The Covid infodemic

Competition and the hyping of virus research

Ken Hyland and Kevin Jiang
University of East Anglia | Jilin University

Covid-19, the greatest global health crisis for a century, brought a new immediacy and urgency to international bio-medical research. The pandemic generated intense competition to produce a vaccine and contain the virus, creating what the World Health Organization referred to as an 'infodemic' of published output. In this frantic atmosphere, researchers were keen to get their research noticed. In this paper, we explore whether this enthusiasm influenced the rhetorical presentation of research and encouraged scientists to "sell" their studies. Examining a corpus of the most highly cited SCI articles on the virus published in the first seven months of 2020, we explore authors' use of hyperbolic and promotional language to boost aspects of their research. Our results show a significant increase in hype to stress certainty, contribution, novelty and potential, especially regarding research methods, outcomes and primacy. Our study sheds light on scientific persuasion at a time of intense social anxiety.

Keywords: Covid research, hype, academic persuasion, scientific writing

1. Introduction: the infodemic

At the end of July 2020, with the number of coronavirus cases worldwide approaching 18 million and over 650,000 deaths, maximum effort was being expended to contain the pandemic. At that time, the World Health Organization (WHO) estimated that doctors and scientists had published over 50,000 studies of Covid-19, ranging from disease transmission to virus-induced anxiety (Allen-Mills & Gregory, 2020). Covid-19 clinical trials grew from 60 in the first week of January to 4,271 (Nature index, 2020). With the urgency of the search for effective medical treatments continuing, tens of billions of dollars had been invested and universities, pharmaceutical companies, research institutes and government laboratories were working on 133 possible vaccines. In this feverish atmosphere of intense, high-stakes competition, the pace of scientific publishing accelerated dramatically. Many journals were receiving up to twice their normal submissions and greatly speeded up their publication times, with *Science*, for example, publishing one paper only nine days after submission (Tingley,

2020). At the same time, scientists were uploading thousands of papers to open-access preprint servers without normal peer-review, and publishers of elite journals such as *Science*, *The Lancet*, *JAMA* and *The New England Journal of Medicine*, made potentially life-or-death coronavirus research available free online (Wellcome Trust, 2020).

This unprecedented growth in the volume and speed of scientific publishing led WHO to speak of:

> a massive 'infodemic' — an overabundance of information, some accurate and some not — that makes it hard for people to find trustworthy sources and reliable guidance when they need it.[1]

Covid-19 rapidly became a highly politicised pandemic, with some scientists arguing that fastidious research standards and core methodological principles should be relaxed (see London & Kimmelman, 2020 for refutation). As a result, the journal *Retraction Watch* listed 33 Covid paper retractions by August 1st, including studies in two of the world's most prestigious medical journals, *The New England Journal of Medicine* and *The Lancet* [2].

In this paper, we explore whether this enthusiasm to understand the virus and discover a vaccine for its control extended beyond the hurried design of experiments, collection of data and submission of findings. We focus here on the rhetorical presentation of the research to discover whether scientists employed language to "sell" aspects of their studies. We do this by examining authors' use of hyperbolic and promotional language to glamorise, publicize, embroider and/or exaggerate aspects of their research – a phenomenon Millar et al. (2019) refer to as 'hype'. Specifically, we address the following questions:

i. To what extent do authors of Covid papers hype their research?
ii. What are the most common hyping devices?
iii. What aspects of research does this language target?
iv. Has the extent of this changed during the course of the pandemic?

2. The hyping of medical research

[1] https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200202-sitrep-13-ncov-v3.pdf
[2] https://retractionwatch.com/retracted-coronavirus-covid-19-papers/

That academics promote the importance or value of their work is not news. We have long known that scientists routinely conceal contingent factors, downplaying the role of social allegiance, self-interest and power to depict research as a disinterested, inductive, democratic and goal-directed activity (e.g. Gilbert & Mulkay, 1984). Moreover, over 20 years ago Fairclough (1993) referred to the 'marketisation' of research and Hyland (1999) spoke of authors 'boosting' statements to strengthen their claims. This can be seen in Examples (1)–(3) from our Covid corpus[3]:

(1) Our investigation provides ***significant*** practical implications for public health decision- and policymakers.

(2) Delineating the proportion of infections is ***critical*** to refining model parameterization.

(3) We have ***established*** an ***essential*** component of the transmission…

The promotion of research findings, however, seems to be growing. A number of studies have noted how there has been an increase in 'a news-oriented text schema' designed to promote the authors' results in research papers in Physics (Swales & Najjar, 1987), biological sciences (Berkenkotter & Huckin, 1995), and computer science (Shehzad, 2010). This increasing promotion of results has developed to accommodate the increasingly selective reading patterns of researchers swamped by the explosion of information in the sciences but is also a feature of articles in the humanities. Lindeberg (2004), for example, found that authors in language and literary studies routinely give positive evaluations of their own work and those that support it while negatively evaluating dissenting views. Similarly, promotional elements that underline the worth and significance of research have also been observed in applied linguistics articles (Wang & Yang, 2015).

In addition to foregrounding newsworthy information, authors also promote themselves and their work in other ways. The increasingly competitive marketplace of academic research means that scholars are now giving more prominence to their own previous research through self-citation (Hyland, 2003; Hyland & Jiang, 2019). This potentially distorts citation counts and reduces their reliability as a measure of quality, exaggerated by the fact

[3] All examples are from our Covid-19 corpus unless otherwise indicated.

that the more one cites oneself, the more one is cited by others (Fowler & Aksnes, 2007). Hyland & Jiang (2019) have also found that scientists are now substantially more "present" in their texts with significant rises in self-mention in the last 50 years.

While such rhetorical tactics have become important strategies for promoting research, we are more specifically concerned with the role of hyperbolic language: the use of words which impose subjective value on claims to embellish results and promote the merits of studies. The effect of these devices can be seen if we remove the boldface items from Examples (4) and (5):

> (4) It is **noteworthy** that the discovered evolutionary relationships of CoVs shown by the whole genome, RdRp gene, and S gene were **highly** consistent with those exhibited by complete genome information.
> (5) Medical imaging and neuropathology will **certainly** play an **important** role to detect abnormalities in … COVID-19 patients.

