that could also inform us about the researchers' writing practices and help us verify or invalidate some of the proposed hypotheses.

AUTHOR CONTRIBUTIONS

Conceptualization: FB, LE and MN; Data curation: FB and LE; Formal analysis: FB; Methodology: FB, LE and MN; Visualization: FB; Writing – original draft: FB; Writing - review & editing: FB, LE and MN.

REFERENCES

ACS Publications. (2020). *Organic Letters Author Guidelines*. https://publish.acs.org/publish/author_guidelines?coden=orlef7

Aull, L. L., Bandarage, D., & Miller, M. R. (2017). Generality in student and expert epistemic stance: A corpus analysis of first-year, upper-level, and published academic writing. *Journal of English for Academic Purposes*, *26*, 29–41. https://doi.org/10.1016/j.jeap.2017.01.005

Beller, E. M., Glasziou, P. P., Altman, D. G., Hopewell, S., Bastian, H., Chalmers, I., Gøtzsche, P. C., Lasserson, T., & Tovey, D. (2013). PRISMA for abstracts: Reporting systematic reviews in journal and conference abstracts. *PLoS Medicine*, *10*(4), e1001419. https://doi.org/10.1371/journal.pmed.1001419

Berkenkotter, C. A., & Huckin, T. N. (1995). *Genre knowledge in disciplinary communication: Cognition/culture/power*. Lawrence Erlbaum Associates.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.

Bongelli, R., Riccioni, I., Burro, R., & Zuczkowski, A. (2019). Writers' uncertainty in scientific and popular biomedical articles. A comparative analysis of the British medical journal and discover magazine. *PLoS One*, *14*(9), e0221933. https://doi.org/10.1371/journal.pone.0221933

Bordignon, F., Ermakova, L., & Noel, M. (2021). A corpus designed to study preprints produced during the Covid-19 crisis and to make comparative studies with the pre-pandemic period. *Mendeley Data*, 1. https://doi.org/10.17632/rn9b93x5d4.1

Boutron, I. (2020). Spin in scientific publications: A frequent detrimental research practice. *Annals of Emergency Medicine*, *75*(3), 432–434. https://doi.org/10.1016/j.annemergmed.2019.11.002

Brennan, M. (2020). So ChemRxiv is currently getting nearly 100k views & downloads *per day* all of a sudden. Trying not to hyperventilate. Twitter. https://twitter.com/Organometallica/status/1246115340570046465

Cao, X., Lei, L., & Wen, J. (2020). Promoting science with linguistic devices: A large-scale study of positive and negative words in academic writing. *Learned Publishing*, *34*, 82–88. https://doi.org/10.1002/leap.1322

Caulfield, T., & Condit, C. (2012). Science and the sources of hype. *Public Health Genomics*, *15*(3–4), 209–217. https://doi.org/10.1159/000336533

Chiarelli, A., Johnson, R., Pinfield, S., & Richens, E. (2019). Preprints and scholarly communication: Adoption, practices, drivers and barriers. *F1000Research*, *8*, 971. https://doi.org/10.12688/f1000research.19619.1

CorTexT Manager Documentation. (2020). *Terms Extraction*. https://docs.cortext.net/lexical-extraction#methodological-background

Dodds, P. S., Clark, E. M., Desu, S., Frank, M. R., Reagan, A. J., Williams, J. R., Mitchell, L., Harris, K. D., Kloumann, I. M., Bagrow, J. P., Megerdoomian, K., McMahon, M. T., Tivnan, B. F., & Danforth, C. M. (2015). Human language reveals a universal positivity bias. *Proceedings of the National Academy of Sciences*, *112*(8), 2389–2394. https://doi.org/10.1073/pnas.1411678112

Doró, K. (2014). Verb tenses and hedging in published and unpublished applied linguistics research paper abstracts. In *Studies in honour of Marianne Nikolov* (pp. 356–369). Pécs: Lingua Franca Csoport.

