



HAL
open science

Definitely Maybe: Hedges And Boosters in the HCI Literature

Lonni Besançon, Yvonne Jansen, Andy Cockburn, Pierre Dragicevic

► **To cite this version:**

Lonni Besançon, Yvonne Jansen, Andy Cockburn, Pierre Dragicevic. Definitely Maybe: Hedges And Boosters in the HCI Literature. 2021. hal-03342756

HAL Id: hal-03342756

<https://inria.hal.science/hal-03342756v1>

Preprint submitted on 22 Nov 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Definitely Maybe: Hedges And Boosters in the HCI Literature

Lonni Besançon
lonni.besancon@gmail.com
Linköping University, Sweden
Monash University, Australia

Andy Cockburn
andrew.cockburn@canterbury.ac.nz
University of Canterbury
New Zealand

Yvonne Jansen
yvonne.jansen@inria.fr
Sorbonne Université, CNRS, ISIR
France

Pierre Dragicevic
pierre.dragicevic@inria.fr
Université Paris-Saclay, CNRS, Inria, LISN
France

ABSTRACT

Introduction: Hedges and boosters are terms respectively used to decrease and increase the strength of statements. They are therefore essential lexical tools in scientific communication, in particular in fields conducting human-subjects experiments such as Human-Computer Interaction (HCI).

Objectives: We present an analysis of the use of hedges and boosters in the proceedings of the main HCI conference, namely CHI, between 2010 and 2018 to better understand how CHI authors report empirical findings.

Methods: We only considered papers reporting a user study and focused our analysis on the sentences in the abstract that describe empirical results. We used a program to detect boosters and hedges and manually adjusted their classifications.

Results: We found that CHI studies reporting inferential statistics are more likely to boost than they hedge.

Conclusion: Our work intends to raise awareness within the HCI community of the importance of hedging and boosting when reporting research results. Further, our results establish a baseline of the current use of these linguistic devices in HCI, adding to the current body of linguistic research. We finally contribute tools, procedures, and materials to facilitate future analyses of hedges and boosters in scientific communication.

Reproducibility: All data and scripts are available on osf.io/kefmr/

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI); User studies; Empirical studies in HCI.**

KEYWORDS

Content analysis; meta-analysis/literature survey; hedges; boosters; scientific communication.

1 INTRODUCTION

The effective interpretation and reporting of research results is a fundamental part of scientific knowledge dissemination. However, in communicating their work, scientists need to address a potential conflict between their desire to clearly and succinctly summarise

results, implications, and importance while also honestly describing uncertainty in the data, limitations in the method, and risks to scope of generalisation.

This potential for conflict is exacerbated by various factors, including reviewers', readers', and the media's desire for work that shows a clear and unambiguous contribution to knowledge. Yet science is inherently uncertain [23, 24, 35], and appropriately communicating uncertainty in a paper risks introducing ambiguity that erodes perceptions of contribution, which consequently may impair the likelihood that the work will be accepted for publication or widely disseminated if accepted.

These challenges to the effective communication of science have been broadly discussed in multiple fields [23, 35, 50], particularly in areas of widespread concern such as health [71], nutrition science [4], education research [105], and climate change [18]. Other areas of science have also raised these concerns. For example, in a recent study of over 1,500 psychology research articles, DeJesus *et al.* [35] indicated that articles tended to sacrifice precision in favour of bolder claims, and that readers rated work that expressed findings with bolder generic language as being more important than work that expressed findings more cautiously.

To better understand how scientists are communicating their findings, linguists and scientists in several disciplines have analysed the use of *hedges* and *boosters* within their respective bodies of literature. Hedging is a rhetorical means to convey the degree of uncertainty associated with a statement or an assertion [14, 23, 36]. For example, in the following sentence, the underlined words are hedges: “this might be due to [...], which suggests that [...]”. Conversely, boosting is a rhetorical means to convey the certainty and commitment from the speaker to a statement or assertion [114], such as the word ‘demonstrated’ in the following sentence: “We have demonstrated that [...]”. Hedging and boosting are important means to express degrees of belief, degrees of available evidence, or degrees of precision in research articles [113].

Researchers have argued that using hedges is a natural part of scientific communication [59, 87, 112], particularly in fields that observe human behaviour [112], as is often the case in Human-Computer Interaction (HCI) research. While ignorant use of hedges and boosters can be detrimental to scientific communication, linguists sometimes argue that authors’ use of hedges and boosters can serve a strategic role in minimising threats of criticism (e.g., [91, 107]), and they have also argued that when used appropriately hedges and boosters can assist proper communication about the true state of results’ certainty [113].