Some scientists feel that such hyping practices have reached a level where objectivity is sacrificed for manufactured excitement. The former Editor of *Cell Biology International,* for example, has complained that the widespread use of 'drama words' such as *drastic decrease, new and exciting evidence, remarkable effect* has turned science into a "theatrical business" (Wheatley, 2014).

There is also evidence to suggest that this is increasing. Thus, Fraser & Martin (2009), for example, found a significant increase in adjectives stressing value (e.g. *important, critical, original*) in clinical research journals between 1985 and 2005. In biology, Hyland & Jiang (2019) report an increase in the boosters *show*, *must* and *know* over the last 50 years, forms which ensure readers are aware of the strength of results. They also note a rise in *essential* with an extraposed *to-clause* or followed by *for*, forms which express judgments of extreme importance or necessity. Research showing the extent of the rise of hyperbole in medical journals is most dramatically shown by Vinkers et al. (2015) who found that the frequency of 25 positive-sounding words such as *novel*, *amazing*, *innovative* and *unprecedented* increased almost nine-fold in the titles and abstracts of papers published in PubMed between 1974 and 2014.

Hype and the use of language to sell research, has also been studied in Randomised Control Trials (RTCs) by Millar et al. (2019), from whom we borrow the term 'hype'. RTCs

are the gold standard used in medicine to measure the effectiveness of a new intervention or treatment, reducing bias and examining the effect of new drugs. Yet, in a small corpus of 24 texts, Millar et al. (2019) found 6.7 occurrences of hype words per-paper, or 2 occurrences per 1,000 words, principally in method and discussion sections and mainly to emphasise the expertise of the authors or the strength of the study design. Millar et al. (2019: 149) argue that this focus on marketability rather than the actual significance of the reported research can impose "judgments on readers that might undermine objective and disinterested evaluation of new knowledge" .

This growth of promotional language (Vinkers et al., 2015; Hyland & Jiang, 2019) is not unrelated to the massive expansion of research and the cutthroat competition engendered by the appraisal culture which has accompanied it. With as many as eight million scholars seeking to publish in English-language journals each year (UNESCO, 2017) and over 3 million peer reviewed papers published (Johnson et al., 2018), there is increasing pressure on scholars. This has been exacerbated not only by the ease of collaboration and access afforded by new technologies, but also by the fact that promotion and career opportunities of scholars across the globe are increasingly tied to an ability to gain acceptance for work in high profile journals. It is not surprising, therefore, that many scientists seem to have come to "assume that results and their implications have to be exaggerated and overstated in order to get published" (Vinkers et al., 2015: 3). This pressure has been massively ramped up, moreover, with the urgency, and opportunities, created by the Covid-19 pandemic. The historian Lorraine Daston, for example, argued in April that the virus would drive scientific research with a ferocity not seen since the 1600s, contending that: "we are living in a moment of ground zero empiricism in which almost everything is up for grabs" (Daston, 2020: 4).

The desperation to make a breakthrough in Covid-19 research has created an "infodemic" of research such that scientists are unable to keep pace with the torrent (e.g. Brainard, 2020), and some of this, moreover, has been hastily conducted and poorly scrutinised (e.g. Redden, 2020). The acceleration of journals' publication processes is laudable in making information available quickly, and in a study of 669 articles, for example, Horbach (2020) found that medical journals have halved their publishing times since the crisis began. Adequate editorial assessment and peer review obviously require some time, however, and quick information dissemination also raises concerns of quality (e.g. Dinis-Oliveira, 2020). In the first four months of the pandemic, for example, there were some 6,000 unreviewed papers posted on *bioRxiv* and *medRxiv*, two preprint servers for biomedical

research (Fraser et al., 2020). Some scientists have raised concerns that such unvetted, early science can spread disinformation and create 'clickbait science' which skews public debate (Heimstadt, 2020). But as the increasing numbers of papers reported on *Retraction Watch* show, there are problematic papers even in the most prestigious journals.

In this fevered context of competition and frantic deadlines, we suspect there is also a clamour for rhetorical precedence: a heightened tendency for medical researchers to emphasize the significance, uniqueness and novelty of their research. We now turn to describe the methods and approach we use to explore this hypothesis before discussing our results.

3. Corpus and analysis

We built a corpus of research articles on Covid-2019 by extracting the 200 most highly cited papers in SCI-indexed journals on the *Web of Science* focusing explicitly on the virus and published between January and July 2020 (see Appendix for list of journals). We identified these papers following Teixeira da Silva et al., (2020) and *Nature Index* (2020), who suggested these terms naming the virus:

> "SARS-CoV-2" OR "COVID-19" OR "Coronavirus 2019" OR "Corona Virus 2019" OR "novel coronavirus" OR "novel corona virus" OR "2019-nCoV"

Searching for papers containing these terms in their titles produced 5,707 articles with a total of 51,688 citations. We took the most highly cited 200 papers from these for our corpus, so while they represent several fields of research, and so contain different experimental designs, they focus explicitly on the virus and exemplify the most influential studies at the time. For purposes of comparison we used a reference corpus of 200 scientific research articles from Science Citation Index (SCI) journals in the same disciplines. Table 1 provides the details of our corpora.

Table 1. Characteristics of corpora

Corpus	disciplines	No. of journals	No. of papers	No. of words
Covid-19	medicine & biology	90	200	1,347,150
Reference corpus	medicine & biology	40	200	1,698,080

To explore these corpora we developed a list of hyperbolic items from a potentially open set. First we included the categories of boosters and positively marked attitude markers from Hyland's (2005) stance framework. 'Boosters' are epistemic devices which express conviction, seeking to assert claims categorically and shut down alternative voices (*demonstrate*, *show*, *clearly*). 'Positive attitude markers', on the other hand, indicate the writer's affective perspectives and include evaluations and personal feelings towards content (*interestingly*, *fascinating*) or on the communication itself (*honestly*, *in truth*). Together these devices convey a writer's personal assessments and comment on either the truth or the value of arguments to express a conviction in claims. This list was supplemented by reference to the literature on this issue, such as those referred to as "positive words" (Vinkers et al., 2015), "superlatives" (McCarthy, 2015) and "hyperbolic terms" (Millar et al., 2019). We also scanned other sources for a wider inclusion of candidate items, such as the *Oxford Thesaurus of English* (Waite, 2009) and corpus-generated wordlists such as the Academic Word List (Coxhead, 2000) and the Academic Vocabulary List (Gardner & Davis, 2013). This procedure produced a list of about 400 hype items.