Dubois, B. L. (1987). "Something on the order of around forty to forty-four": Imprecise numerical expressions in biomedical slide talks. *Language in Society*, *16*(4), 527–541. https://doi.org/10.1017/S0047404500000361

Fraser, N. (2020). Covid19_preprints. *Dataset on Figshare*. https://doi.org/10.6084/m9.figshare.12033672

Fraser, V., & Martin, J. (2009). Marketing data: Has the rise of impact factor led to the fall of objective language in the scientific article? *Respiratory Research*, *10*(1), 35. https://doi.org/10.1186/1465-9921-10-35

Gillaerts, P., & Van de Velde, F. (2010). Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes*, *9*(2), 128–139. https://doi.org/10.1016/j.jeap.2010.02.004

Goodman, N. W. (1993). Paradigm, parameter, paralysis of mind. *British Medical Journal*, *307*(6919), 1627–1629. https://doi.org/10.1136/bmj.307.6919.1627

Gragson, G., & Selzer, J. (1990). Fictionalizing the readers of scholarly articles in biology. *Written Communication*, *7*(1), 25–58. https://doi.org/10.1177/0741088390007001002

Harwood, N. (2005). "*Nowhere has anyone attempted… In this article I aim to do just that*" A corpus-based study of self-promotional I and we in academic writing across four disciplines. *Journal of Pragmatics*, *37*(8), 1207–1231. https://doi.org/10.1016/j.pragma.2005.01.012

Hewings, M. (2001). In M. Hewings (Ed.), *Academic writing in context*. University of Birmingham Press.

Highfield, R. (2021). The value of media hype for science—Embrace the hype. *Chemistry World*. www.chemistryworld.com/opinion/the-value-of-media-hype-for-science/4013552.article

Hindle, S. (2020). @biorxivpreprint: Submissions: >100,000 #preprints, ∼3500 monthly posts, ∼30% of preprints revised Engagement: ∼8 million views, 3—4 million downloads, 152,000 citations, ∼15,000 tweets per week. Twitter. https://twitter.com/HindleSamantha/status/1344744581154402312

Hinkel, E. (2005). Hedging, inflating, and persuading in L2 academic writing. *Applied Language Learning*, *14*, 29–54.

Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., Altman, D. G., Schulz, K. F., Barbour, V., Berlin, J., Boutron, I., Devereaux, P. J., Dickersin, K., Elbourne, D., Ellenberg, S., Gebski, V., Goodman, S., Gøtzsche, P. C., Groves, T., Grunberg, S., … Simera, I. (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and elaboration. *PLoS Medicine*, *5*(1), 48–56. https://doi.org/10.1371/journal.pmed.0050020

Horbach, S. P. J. M. (2020). Pandemic publishing: Medical journals strongly speed up their publication process for COVID-19.

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

Quantitative Science Studies, *1*(3), 1056–1067. https://doi.org/10.1162/qss_a_00076

Hu, G., & Cao, F. (2011). Hedging and boosting in abstracts of applied linguistics articles: A comparative study of English- and Chinese-medium journals. *Journal of Pragmatics*, *43*(11), 2795–2809. https://doi.org/10.1016/j.pragma.2011.04.007

Hyland, K. (1996a). Talking to the academy. *Written Communication*, *13*(2), 251–281. https://doi.org/10.1177/0741088396013002004

Hyland, K. (1996b). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, *17*(4), 433–454. https://doi.org/10.1093/applin/17.4.433

Hyland, K. (1998). Boosting, hedging and the negotiation of academic knowledge. *Text*, *18*(3), 349–382. https://doi.org/10.1515/text.1.1998.18.3.349

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, *20*(3), 207–226. https://doi.org/10.1016/S0889-4906(00)00012-0

Joly, P. B. (2015). Le régime des promesses technoscientifiques. In *Sciences et technologies émergentes. Pourquoi tant de promesses ?* (pp. 31–48). Editeurs des Sciences et des Arts.