In HCI, these issues have received little attention. However, evidence from a study by Besançon *et al.* [15] suggests that HCI authors are responding to these challenges in a similar manner to that observed in psychology and other fields – with authors tending to make simple, clear, strong, boosted statements, at the cost of more nuanced and cautious hedged summaries. Through analysis of the presence of terms “significant” and “significantly” in papers, Besançon’s study indicated the prevalence of dichotomous interpretations of empirical results in the CHI proceedings from 2010 to 2018. Such dichotomous interpretations have long been considered problematic [5, 6, 29, 45, 46], but they are likely appealing to authors and many readers through their seemingly simple, unambiguous, clear and generic statement of experimental outcome.

However, Besançon *et al.*’s study only considered the terms “significant” and “significantly” through automated text processing, and it is reasonable to attribute their finding to the prevalent use of null-hypothesis significance testing (NHST) in empirical HCI work. A more comprehensive analysis is required to better understand whether authors and the HCI community have a tendency to oversimplify or augment their findings, importance, and generality (or the inverse) through use of hedges and boosters.

One barrier to the analysis of any scientific corpus is the difficulty of identifying and interpreting potential hedges and boosters [2, 51]. Previous analyses of corpora from other disciplines have therefore typically been manually conducted, applied to small samples of papers, and difficult to replicate. To ease these problems we developed tools and procedures for our own analyses, supporting a fully reproducible semi-automated approach. We then applied our tools and procedures to the analysis of 165 abstracts across five years of the CHI proceedings, with results suggesting that CHI studies which make use of inferential statistics are particularly likely to boost more than they hedge.

In this research we do not provide definitive guidance on ‘correct’ or ‘erroneous’ use of hedging and boosting in the HCI literature. Rather, our objective is, as HCI researchers ourselves, to highlight to the HCI community the importance of hedging and boosting as rhetorical means, and to establish baseline understanding of their use in our literature. The primary contributions of our work are as follows: (1) a replicable study of the use of boosters and hedges across a decade of abstracts in the CHI proceedings, (2) a suite of tools, procedures, and materials to facilitate future analyses of hedging and boosting, and (3) a companion website which demonstrates the coding tools and lets visitors try them out ¹.

2 TERMINOLOGY

In linguistics, *modality* refers to different kinds of terms which allow a writer to express “psychological or mental states regarding [a] proposition” [27]. Relevant in our context is one specific type of modal expression, *epistemic modality*, which refers to the “*degree of certainty on the part of the speaker about the truth of [a] proposition*” [27]. Many hedges and boosters fall into this class of epistemic modifiers. Other linguists have also considered that terms used to approximate or define precisely quantities, frequencies, strength, or time, and these should also be considered as hedges or boosters [42, 98, 101, 114].

The use of hedges and boosters has been identified as important in scientific communication [102], yet none of the earlier proposed definitions (e.g., [22, 62, 77, 78]) has been widely accepted, and only a few recent research articles use them (e.g., [10]). Consequently, new definitions are continually proposed (see e.g., [3, 12, 40, 88]). The difficulty in properly defining hedges and boosters partially lies in the fact that the list of linguistic expressions that can be considered hedges or boosters is extensive and diverse [28, 42]. In this article we will be assuming the following definitions, inspired from the past literature:

Hedges are understood to be expressions that are used to tone down the force of statement by making it fuzzier or more uncertain while boosters are expressions that are used to increase the strength of a statement by making it more certain or sharper.

3 RELATED WORK

Our work directly relates to research on epistemic modality, on the detection and characterisation of hedges and boosters in academic articles, on authors’ goals in employing hedges and boosters, and on their impact on readers.

3.1 Dictionaries and Typologies of Hedges and Boosters

A standardised list of hedges and boosters could greatly facilitate the analysis of hedging and boosting in text corpora. Unfortunately, defining a complete list of hedges or boosters is infeasible due to the malleable nature of natural language [28]. However, linguistics researchers who analyse the use of hedges or boosters in specific texts typically rely on custom lists of terms [14], or on lists adapted from previous work [40, 62, 108, 110]. For instance, Farrokhi and Emami [40] combined lists from several influential papers in linguistics (e.g., [55, 60, 64, 97]), while Dobakhti [37] compiled a list from more recent papers [16, 61, 63, 95, 111], but they did not publish their list or make it available to others. Additional advantages of Farrokhi and Emami’s [40] combined list of terms is that they have classified each term into its grammatical category, and they have separated boosters from hedges (which is not always done by other linguists, e.g., [108]).

The broad range of hedging or boosting terms complicates the definition of an inclusive set of grammatical categories. However, most analyses in linguistics consider at least the use of the basic grammatical categories that are adverbs, lexical verbs, modal verbs, adjectives and nouns [40, 51, 102, 110]. Recent work has proposed further expanding the grammatical categories of hedges and boosters to include other linguistic terms, such as progressive forms and agentless passives [42]. The basic grammatical categories highlighted above are often used in linguistics research to present quantitative analyses of a corpus (as done in e.g., [40, 51, 110]). In the case of hedges, researchers have additionally examined the authors’ goal in their use (i.e., whether it is content-oriented or reader-oriented) [59] and the use of compound hedges [102] (e.g., ‘results may possibly suggest’). We are unaware of prior studies into the use of compound boosters.