We searched our corpora for these items using *AntConc* (Anthony, 2019), and then manually examined and counted each concordance to establish that the feature was performing a hyping function. For example, the word *major* was excluded in contexts such as "Substrate A targets the major binding site" but seen as hyping when used to modify a claim such as "Complex A plays a major role in …". Similarly words such as *important* and *definitive* were ignored in the negative (not important/ no definitive conclusion) (see Fraser & Martin, 2009). Both authors worked independently and achieved a high inter-rater agreement ($\kappa > .08$) before resolving disagreements. In order to explore what aspects of the paper the items targeted we adopted a modified version of the categories proposed by Millar et al. (2019), namely: "Broad Research Area", "Specific Research Topic", "Authors' Prior Research", "Research Methods", "Research Outcome" and "Research Primacy". In order to answer the fourth RQ, concerning changes in hyping practices during the course of the pandemic, we also classified the Covid-2019 corpus by month, the details of which are shown in Table 2.

Table 2. Composition of Covid-2019 corpus by month

	January	**February**	**March**	**April**	**May**	**June**	**July**
papers	4	22	25	42	51	28	28
words	26,940	148,170	168,375	282,870	343,485	188,580	188,730

All the results were then normalised to 10,000 words to allow comparison across the monthly corpora. To determine statistical significances, the log-likelihood (LL) test was run using Rayson's log-likelihood spreadsheet (Rayson, 2016), and effect size for log-likelihood test (ELL) was also considered according to Johnston et al. (2006). The test was applied to compare frequencies across the Covid-19 corpus and the reference corpus (as indicated in Section 4), and to compare frequencies in the first and last monthly sub-corpora of the Covid-19 corpus to identify any change across the pandemic (see Section 4.4).

4. Findings

This section is divided into four subsections which correspond to the research questions set out in the Introduction. As a reminder, these are: How much hyping is there? What are the most common hyping devices? What aspects of research are most hyped? And, has hyping behaviour changed during the pandemic?

4.1 How much hyping is there?

We identified 10,929 instances of hyping overall, averaging 27.3 cases in each paper and 35.9 per 10,000 words. The extent of this phenomenon is perhaps surprising and points to a widespread sensitivity on the part of writers to promote their work and ensure their voice is heard in an increasingly competitive academic context. While writer-reader interactions have been widely identified in the literature as a means to establish a writer's credibility, enhance reader engagement, and strengthen research claims (e.g Hyland, 2005), the presence of hype now seems to be widely established.

The Covid-19 corpus contained significantly more devices than the reference corpus of texts from the same fields (LL=80.5, p<0.0001, ELL<0.001). We also found a significant difference between the use of boosters and affect in the two corpora. Boosters, which confirm the writer's commitment to the truth and reliability of a proposition, comprise a greater proportion of items in the reference corpus compared with affect devices, which express personal and professional attitudes towards what is said. Table 3 shows these differences.

Table 3. Comparison of hyping devices in the two corpora

Corpus	Cases	Per 10,000 words	Boosters %	Affect %
Covid-19	5302	39.4	24.7	75.3
Reference	5627	33.1	37.5	62.5
Totals	10,929	35.9	31.3	68.7

There is, then, a higher ratio of words marking positive affect in the Covid-19 papers, as authors inject emotional colour into their texts, than items which promote the epistemic strength of statements. This is almost certainly related to the novel characteristics of the virus in the first months of the pandemic making certainties more difficult to determine. So, a well-established research subject such as a specific glycolate oxidase in (6) and vaccination in (7) allow writers to present results as allowing little disagreement. The Covid scientists, on the other hand, dealing with a new and little understood subject, may choose to be more circumspect in how they present their findings, instead preferring to stress the prospective benefits of their work, as in (8) and (9).

(6) Both these experiments **clearly showed** that GOX3 preferentially metabolizes l-lactate in vivo. (Reference corpus)

(7) The total number of cases averted by vaccination over the child's lifetime is ***always*** predicted to be positive, in agreement with data from the Phase III trial of RTS,S. (Reference Corpus)

(8) Such clinical research on the treatment of SARS, MERS, and the new coronavirus 2019-nCoV is ***necessary*** if we are to understand the potential risks and benefits.

(9) This work can ***greatly contribute*** to an improved understanding of how 2019-nCoV invades human body systems.

4.2 What are the most common hyping devices?

Table 4 presents the most frequently occurring items used to promote the value of research in the corpora. The word *significant* appears at the head of both lists (after removing all cases of

statistical uses) to mean "sufficiently important to be worthy of attention". The fact that it comprises only half of the cases of hype in the monitor texts, however, suggests that there are far more different types (items) in the Covid texts. Such types include many that are not in the comparison texts at all, including *fascinating, extremely* and *definitive*. It is interesting to note that among the top 10, *first*, *contribute*, *potential*, *promising* and *necessary* are significantly more prominent in the Covid papers (*LL*=302.98, *LL*=98.56, *LL*=140.56, *LL*=45.36, *LL*=53.39, $p<0.001$, *ELL*<0.001).