Kupferschmidt, K. (2020). 'A completely new culture of doing research.' Coronavirus outbreak changes how scientists communicate. *Science*. https://doi.org/10.1126/science.abb4761

Lakoff, G. (1973). Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, *2*(4), 458–508.

Lee, E. (2007). A study on error analysis & Hedging expressions of medical research abstracts. *English Language & Literature Teaching*, *20*, 47–66.

Lewin, B. A. (2001). From hedging to heightening: Toning down and up in scientific texts · personal pronouns · personal pronouns. *Melbourne Papers in Linguistics and Applied Linguistics*, *1*, 17–28.

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

Lindeberg, A.-C. (2004). *Promotion and politeness. Conflicting scholarly rhetoric in three disciplines*. Åbo Akademi University Press.

MacGregor, H., Ripoll, S., & Leach, M. (2020). Disease outbreaks - navigating uncertainties in preparedness and response. In I. Scoones & A. Stirling (Eds.), *The politics of uncertainty: Challenges of transformation*. Routledge.

Martín, P. M. (2003). A genre analysis of English and Spanish research paper abstracts in experimental social sciences. *English for Specific Purposes*, *22*(1), 25–43. https://doi.org/10.1016/S0889-4906(01)00033-3

Martin-Martin, P., & Burgess, S. (2004). The rhetorical management of academic criticism in research article abstracts. *Text - Interdisciplinary Journal for the Study of Discourse*, *24*(2), 171–195. https://doi.org/10.1515/text.2004.007

Milanović, M., & Milanović, A. (2010). Softening the words of crisis: Hedging in economics-related articles published on the internet. *Ekonomski Horizonti*, *12*(1), 121–129.

Milard, B., & Tanguy, L. (2018). Citations in scientific texts: Do social relations matter? *Journal of the Association for Information Science and Technology*, *69*(11), 1380–1395. https://doi.org/10.1002/asi.24061

Millar, N., Salager-Meyer, F., & Budgell, B. (2019). "It is important to reinforce the importance of…": 'Hype' in reports of randomized controlled trials. *English for Specific Purposes*, *54*, 139–151. https://doi.org/10.1016/j.esp.2019.02.004

Morin, H. (2020, Septembre 16). Covid-19: La frénésie des prépublications scientifiques retombe. *Le Monde*.

Myers, G. (1989). The pragmatics of politeness in scientific articles. *Applied Linguistics*, *10*(1), 1–35. https://doi.org/10.1093/applin/10.1.1

Myers, G. (1996). Strategic vagueness in academic writing. In E. Ventola & A. Mauranen (Eds.), *Academic writing: Intercultural and textual issues* (p. 3) Pragmatics & Beyond New Series 41. Amsterdam, Philadelphia: John Benjamins Publishing Company. https://doi.org/10.1075/pbns.41.04mye

Nerlich, B. (2013). Moderation impossible? On hype, honesty and trust in the context of modern academic life. *The Sociological Review*, *61*(2_suppl), 43–57. https://doi.org/10.1111/1467-954X.12099

Orasan, C. (2001). Patterns in scientific abstracts. *Proceedings of Corpus Linguistics*, 433–445.

Pérez-Llantada, C. (2008). Stance and academic promotionalism: A cross-disciplinary comparison in the soft sciences. *Atlantis*, *30*(1), 129–145.

Pool, R. (2020, Octobre 29). Closed countries, open data. *Research Information*.

Rinaldi, A. (2012). To hype, or not to(o) hype. Communication of science is often tarnished by sensationalization, for which both scientists and the media are responsible. *EMBO Reports*, *13*(4), 303–307. https://doi.org/10.1038/embor.2012.39

Rubin, A., & Parrish, D. (2007). Problematic phrases in the conclusions of published outcome studies: Implications for evidence-based practice. *Research on Social Work Practice*, *17*(3), 334–347. https://doi.org/10.1177/1049731506293726

Salager-Meyer, F. (2017). I think that perhaps you should: A study of hedges in written scientific discourse. In T. Miller (Ed.), *Functional approaches to written text: Classroom applications* (pp. 105–118). United States Information Agency.