¹<https://hedgesandboosters.github.io>

3.2 Past Studies on Hedges and Boosters

Several studies have examined writer characteristics that may influence the use of hedging or boosting (e.g., [25, 106]). Among the important influencing factors we can find educational level and background [54, 81] as well as cultural aspects (the author’s mother tongue for instance) which have been extensively studied [25, 57, 70, 84, 85, 90, 103, 106]. To mitigate the influence of these factors, some linguists have constrained their analyses to only include papers written by native English speakers (e.g., [108]), but doing so provides only a narrow view of publication practices within the analysed domain, especially in scientific writing where papers are written for an international audience.

While automated uncertainty detection in scientific articles based on machine learning approaches has shown promise [47, 66, 82, 94, 117], linguists have argued that hedge and booster detection is inherently difficult and that automated analysis can lead to errors [2, 51]. Most linguists therefore rely on semi-automated analysis, typically combining automated term detection with validation by human coders (see e.g., [51, 60, 72, 76, 100, 110, 114]).

To the best of our knowledge, no study has focused so far on human-computer interaction. Farrokhi and Emani’s work [40] on electrical engineering research articles may be the closest to the field of human-computer interaction, but it is very unlikely that their findings can be directly applied to, e.g., CHI proceedings. Similar to our study, others have examined the use of hedging and boosting in qualitative research papers compared to quantitative ones [14, 37], noting a tendency for more hedging in qualitative work.

Past linguistics studies on hedging and boosting suffer from several limitations. One notable limitation is that few researchers provide or identify their source materials with sufficient clarity to replicate their work (e.g., [60, 72, 100, 108, 110]). In particular, details are often lacking on the hedge/booster dictionary used, and on the source texts analysed. Consequently, researchers conducting new research are impeded in their ability to use past results to contextualise their own findings (e.g., compare new findings with old ones). This difficulty is further exacerbated by unclear documentation around the procedures used for manual coding of hedge/booster use. Furthermore, the need for manual coding has limited prior studies to small samples of papers: typically 20–90 in total, and often far fewer per group. For example, in order to compare the use of hedges and boosters among eight academic disciplines, Takimoto [108] analysed seven papers per discipline. With such small sample sizes it is very difficult to carry out statistical inference and properly assess the uncertainty in the estimates. Furthermore, the selection of papers is often non-random, leading to potential bias in sample selection and further complicating the use of statistical inference methods.

The studies discussed above have focused their work on characterising researchers’ discourse in general to better understand how they convey their research. In contrast to their approach, we focus on hedges and boosters to understand the (un)certainly that authors of CHI studies convey when discussing their results. This is particularly important in research disciplines that focus on observation of human performance or patterns from small samples [87, 112]. The clear communication of such results, in particular of statistical results, has been emphasised [50] and the use of ambiguous [41, 44] or dichotomous [15] language in statistical reporting has been previously

highlighted. Recent work looking at the use of dichotomous-thinking markers “significant” and “significantly” in CHI papers showed that they were predominant, thus possibly highlighting binary interpretations of statistical results.

Finally, beyond the analysis of authors’ use of hedges and boosters to summarise their own results, there are interesting and relevant issues around how authors use these terms to characterise the work of others, when summarising related work (for example, see [34, 36, 56, 79]). While Marshall et al. [86] have examined citation practices at CHI, we are unaware of any equivalent hedge/booster analyses in the HCI literature.

3.3 Paper Abstracts

Linguists often focus their analysis on specific sections of research articles such as abstracts [76, 90], discussions (e.g., [14, 84]), conclusions (e.g., [85, 106]) or a combination of the above (e.g., [40, 89]). In fact, among linguistics studies we have previously discussed, only a handful look at complete research articles [51, 57, 60, 100, 108, 110, 112, 114]². Focusing on specific sections of articles is particularly interesting since modal terms might serve different functions in different parts of a research manuscript. In addition, limiting the analysis to a specific section helps reduce the amount of manual analysis needed and thus helps increase sample size. We decided to also restrict our analysis to the abstract of research papers. Abstracts are among the first pieces of information that researchers inspect when seeking a quick summary of relevance and findings [92]. Although they can serve a role as an “advertisement” to attract readers to the full article [47, 63, 115], in many cases, they represent the only part of a manuscript that is read [26, 33, 67, 80]. Abstracts are also often used to establish the relevance of an article [93]. Because of their advertising role, they sometimes misrepresent the findings of an article with exaggeration or misreporting (see the work on “spin” in the medical literature, e.g., [11, 68]). Additionally, previous work has found that when citing articles, authors do not necessarily read the entire paper [74] and after the methods and results section, the abstract was the one that gathered the most attention. Considering this tendency for selective and sometimes hasty reading by researchers [99], it is reasonable to assume that abstracts’ misrepresentation of results could likely result in misrepresentation of results in the trail of citations of a paper. Finally, abstracts are rarely hidden behind a paywall, so they are more broadly accessible to the public [9, 116].