Table 4. The most common hyping items (per 10,000 words & percentage of total cases)

Covid-19 Corpus			**Reference Corpus**		
item	**normed freq**	**%**	**item**	**normed freq**	**%**
significant	4.2	9.5	significant	6.9	18.8
important	3.1	6.9	important	3.5	9.2
effective	1.8	4.0	strong	2.5	6.6
first	1.6	3.4	clear	1.4	3.6
contribute	1.2	2.5	interesting	1.4	3.6
potential	1.0	2.0	novel	1.4	3.5
very	1.0	2.0	effective	1.1	2.8
promising	0.9	1.7	show	0.9	2.1
highly	0.8	1.7	unique	0.7	1.7
strong	0.8	1.5	very	0.7	1.7
necessary	0.7	1.4	best	0.7	1.5
essential	0.7	1.4	essential	0.6	1.3
notable	0.7	1.3	crucial	0.6	1.2
critical	0.6	1.1	useful	0.6	1.2
facilitate	0.5	1.1	notable	0.5	1.1
substantial	0.5	1.1	demonstrate	0.5	1.0
timely	0.5	1.0	apparent	0.5	1.0
clear	0.3	0.8	robust	0.5	0.9
unique	0.3	0.8	substantial	0.5	0.9
successful	0.3	0.8	enhance	0.4	0.7

Clearly, expressing a strong evaluation in a text involves both a statement of personal judgement and an appeal to shared values. Hype is therefore interpersonal; it requires writers to draw on their knowledge of what is prized by the community and how to appeal to this. From the top 20 items in Table 4 we identified four broad categories of value which writers seemed to be promoting in their papers:

i. Certainty (concerns the strength or importance of the statement – *significant*, *important*, *strong*, *crucial*, *clear*)

ii. Contribution (refers to the immediate value or use of the issue – *necessary*, *essential*, *effective*, *useful*)
iii. Novelty (stresses the originality of the claim – *first*, *timely*, *novel*, *unique*)
iv. Potential (comments on the possible future value of something – *promising*, *potential*, *apparent*)

Like the texts in the reference corpus, the top two items in the Covid papers related to a clear assurance of importance or benefit of the current research, and such declarations of certainty comprise half of the most frequent items in the Covid corpus, as can be seen in Examples (10)–(12):

(10) Appropriate use of glucocorticoids is able to **significantly** improve the clinical symptoms of patients, reduce the degree of disease progression, and accelerate the absorption of lung lesions.

(11) […] early identification of the specific features of severe paediatric patients and timely treatment are of **crucial importance**.

(12) This marks a **substantial** improvement on previous results conducted elsewhere.

The function which occurs next most frequently in the top 20 is the contribution the study is claimed to make to understanding or overcoming the pandemic. Here we find items which assert the direct and current benefits of the research, as in Examples (13) and (14):

(13) The temporal and spatial distribution of infectious SARS-CoV-2 RNA strands is a **necessary** research area.

(14) Delineating the proportion of infections is **critical** to refining model parameterization.

The remaining two areas which writers address to hype their research concern novelty and potential. Clearly, the originality of research is at a premium in a context where scientists are searching for ideas which might impact the spread of the virus, as in (15) and (16), although

often it is the potentiality of this impact, rather than its confirmed value, which is all that can be said for it, as in (17) and (18).

> (15) this study provides the **first** report on a potential closely related kin (Pangolin-CoV) of SARS-CoV-2 […]
>
> (16) we therefore describe for the **first** time the 18F-FDG PET/CT findings of four patients with COVID-19.
>
> (17) The results are **promising** and suggest possible inhibition for the currently available therapeutics against the newly emerged coronavirus.
>
> (18) We hope that these results will **contribute** to the prevention, diagnosis, and treatment of VTE.

The scientists working on Covid-19 research, driven by the urgency of the search for viable treatments and understandings of the disease, thus seem eager to promote the value of their work. Equally, however, they recognise the limitations of what they can demonstrably prove so statements of certainty represent a smaller proportion of hyping items than the reference corpus. Nevertheless, writers seek to emphasise the contribution of their research, its novelty and its potential to make a significant difference in the fight to overcome the disease.

4.3 What aspects of research are most hyped?

In addition to the specific rhetorical functions which writers boost or positively evaluate with hyping terms, we also examined the broad functional categories targeted by the hyping words and expressions. Following a modified version of Millar et al.'s (2019) classification, we employed the following categories to determine which aspects of the study hypes served to embellish.

i. "Broad Research Area": hypes targeted at the general field of study
ii. "Specific Research Topic": relates to the particular area under investigation

iii. "Authors' Prior Research": hypes where the authors identify themselves as the source of related research
iv. "Research Method": targeted at how the study was designed or conducted
v. "Research Outcome": concerned with the results of the study or its interpretations
vi. "Research Primacy": hypes that describe the research as superior or assign it priority in some way.

To some extent, but not entirely, these rhetorical functions correspond with the IMRD (Introduction, Methods, Research, Discussion) structure of the conventional science research paper. Millar et al. (2019), for example, found that hypes related to both broad and specific research topics mainly occurred in the introduction to establish the centrality of the topic and the purpose of the research; the authors' prior research was mainly found in the discussion; hypes related to research methods were in the methods and discussion sections; and those boosting research outcomes and primacy mainly in the discussion.

Table 5 shows that four functions were hyped significantly more by writers working on coronavirus research than their peers more generally. These are "Broad Research Area", "Research Methods", "Research Outcome" and "Research Primacy". In contrast, "Specific Research Topic" and "Authors' Prior Research" are significantly less hyped by the coronavirus writers.

Table 5. Hyped targets in the Covid-2019 and Reference corpora (per 10,000 words)

Target Focus	**Covid-19**	**Reference Corpus**	***LL***	***p***
Broad Research Area	4.9	3.2	53.1	0.0001
Specific Research Topic	3.5	4.1	7.0	0.001
Authors' Prior Research	3.8	5.7	54.5	0.001
Research Methods	9.6	6.4	96.5	0.0001
Research Outcome	10.8	8.2	53.6	0.0001
Research Primacy	9.8	8.8	8.7	0.001
Total	**42.4**	**36.4**	**68.91**	**0.0001**

* *ELL*<0.001

The *broad research area* concerns the general topic being addressed in the paper and is often used to claim centrality for the research in the Introduction, although hyping this area of the paper was also found in the discussion section of some papers. Here, in Examples (19)–(21), the Covid scientists are seeking to promote the field they deem important:

(19) Tracheal intubation is a **potentially** high-risk procedure for the airway manager, particularly as it risks exposure to a high viral load. This is **clearly** an area of **great importance**.

(20) The ongoing 2019-nCoV outbreak has **undoubtedly** brought back memories of the severe acute respiratory syndrome (SARS)–coronavirus (CoV) outbreak.

(21) Vigilant epidemiological control in the community and health-care facilities is **very important** to prevent another SARS-like epidemic.