Samar, R. G., Talebzadeh, H., Kiany, G. R., & Akbari, R. (2014). Moves and steps to sell a paper: A cross-cultural genre analysis of applied linguistics conference abstracts. *Text & Talk*, *34*(6), 759–785. https://doi.org/10.1515/text-2014-0023

Scott, S. L., & Jones, C. W. (2017). Superlative scientific writing. *ACS Catalysis*, *7*(3), 2218–2219. https://doi.org/10.1021/acscatal.7b00566

Shehzad, W. (2010). Announcement of the principal findings and value addition in computer science research papers. *Iberica*, *19*(2004), 97–118.

Shinohara, K., Suganuma, A. M., Imai, H., Takeshima, N., Hayasaka, Y., & Furukawa, T. A. (2017). Overstatements in abstract conclusions claiming effectiveness of interventions in psychiatry: A meta-epidemiological investigation. *PLoS One*, *12*(9), e0184786. https://doi.org/10.1371/journal.pone.0184786

Sun, J., & Hyland, K. (2001). Disciplinary discourses: Social interactions in academic writing. *TESOL Quarterly*, *35*(2), 344. https://doi.org/10.2307/3587657

Swales, J. (2004). The research article revisited. In *Research genres* (pp. 207–240). Cambridge University Press. https://doi.org/10.1017/CBO9781139524827.008

© 2021 The Authors.
Learned Publishing © 2021 ALPSP

Swales, J., & Feak, C. (2009). Abstracts and the writing of abstracts. In *Abstracts and the writing of abstracts* (Vol. *I*). University of Michigan Press. https://doi.org/10.3998/mpub.309332

Thapa, D. K., Visentin, D. C., Hunt, G. E., Watson, R., & Cleary, M. (2020). Being honest with causal language in writing for publication. *Journal of Advanced Nursing*, *76*(6), 1285–1288. https://doi.org/10.1111/jan.14311

Vassileva, I. (2001). Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes*, *20*(1), 83–102. https://doi.org/10.1016/S0889-4906(99)00029-0

Vinkers, C. H., Tijdink, J. K., & Otte, W. M. (2015). Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: Retrospective analysis. *BMJ*, *351*(dec14_13), h6467. https://doi.org/10.1136/bmj.h6467

Wang, W., & Yang, C. (2015). Claiming centrality as promotion in applied linguistics research article introductions. *Journal of English for Academic Purposes*, *20*, 162–175. https://doi.org/10.1016/j.jeap.2015.05.002

Warr, W. A. (2003). Evaluation of an experimental chemistry preprint server. *Journal of Chemical Information and Computer Sciences*, *43*(2), 362–373. https://doi.org/10.1021/ci025627a

Wheatley, D. (2014). Drama in research papers. *European Science Editing*, *40*(1), 14.

World Health Organization. (2020). *Infodemic management of WHO Information Net Work for Epidemics*. www.who.int/teams/risk-communication/infodemic-management

World Medical Association. (2001). World Medical Association Declaration of Helsinki. Ethical principles for medical research involving human subjects. *Bulletin of the World Health Organization*, *79*(4), 373–374. https://doi.org/10.1016/0300-9629(89)90145-x

Xing, W., Hejblum, G., Leung, G. M., & Valleron, A.-J. (2010). Anatomy of the epidemiological literature on the 2003 SARS outbreaks in Hong Kong and Toronto: A time-stratified review. *PLoS Medicine*, *7*(5), e1000272. https://doi.org/10.1371/journal.pmed.1000272

Zhang, L., Zhao, W., Sun, B., Huang, Y., & Glänzel, W. (2020). How scientific research reacts to international public health emergencies: A global analysis of response patterns. *Scientometrics*, *124*(1), 747–773. https://doi.org/10.1007/s11192-020-03531-4

© 2021 The Authors.
Learned Publishing © 2021 ALPSP