4 RESEARCH QUESTIONS

Our goal is to conduct exploratory research on the use of hedges and boosters in the proceedings of the main venue for Human-Computer Interaction (HCI) research, the CHI conference. We want to better understand how hedges and boosters are used in HCI when summarising empirical results (and how this use may change across years), which includes examining whether there are stereotypical patterns of use, and whether any detected patterns vary across different types of papers (such as those using inferential statistics versus purely descriptive studies). Among the motivations for conducting this analysis is our hope that better understanding of hedge/booster use can contribute to awareness, guidelines, and cautions for the

²A breakdown of each paper’s focus is available in supplementary materials.

ways in which the HCI research community reports its research findings, thereby improving the quality of research dissemination.

Our major focus will be on comparing studies that report statistics such as p -values or confidence intervals (which we will refer to as “inferential/mixed”) with studies that do not (which we will refer to as “purely descriptive”). There are reasons to expect authors who report inferential statistics to be *more* careful in their conclusions, because reporting inferential statistics is a way of conveying uncertainty, and can be an indication that authors recognise the existence of uncertainty in their data. However, there are equally plausible reasons to expect inferential reports to be *less* careful and more assertive in how they report their conclusions, because inferential statistics are frequently misused to convey a false sense of certainty [6, 15, 45]. This would be consistent with a past study suggesting that quantitative studies are more assertive than qualitative ones [14]. We are also interested in seeing whether the use of hedges and/or boosters has evolved over the years. The replication crisis had led many methodologists to call for a more nuanced reporting of findings over the past few years [7, 31, 48]. Perhaps these methodologists have had some influence in HCI, although one prior analysis suggests that adoption of new statistical reporting habits has been slow at CHI [15].

5 METHOD

We analysed paper abstracts from the main proceedings of ACM CHI for each of the years 2010, 2012, 2014, 2016 and 2018. We chose even years to reduce coding workload and to permit a wider range of years analysed. The analysis was conducted by three coders, all co-authors of this article. The three coders are nationals of non-English European countries but are all proficient in English and have 6, 13 and 21 years of experience writing and reading scientific articles in the field of HCI. The analysis procedure described below as well as the analysis in the section “Results” were planned ahead of time and pre-registered on the Open Science Framework website³.

The procedure is summarised in Figure 1 and details are provided on our [companion website](#). We started by randomly sampling 33 CHI papers for each year, using a pseudo-random generator (see python code in the [OSF repository](#)). All papers from the main proceedings were considered, which includes long and short papers. The rationale for $N=33$ was the following: we planned to use non-parametric bootstrapping to derive interval estimates, and these intervals are known to be reasonably accurate for sample sizes of approximately 20 or more [73]. Since previous analyses of CHI proceedings suggest that 80–90% of all CHI papers report an empirical study [13, 19, 75], we decided to use $N = 30 \div 0.9 \approx 33$ to get about 30 papers with a study. With 30 papers per year we would analyse 150 papers, which is more than twice the sample size of a typical study on booster/hedges in the linguistic literature. Indeed, among the many studies we examined, the mean number of articles analyzed is 58 (median 56, range [20–200], full data available on our [OSF repository](#)). The list of selected papers and their DOIs is available on the [OSF repository](#).

5.1 Coding step 1: Finding papers reporting a study

This first step was carried out with the help of a Google spreadsheet, which was populated with the 165 (5×33) papers for each analysed

year. Each coder then scanned the entire PDF of each paper to determine whether or not it reported a study (binary coding, Yes or No). Additionally, to answer our research question about the possible differences between qualitative and quantitative research papers, each coder had to report whether the study (if any) reported inferential statistics (again, binary coded). Coders relied on the following definitions:

A *study* is a paper that reports research involving human participants. The report must be on an original study: surveys of previous studies, discussions of previous studies, and re-analyses of previous studies do not qualify.

An *inferential/mixed study* is a paper that reports results using measures of statistical uncertainty such as test statistics, p -values, interval estimates (confidence intervals, credible intervals, standard errors, but *not* standard deviations), Bayesian posteriors, or Bayes factors. All of these can be reported either numerically or graphically. It is not necessary that the paper *only* reports inferential statistics.

All studies that were not marked as inferential/mixed were automatically considered *purely descriptive*. Note that we used the shorter terms *inferential* and *descriptive* in our codebook, but here we use *inferential/mixed* and *purely descriptive* for clarity reasons.