In these examples, for instance, the writers are encouraging readers to accept that the domain they have identified offers a significant or worthwhile area to traverse. The reason this hard sell involves greater rhetorical work than in papers from the reference corpus is perhaps due to the very novelty and fast-moving nature of the area. Unlike more established fields, the urgency and uncertainty of the pandemic requires different paths to be followed, and authors promote their favourites.

The greatest differences in hyping between the two corpora are found in the promotion of methodological decisions and practices. Although Methods are generally considered to be the most expository, factual and least overtly persuasive part of research articles (e.g. Samraj, 2016), scientists consistently use hypes in this section. This is particularly true of the Covid writers who, in the competitive context of the search for answers to the pandemic, are seeking to establish that their approach is both valid and robust. The peculiar nature of the current situation thus perhaps encourages the use of unusual and innovative research designs and methods, but it also places pressure on writers to demonstrate the value of their study in terms of the rigor of their methodology; see Examples (22)–(24).

(22) To **facilitate** rapid development of recommendations, we performed a **novel systematic** prioritization of outcomes by the ongoing SSC guideline 2020 work and expert input.

(23) In this study, we did a **comprehensive** evaluation of deceased patients and patients recovered among those with confirmed covid-19 who were previously transferred or admitted to the isolation ward of Wuhan Tongji Hospital.

(24) With a **robust** and **complete** line list for characterising the epidemiology of this novel pathogen, we **effectively** inferred the outbreak size of 2019-nCoV in Wuhan from the number of confirmed cases that have been exported to cities outside mainland China.

The greatest number of hype items was found in claims made for research outcomes. This is not surprising as it is here that authors seek to underline the importance of their findings and the weight of their interpretations. Hyping of outcomes was mainly achieved by the use of adverbs and adjectives to impart a personal take on the results, see Examples (25)–(28), highlighting how the reader should understand what was being presented to them and ensuring they got the assessment of significance:

(25) **Strikingly**, the 2019-nCoV S-protein sequence contains 12 additional nucleotides upstream of the single Arg cleavage site 1 (Fig. 1, Fig. 2) **revealing** a predictively solvent-exposed PRRAR ↓ SV sequence

(26) These data **clearly demonstrate** that hrsACE2 can **significantly** block early stages of SARS-CoV-2 infections.

(27) **Importantly**, VSV-G-pseudotyped particles were able to produce luciferase signal in all cell lines tested in this study.

(28) It is **notable** that two related lineages of coronaviruses are found in pangolins that were independently sampled in different Chinese provinces and that both are also related to SARS-CoV-2. This **shows** that these animals may be **important** hosts for these viruses, which is **surprising** as pangolins are solitary animals that have relatively small population sizes

The terms *striking*, *notable*, *important*, and *surprising* and are examples of what Wheatley (2014) calls drama words which conjure up a sense of theatre. They certainly add a hyperbolic dimension to the text in order to send a clear signal of the significance of the result and the take-away message for the reader.

The final category which recorded more hypes in the Covid papers is that of the primacy given to the research. These promote the research itself rather than the results and often refer to the likely future value of pursuing this line of work or assigning it a superior status in terms of its novelty. Examples (29)–(31)are typical:

(29) The results are ***promising*** and suggest inhibition for the currently available therapeutics against the newly emerged coronavirus.

(30) This study provides a ***detailed*** and ***robust*** analysis of essential residues and ligand-receptor interactions for the development of peptide-like structures as SARS-CoV-2 main protease inhibitors.

(31) Our study is the ***first*** nationwide investigation that systematically evaluates the impact of comorbidities on the clinical characteristics and prognosis in patients with COVID-19 in China.

Table 5 indicates that promoting the overall value of the current research is also a common feature of work in biology and medicine in more normal times. In the cut-throat competitive world in which academics now work, publication and the citations which those publications attract, determine reputations and careers. As we have said, this intense and fevered context is aggravated by the uncertainties of the pandemic and the desire to make an impact on its eventual defeat.

Hyping features in the remaining rhetorical categories of specific topic and authors' prior research were, however, statistically more frequent in the reference corpus taken from 2015. Clearly the recency of the crisis, the fact that the virus emerged only in January, restricts the opportunities for scientists to cite and promote their own work on the virus. Indeed, citations appear to be lower in the Covid texts than in those in the reference corpus. Scientists are therefore less well-positioned to hype their prior work in the same way as these writers in the reference corpus; see Examples (32) and (33):

(32) **We recently showed** that Fstl1 acts as a BMP4 antagonist to play a key role in lung development (Geng et al., 2011). (Reference corpus)

(33) **Our earlier work importantly expanded** upon previous research by employing DI models to test how interactions between species pairs within communities influence diversity. (Reference corpus)

This is not to say that Covid researchers are unable to promote their previous work or establish links to it. The frenetic pace at which research was being conducted at the time, and the speed at which it was being published, allowed some academics to build on their work very quickly. Self-citation has been shown to be particularly heavy among authors who have a long history of engagement in an area (Pichappan & Sarasvady, 2002) and increases as scholars move through their careers and publish more research (Chang, 2006). Referring to one's own work seems to be relatively low in medical and life sciences compared with other fields (Public Policy Group, 2011). However, the intensive and serial pursuit of work on the virus is likely to increase the extent to which writers are able to specialize and build on their research, so we may see more hyping of their prior work as in Examples (34) and (35):

(34) Our recent study **highlighted** the **important** role of ACE2 in mediating entry of SARS-CoV-21.

(35) However, our previous clinical trial of influenza treatment **showed** that a triple antiviral combination could **significantly** improve the clinical outcome and viral load profile and **greatly** reduce emergence of resistant virus quasispecies.