The decision to classify papers as inferential/mixed vs. purely descriptive instead of, for example, quantitative vs. qualitative stems from the difficulties we experienced operationalising the concepts of qualitative and quantitative. Previous analyses of HCI research have used a similar approach, classifying papers as using statistical analyses or not [19, 20]. Another difficulty with the notions of quantitative and qualitative is that many so-called quantitative studies report some qualitative data, and vice versa. Furthermore, the classification of papers as quantitative or qualitative will vary dependent on the exact criteria used [8, 17, 96], including the acquisition methodology, the nature of data, the presentation of data, and the nature of analysis or even the original philosophical mindset and approach behind the tested theory (positivism or interpretivism) [8, 38, 104].

In contrast, it is relatively easy for a researcher with experience in statistics to find whether a paper reports inferential statistics. Nevertheless, the distinction between inferential/mixed and purely descriptive correlates in practice with the one between quantitative and qualitative. This is because many qualitative papers (e.g., ethnographic studies) do not report inferential statistics and will therefore be classified as purely descriptive. Conversely, quantitative studies generally report inferential statistics and will therefore be classified as inferential/mixed. One consequence of our definition is that a mixed-methods paper that focuses on qualitative reporting but also reports a couple of p -values will be classified in the same category as a purely quantitative paper.

Once all three coders completed their analysis, we retained only those papers that at least two coders marked as including a study. We then used voting to decide whether a study was inferential/mixed or purely descriptive. Papers that were marked {non-study, inferential/mixed, purely descriptive} by the three coders (two instances)

³OSF project: <https://osf.io/kefmr/>; Preregistration: <https://osf.io/up6kw>

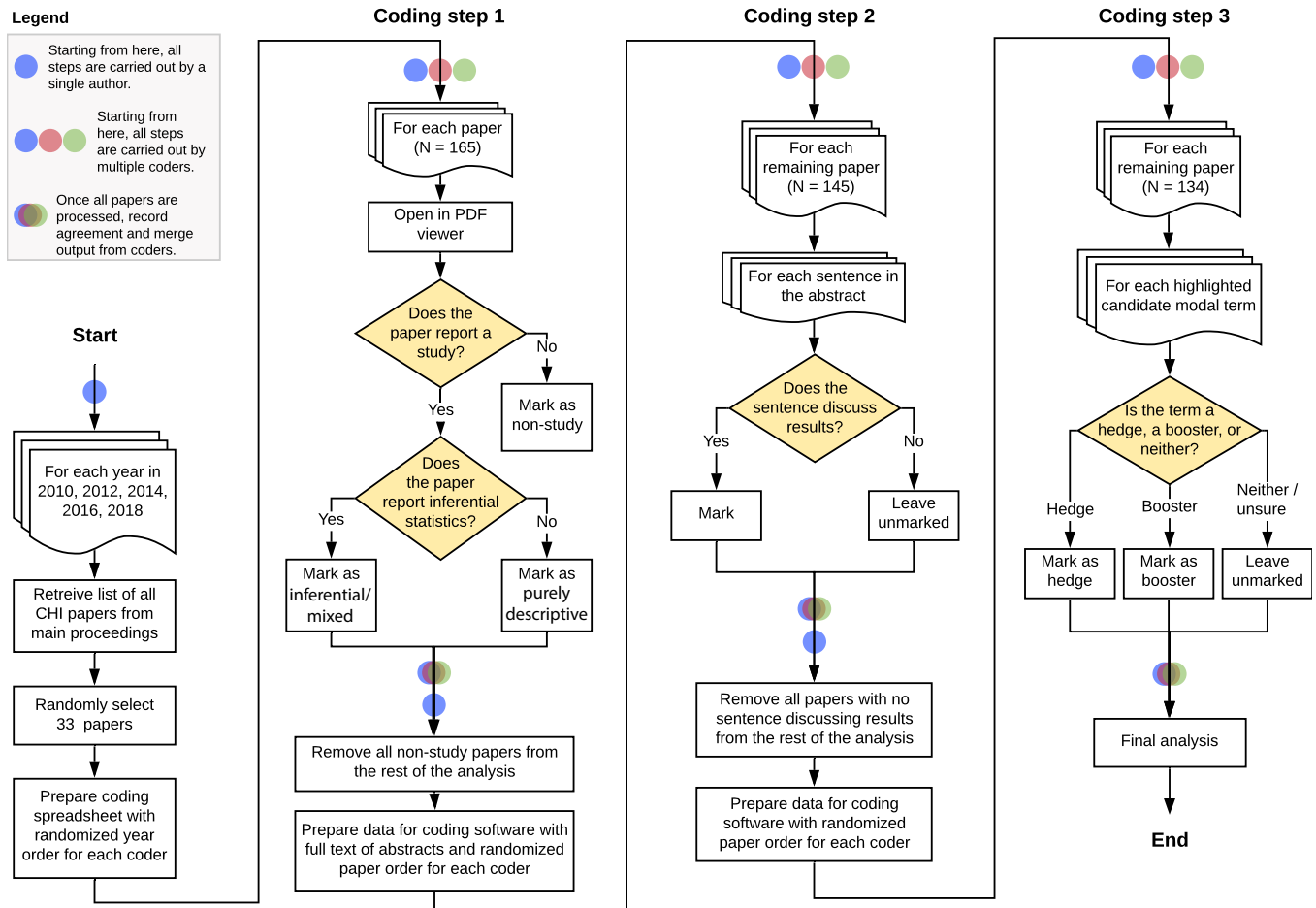


Figure 1: Flow chart describing the entire coding procedure. The option to report extra hedges or boosters in Coding step 3 is left out for simplicity.

were also excluded from later analyses. This left 145 papers from the initial 165 (12% removed).

5.2 Coding step 2: identifying result-discussion sentences

Coding step 2 used custom web apps that we built to assist the analysis of paper abstracts (see Figure 2a). The custom apps built for our analyses are available in the supplementary materials. In step 2, each coder read every sentence of each abstract, and applied a binary tag for whether or not it discussed study results. Paper order was randomized in order to avoid interactions between publication year and coder fatigue or learning. Sentence order was not randomized within abstracts because we thought context would assist understanding individual sentences. Coders used the following definition (excerpts – for the full text, please see the supplementary material):

A result description is a sentence that describes or summarises the results of a study, explains them, or discusses their implications. As a general rule, it is a sentence that may have been different had the results been different. Both

sentences in the following excerpt are result-relevant: “We found that touch interaction is slower but more precise than tangible interaction. This suggests that designers and developers should consider this tradeoff when designing applications”. [...] However, we exclude the description of further research steps that were carried out based on the results of a study and reported in the same paper [...].

At the end of step 2, we only kept sentences that at least two coders identified as discussing results. We removed papers in which the abstract did not have any sentence discussing results. This left us with 134 papers for step 3 (8% removed).

5.3 Coding step 3: identifying boosters and hedges

In this step, the coders used a second custom web app⁴ (Figure 2b) to identify hedges and boosters in the sentences describing results. The app initially highlighted terms within each sentence, based on the predefined list of hedges and boosters from Farrokhi and Emami

⁴The web app can be tried out under hedgesandboosters.github.io/step3.html.

Sentences	Contains a result description?
We first conduct a study using abstracted video content to measure the effects of latency on video scrubbing performance and find that even very small amounts of latency can significantly degrade navigation performance.	N
Based on these results, we present Swift, a technique that supports real-time scrubbing of online videos by overlaying a small, low resolution copy of the video during video scrubbing, and snapping back to the high resolution video when the scrubbing is completed or paused.	N
A second study compares the Swift technique to traditional online video players on a collection of realistic live motion videos and content-specific search tasks which finds the Swift technique reducing completion times by as much as 72% even with a relatively low latency of 500ms.	Y

(a)

We find that this juxtaposition of concepts helps explain a highly generative creative practice—toolmaking—within the hackerspace we studied.

2018 - paper443

Results highlighted between-diagnostic-group differences in playing patterns with children with ASD showing broad patterns of EF deficits, but with relative strengths in nonsocial short-term memory, and preserved response to emotional inhibition cues.

We showed the system could predict IQ, an important target for clinical treatment, towards the goal of developing platforms to act as long-term, efficient, and effective behavioral biomarkers for ASD.

(b)

Figure 2: Screenshots of the web-app used for the coding step 2 (Figure 2a) and step 3 (Figure 2b). During step 2, sentences containing results discussions are highlighted by the coder in blue. During step 3, possible modal terms are highlighted in gray, which the coder could turn red to indicate a booster, or blue to indicate a hedge.

[40]. This list, which consists of 170 hedges and 85 boosters, was corrected for a few obvious errors (see supplementary material) and augmented to include all inflections of each word (e.g., plural of nouns, conjugations of verbs, and contraction of negative modal verbs such as couldn't and could not). Adding inflections extended this list to 400 hedges and 176 boosters.

To prevent coders from being influenced by the initial hedge/booster classification, all candidate hedges and boosters were highlighted in the same gray color. Each coder had to determine whether every highlighted term was a booster, a hedge, or neither (Figure 2b). Coders were instructed to focus on the terms that were already highlighted, but they were also invited to report possible hedges and boosters that were not highlighted, i.e., that were absent from Farrokhi and Emami's list (analysed separately). Like before, the order in which papers were presented to coders was randomised for each coder.

Coders used the following definitions (see supplementary material for full text):

A modal term is a term that is used either to tone down a statement, or to increase its strength. Examples are the word may in "The drug may have an effect on sleep" or the word definitely in "The drug definitely has an effect on sleep". However, the word definitely is not considered a modal term in the sentence "One participant said she would definitely use the tool in the future", because changing it would alter the meaning of the statement. As a general rule, if a term

can be removed or replaced by another term in a way that alters the subjective strength of the statement made by the authors without altering its objective meaning, then it is a modal term. [...]