We also found significantly more hyping behaviour in the reference corpus related to the category of "Specific Research Topic". Promoting the topic of the paper allows writers to establish the significance of their research and specify the gap they are filling in the knowledge of the broader research area. In doing so they are not only furthering readers' probable acceptance of the value of the research but also establishing their expertise in the area and right to be heard. Interestingly, Hyland & Jiang (2019) have shown that in the last 50 years biologists have adopted more involved, stance-laden discourses which promote the importance of their research and emphasize the role of the interpreting researcher. This authorial repositioning from traditional faceless discourses, not only makes the author's role clearer and strengthens claims, but highlights the significance of the research area; see Examples (36) and (37):

(36) VC is one of the main components of the cardiovascular remodelling process that ***substantially contribute to*** cardiovascular events and increased morbidity and mortality of patients with CKD. (Reference corpus)

(37) There are, however, many ***important*** reasons to study these animals, which play a ***critical*** role in the ecology of the seas, with a ***major*** impact on the atmosphere. (Reference corpus)

It is perhaps the very urgency and importance of the Covid virus which means that specific research areas are less hyped than topics in the Reference corpus. The interdisciplinary audiences who read papers in the life and biological science journals may be less aware of a specific topic than specialists and need to be shown their importance. On the other hand, we might speculate that there is less hyping of "Specific Research Topics" in the Covid papers simply because there is less need to do so. Covid research is far more specialised and, for those working in areas relevant to it, there may be greater awareness of the tracks on which research is running so less promotion is required. Having said that, however, the frequency of hypes in the Covid texts is not far short of those in the Reference corpus, and there is no shortage of examples where writers are prepared to ensure readers are left in no doubt of the value of what the topic offers; see Examples (38)–(40):

(38) Therefore, binding to the ACE2 receptor is a **critical** initial step for SARS-CoV to enter into target cells.

(39) Reverse transcription-polymerase chain reaction (RT-PCR) remains the most **useful** laboratory diagnostic test for COVID-19 worldwide.

(40) Therefore, neutralising key inflammatory factors in cytokine release syndrome (CRS) will be of **great value** in reducing mortality in severe cases.

4.4 Has hyping behaviour changed during the pandemic?

Finally, in order to better understand authors' hyping behaviour in relation to Covid-19 research, we tracked changes in our list of features over the first seven months of the pandemic.

We were not surprised to find that scientists have become more assertive and definite in presenting their results over this period, doubling their use of hyping devices (per 10,000 words) between January and July (*LL*=3174.0, *p*<0.001, *ELL*=0.002). Interestingly, both boosters (*LL*=320.4, *p*<0.001, *ELL*<0.001) and affective markers (*LL*=3122.6, *p*<0.001, *ELL*=0.002), showed significant increases over the period. As we can see from Table 6, there is a steady growth in hyping features over the period as authors gradually increase the urgency with which they press their case for their work.

Table 6. Changing frequency of hyping items in the Covid-19 corpus (per 10,000 words)

	January	**February**	**March**	**April**	**May**	**June**	**July**
boosting markers	6.2	7.4	10.1	11.8	13	17.6	20.3
affective markers	20.1	22.5	25.2	29.4	30.2	30.9	31.3
total	**26.3**	**29.9**	**35.3**	**41.2**	**43.2**	**48.5**	**51.6**

Affective markers, expressing positive evaluations of what is presented, have increased by 50% per 10,000 words, adding a considerable personal complexion to the papers as time goes on. These makers of positive affect indicate the writer's responses to material, pointing out what is important, unusual or valuable and encouraging readers to engage with the topic. As we mentioned in Section 4.3, such hypes often promote the future value of what the researchers have done and the contribution the study may make to the eventual resolution of the pandemic; see (41)–(42):

> (41) We have also found that SARS-CoV RBD-specific polyclonal antibodies could cross-neutralize SARS-CoV-2 pseudovirus infection, showing the ***potential*** to develop SARS-CoV RBD-based vaccine for prevention of infection by SARS-CoV-2 and SARS-CoV.

> (42) With these ***promising*** data in mind, we tested the prophylactic and therapeutic efficacy of remdesivir treatment in a nonhuman primate model of MERS-CoV infection, the rhesus macaque.

Increasingly over the period, however, we also see more authors stressing the importance of their work. This may indicate their growing confidence in the direction they have taken and how the potential is being profitably realized, but perhaps it simply

reflects a growing desperation among scientists to persuade others of their direction towards the collective goal, as in Examples (43)–(44):

(43) We identified the presence of an **unexpected** furin cleavage site at the S1/S2 boundary of SARS-CoV-2 S, a **novel** feature setting this virus apart from SARS-CoV and SARSr-CoVs.

(44) Delineating the proportion of infections is **critical** to refining model parameterization. In turn, estimates of both the observed and unobserved infections are **essential** for informing the development and evaluation of public health strategies.

Overall, however, it is boosters which have increased the most over the seven months, becoming over three times more frequent and becoming far more prominent in the expression of Covid arguments. As mentioned, boosters have an epistemic function, encouraging readers to accept the categorical truth of statements and displaying how the writer wishes to stand behind claims. Examples (45)–(47) are from papers published in July:

(45) Medical imaging and neuropathology will **certainly** play an important rule to detect abnormalities in olfactory bulb, cranial nerves, and brain of COVID-19 patients.

(46) the daily incidence rate would **never** surpass 1 infection per 1000 people and the number of imported infections at international destinations would **always** be in the single-digit range.

(47) This observation is **clearly** different from the pattern of hospitalisation that occurred in China, where hospitalisation was also used to ensure case isolation.

The strong assertions in these papers draw attention to the fact that statements don't just communicate ideas, but also the writer's attitude to them and to readers. In these examples the writers are weighting their commitment to show the status of propositions as accredited facts.

They therefore not only carry the writer's confidence but also an attitude to the audience, assuming that readers may be prepared to accept these assertions.

The increase in certainty, or hyperbole, over the first seven months of the crisis thus perhaps maps science's growing assurance about how to tackle the pandemic. However, the retractions of overstated papers, the climbing numbers of infected cases and the increasing fatalities from the virus, tell another story. Boosters, and the positive affect markers which together mark the hyping of research, also represent a growing concern with the continuing search for information and the desire to make a name by gaining support for certain lines of work.

5. Conclusions

In this paper we have discussed the use of hyperbolic language in the most cited research papers on Covid-19 during the first seven months of the pandemic. These features glamorise, promote or exaggerate research, helping to enliven text, personalise commentary, engage readers and boost aspects of a study. But while they may add to the readability and persuasiveness of an argument by enhancing interactivity, they can also undermine arguments and compromise the veracity of the information being presented. Such practices can challenge objective and disinterested interpretation and bias readers' evaluation of new knowledge to gain advantages over competitors. Our findings show that (i) there has been a significant increase in the use of hyperbolic language during the Covid pandemic (compared with our reference corpus); (ii) that the most common forms stress certainty, contribution, novelty and potential; (iii) that the most hyped aspects of texts are research methods, outcomes and primacy; and (iv) that the practice of hyping research increased in each of the first seven months of the pandemic.