A *hedge* is a term that is used to tone down a statement, for example by making it appear more fuzzy or more uncertain. Examples of hedges include possibly, approximately, sometimes or speculate. For example, the sentence "The drug may have an effect on sleep" is weaker than the simpler sentence "The drug has an effect on sleep", so may is a hedge in the first sentence. [...]

A *booster* is a term that is used to increase the strength of a statement, for example by making it appear sharper or more certain. It has the opposite effect of a hedge. Examples of boosters include definitely, strongly, exactly or prove. For example, the statement "The drug has a clear effect on sleep" appears stronger than "The drug has an effect on sleep", so clear acts as a booster in the first sentence. [...]

In addition to these definitions, coders were instructed to leave the term as unmarked (in gray) for cases where it was difficult to decide whether or not the term was a modal, or whether the term was a hedge or a booster.

Once this step was completed, we used a voting scheme whereby a term was considered a hedge or a booster if at least two coders marked it as such. We kept all 134 papers for the final analysis of hedge and booster frequency, irrespective of whether or not the papers contained hedges/boosters.

The complete three-step coding procedure took about four hours per coder (≈ 1.5 minutes per paper). Our web apps were critical in streamlining the analysis.

6 RESULTS

As mentioned previously, all the analyses reported in this section were planned and preregistered before data collection. Due to the limits of null hypothesis significance testing for assessing statistical evidence [6, 31, 39], and as indicated in our preregistration, we report and interpret all our results using an estimation approach, i.e., with a focus on effect sizes and their interval estimates.

6.1 Analysis of Study types

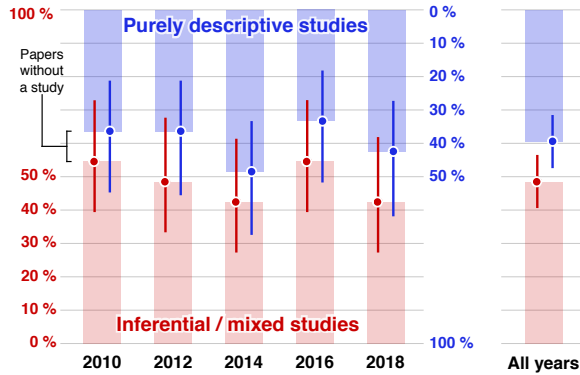


Figure 3: Proportion of CHI papers reporting inferential/mixed studies (red) and purely descriptive studies (blue). Error bars are 95% confidence intervals.

The first step of the coding process aimed at determining whether (1) a paper reports a study involving human participants and (2) if so, whether the paper reports inferential/mixed or purely descriptive results.

In all the following, the reported confidence intervals for Fleiss kappa were computed using Gwet’s normal approximation assuming random items and fixed raters [49]. The three coders reached a relatively high agreement on classifying papers, with a Fleiss kappa of 0.80, 95% CI = [0.74, 0.87]. More specifically, for the coding of whether the paper included a study, we obtained a Fleiss kappa of 0.73, 95% CI = [0.59, 0.86]; and for whether a study paper was inferential/mixed or purely descriptive, we obtained a Fleiss kappa of 0.86, 95% CI = [0.79, 0.93].

Figure 3 summarises our estimates for the proportion of papers reporting studies of each type across the 5 years analysed, as well as an average for all years. Overall, about 90% of CHI papers report a study. The proportion of papers including inferential/mixed results seems to be similar to the proportion reporting purely descriptive results, with weak evidence that inferential/mixed papers are slightly more numerous. There does not seem to be any clear trend over

years, although per-year interval estimates are wide and thus larger sample sizes are needed to examine possible trends. These estimates align with previous studies [13, 20, 75].

6.2 Analysis of abstract sentences

As mentioned in the previous section, the second step of the coding process consisted of examining sentences in the abstracts and coding whether or not they discussed results. Again, inter-coder agreement was high: Fleiss kappa 0.85, 95% CI = [0.82, 0.88].

Figure 4 summarises the proportion of sentences in the abstract discussing results (roughly 30%), with little discernible variation across years, and no striking difference between papers reporting inferential/mixed and purely descriptive results.

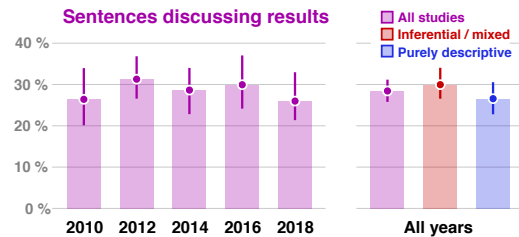


Figure 4: Proportion of sentences in the abstracts of CHI studies that discuss the results of the study. Error bars are 95% CIs.