We have suggested that academics are driven by pressures to gain visibility through their publications and citations as much as by the desire to get their research accepted and that promoting their work often plays a role in this. During the most serious global pandemic for a century, the temptation to hype their work even more seems to have been too hard for researchers to resist. It should be remembered, moreover, that our results are based on the most cited papers in top SCI journals over the period and therefore represent the cream of what is published. This, then, is the "hard case" of demonstrating hype. These are the papers

which are most likely to be circumspectly written and meticulously scrutinised to gain the acceptance of the most demanding peer reviewers. In other words, this is likely to be the tip of a larger iceberg and there are almost certainly far more papers in which research is less rigorously checked and more strenuously hyped than our data show. At the same time, there are limitations to our study. We have not, for example, sought to corroborate our interpretations of authors' decisions through interviews nor to extend it beyond the first seven months of the crisis. Both are likely to be profitable future directions in which to take the research.

The significance of our study only partly lies in its value for applied linguists and other students of academic writing. We hope to have contributed to the literature on scientific persuasion and the linguistic realisation of hype which, as we have argued, is now a common feature of academic writing in a context where publication and securing grants are crucial to a successful career. In addition, however, we believe these findings have wider implications, speaking to the relationship between social anxiety and research communication. Hype is part of researchers' desire to get there first and at moments of extreme scientific uncertainty, this desire is nakedly apparent. Perhaps it comes into its own because we recognise that we know so little and do not have the luxury of time to resolve that. This is what Daston (2020: 8), mentioned earlier in the paper, refers to as ground-zero empiricism:

> Suggestive single cases, striking anomalies, partial patterns, correlations as yet too faint to withstand statistical scrutiny, what works and what doesn't: every clinical sense, not just sight, sharpens in the search for clues.

Eventually, some of those clues will guide better experiments, more persuasive statistics and clearer directions of research so that uncertainty declines and understandings clarify. But for now, scientists seem to be hyping what they have to hand.

References

Allen-Mills, T., & Gregory, A. (2020, July 26). Bit by bit, British scientists are building an anti-Covid arsenal. *Sunday Times*. https://www.thetimes.co.uk/article/coronavirus-vaccine-bit-by-bit-british-scientists-are-building-an-anti-covid-treatment-arsenal-t5ngcv2w5

Anthony, L. (2019). *AntConc* (Version 3.5.8) [Computer software]. Waseda University. https://www.laurenceanthony.net/software

Berkenkotter, C., & Huckin, T. (1995). *Genre Knowledge in Disciplinary Communication.* Lawrence Erlbaum.

Brainard, J. (2020) Scientists are drowning in COVID-19 papers. Can new tools keep them afloat? *Science* https://www.sciencemag.org/news/2020/05/scientists-are-drowning-covid-19-papers-can-new-tools-keep-them-afloat

Chang, Y-Y. (2006, June 22–23). *How do well-established scholars cite themselves over their academic careers? Six stories of six well-established scholars* [Paper presentation]. Conference in Honour of John Swales, Ann Arbor, USA.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, *34*(2), 213–238.

Daston, L. (2020). Ground zero empiricism. *Critical Inquiry*. https://critinq.wordpress.com/2020/04/10/ground-zero-empiricism/

Dinis-Oliveira, R.J. (2020). COVID-19 research: Pandemic *versus* "paperdemic", integrity, values and risks of the "speed science". *Forensic Sciences Research*, *5*(2), 174–187.

Fairclough, N. (1993). Critical discourse analysis and the marketisation of public discourse: The universities. *Discourse & Society*, *4*(2), 133–168.

Fowler, J. H., & Aksnes, D. W. (2007). Does self-citation pay? *Scientometrics*, *72*(3), 427-437.

Fraser, N., Brierley, L., Dey, G., Polka, J., Pálfy, M., & Coates, J. (2020). Preprinting a pandemic: The role of preprints in the COVID-19 pandemic. *BioRxiv* https://doi.org/10.1101/2020.05.22.111294

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, 35. https://doi.org/10.1186/1465-9921-10-35

Gardner, D., & Davies, M. (2013). A new academic vocabulary list. *Applied Linguistics*, *35*(3), 305–327.

Gilbert, G. N., & M. Mulkay. (1984). *Opening Pandora 's box: A Sociological Analysis of Scientific Discourse*. Cambridge University Press.

Heimstadt, M. (2020). Between fast science and fake news: Preprint servers are political. *LSE Impact Blog*. https://blogs.lse.ac.uk/impactofsocialsciences/2020/04/03/between-fast-science-and-fake-news-preprint-servers-are-political/

Horbach, S. (2020). Pandemic Publishing: Medical journals drastically speed up their publication process for Covid-19. *bioRxiv* 2020.04.18.045963; doi: https://doi.org/10.1101/2020.04.18.045963

Hyland, K. (1999). Disciplinary discourses: Writer stance in research articles. In Candlin, C. & Hyland, K. (Eds.), *Writing: Texts, Processes and Practices* (pp. 99–121). Longman.

Hyland, K. (2003). Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of American Society for Information Science and Technology*, 54(3), 251–259.

Hyland, K. (2005). Stance and engagement: A model of interaction in academic discourse. *Discourse Studies*, *7*(2), 173–191.

Hyland, K., & Jiang, K. (2019). *Academic Discourse and Global Publishing: Disciplinary Persuasion in Changing Times*. Routledge

Johnson, R, Watkinson, A., & Mabe, M. (2018). *The STM Report 5th ed*. International Association of Scientific, Technical and Medical Publishers, Holland.

Johnston, J. E., Berry, K. J., & Mielke, P. W. (2006). Measures of effect size for chi-squared and likelihood-ratio goodness-of-fit tests. *Perceptual and Motor Skills*, *103*(2), 412–414.

Lindeberg, A.-C. (2004). *Promotion and Politeness: Conflicting Scholarly Rhetoric in Three Disciplines*. Åbo Akademi University Press.

London, A.J., & Kimmelman, J. (2020). Against pandemic research exceptionalism. *Science*, *368*(6490), 476–477.

Martin, P., & León Pérez, I. K. (2014). Convincing peers of the value of one's research: A genre analysis of rhetorical promotion in academic texts. *English for Specific Purposes*, *34*, 1–13.

McCarthy, M. (2015). Superlatives are commonly used in news coverage of cancer drugs, study finds. *British Medical Journal*, *351*, h5803

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.

Nature Index (2020). COVID-19 research update: How many pandemic papers have been published? *Nature Index*. https://www.natureindex.com/news-blog/how-coronavirus-is-changing-research-practices-and-publishing

Pichappan, P., & Sarasvady, S. (2002). The other side of the coin: The intricacies of author self-citations. *Scientometrics*, *54*(2), 285–290.

Public Policy Group. (2011). *Maximizing the Impacts of your Research: A Handbook for Social Scientists* (Consultation draft 3). LSE Public Policy Group. http://eprints.lse.ac.uk/35758/1/Handbook_PDF_for_the_LSE_impact_blog_April_2011.pdf

Rayson, P. (2006). *Log-likelihood spreadsheet*. http://ucrel.lancs.ac.uk/llwizard.html.

Redden, E. (2020). Rush to publish risks undermining COVID-19 Research. *Inside Higher Ed.* June 8, 2020. https://www.insidehighered.com/news/2020/06/08/fast-pace-scientific-publishing-covid-comes-problems

Samraj, B. (2016). Research articles. In K. Hyland, & P. Shaw (Eds.), *The Routledge Handbook of English for Academic Purposes* (pp. 403–415). Routledge.

Shehzad, W. (2010). Announcement of the principal findings and value addition in computer science research papers. Iberica, 19, 97-119.

Swales, J. M., & Najjar, H. (1987). The writing of research article introductions. *Written Communication*, *2*(4), 175–192.

Teixeira da Silva, J.A., Tsigaris, P. & Erfanmanesh, M. Publishing volumes in major databases related to Covid-19. *Scientometrics* (2020). https://doi.org/10.1007/s11192-020-03675-3

Tingley, K. (2020). Coronavirus is forcing medical research to speed up. *New York Times Magazine.* April 21, 2020.

UNESCO (2017) *Science Report: Towards 2030.* https://en.unesco.org/unesco_science_report/

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. *British Medical Journal, 351*, h6467.

Waite, M. (2009). *Oxford Thesaurus of English* (3rd ed.). Oxford University Press.

Wang, W., & Yang, C. (2015). Claiming centrality as promotion in applied linguistics research article introductions. *Journal of English for Academic Purposes, 20*, 162–175.

Wellcome Trust (2020). Publishers make coronavirus (COVID-19) content freely available and reusable. *Welcome Trust*. https://wellcome.ac.uk/press-release/publishers-make-coronavirus-covid-19-content-freely-available-and-reusable

Wheatley, D. (2014). Drama in research papers. *European Science Editing, 40*(1), 14–16.

Appendix: Journals in the Covid-19 Corpus.

Acs Central Science
Acta Pharmaceutica Sinica B
Aging and Disease
American Journal of Gastroenterology
American Journal of Roentgenology
American Journal of Transplantation
Anaesthesia
Annals of Internal Medicine
Annals of Oncology
Annals of Translational Medicine
Antiviral Research
Biochemical And Biophysical Research Communications
Bioscience Trends
British Medical Journal
British Journal of Dermatology
British Journal of Surgery
Cardiovascular Research
Cell
Cell Discovery
Cellular & Molecular Immunology
Chembiochem
Chinese Medical Journal
Circulation
Circulation Research
Clinica Chimica Acta
Clinical Chemistry
Clinical Chemistry and Laboratory Medicine

Critical Care Medicine
Current Biology
Diabetes-Metabolism Research and Reviews
Emerging Microbes & Infections
European Archives of Oto-Rhino-Laryngology
European Journal of Nuclear Medicine and Molecular Imaging
European Radiology
European Respiratory Journal
Eurosurveillance
Frontiers of Medicine
Gastrointestinal Endoscopy
Graefes Archive for Clinical and Experimental Ophthalmology
Gut
Infection Genetics and Evolution
Intensive Care Medicine
International Journal of Antimicrobial Agents
International Journal of Environmental Research and Public Health
International Journal of Infectious Diseases
International Journal of Oral Science
Jama Network Open
Jama Neurology
Jama Ophthalmology
Jama-Journal of The American Medical Association
Journal of Advanced Research
Journal of Biological Chemistry
Journal of Biomolecular Structure & Dynamics
Journal of Clinical Medicine
Journal of Clinical Microbiology
Journal of Dental Research
Journal of General Internal Medicine
Journal of Infection
Journal of Infectious Diseases
Journal of Internal Medicine
Journal of Korean Medical Science
Journal of Medical Virology
Journal of Telemedicine and Telecare
Journal of The American College of Cardiology
Journal of The National Comprehensive Cancer Network
Journal of Thoracic Oncology
Journal of Thrombosis and Haemostasis
Journal of Virology
Kidney International
Korean Journal of Radiology
Lancet
Lancet Global Health
Lancet Infectious Diseases
Lancet Psychiatry
Lancet Public Health
Lancet Respiratory Medicine
Laryngoscope
Microbes and Infection
Military Medical Research
Morbidity and Mortality Weekly Report
National Science Review
Nature
Nature Communications
Nature Medicine
Nature Microbiology
New England Journal of Medicine
Obesity
Paediatric Pulmonology
Proceedings of The National Academy of Sciences of The United States of America

Psychiatry Research
Radiology
Science
Science China-Life Sciences
Translational Paediatrics
Translational Research
Viruses-Basel
World Journal of Paediatrics

Addresses for correspondence

Ken Hyland
School of Education and Lifelong Learning
University of East Anglia
Norwich, NR4 7TJ
UK

k.hyland@uea.ac.uk

Co-author information

Feng (Kevin) Jiang
School of Foreign Language Education
Jilin University

kevinjiang@jlu.edu.cn