Somewhat surprisingly, 11 papers (approx. 8%) that included a study made no mention of their results in the abstract. As explained in the previous section, we removed these 11 papers from the rest of the analysis⁵, leaving 134 papers.

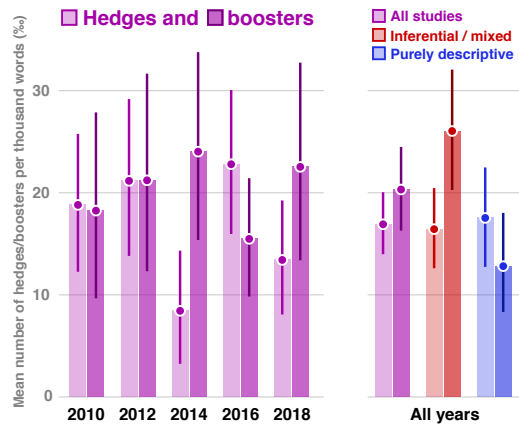


Figure 5: Average number of hedges and boosters in sentences discussing study results, for every 1000 words. Error bars are 95% CIs.

6.3 Analysis of hedging and boosting

The outcomes of step 3 – the coding of hedges and boosters – were our main research objective. For this final coding step we obtained a good inter-coder agreement, with a Fleiss kappa of 0.72, 95% CI = [0.67, 0.76]. More specifically, we obtained a medium agreement of

⁵See our [OSF repository](#) for the list of papers.

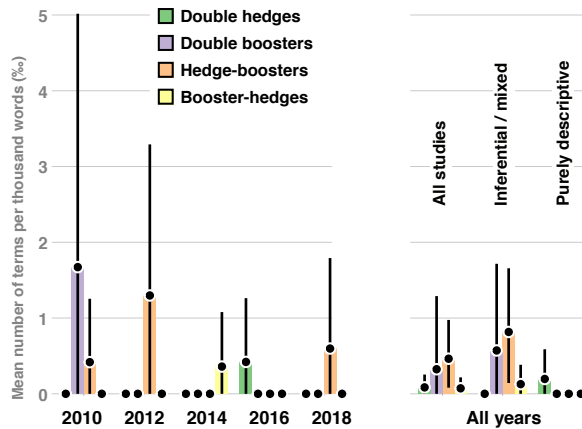


Figure 6: Average number of compound modal terms in sentences discussing study results, for every 1000 words. Error bars are 95% CIs.

0.66, 95% CI = [0.61, 0.72] for deciding whether a term is modal (hedge/booster) or not; and a much higher agreement of 0.92, 95% CI = [0.88, 0.97] for identifying whether a modal term is a hedge or a booster.

Figure 5 shows the mean number of hedges and boosters per thousand words in the sentences that discuss study results (words per thousand words is a widely used metric in linguistics, e.g., [40, 60, 108]). While mean values are relatively stable across years (at around 20 words per thousand), there is a disproportionately low use of hedges in 2014 – more data would be needed to determine whether this was a systematic effect at CHI that year, or merely a random sampling anomaly. Given that this dip is hard to explain, a sampling anomaly is more likely. A more robust suggestion in the results (shown in the ‘All years’ summary at the right of Figure 5) is that inferential/mixed studies use more boosters than hedges. Even more clearly, inferential/mixed studies use more boosters in their abstract than purely descriptive studies on average (approximately twice as much). Meanwhile, there is no evidence that the frequency of hedge usage differs between inferential/mixed and purely descriptive studies.

Figure 6 shows the mean number (per 1000 words) of compound hedges and boosters in abstract result sentences. Unsurprisingly, the means are much lower, and the uncertainty around estimates is high. While patterns across years are noisy, the combined ‘All years’ data suggests that papers reporting inferential/mixed studies use more compound hedges and boosters than papers reporting purely descriptive studies.

6.4 Additional Analyses

In order to better characterise the quantitative summaries reported previously, we provide complementary analyses, specific examples and observations from our dataset. The analyses in this section are not part of our preregistered plan.

The histograms in Figure 7 show the underlying distributions of hedges and boosters for the two study types, inferential/mixed and purely descriptive. The variability across papers is very high, explaining why the interval estimates in Figures 5 and 6 are so wide. We observe, in particular, that almost half of all articles make no use

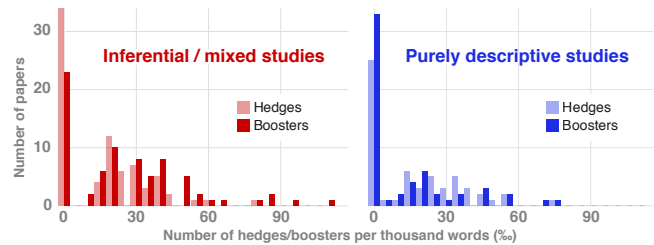


Figure 7: Distributions of hedge count and booster count for all inferential/mixed studies (N=76) and all purely descriptive studies (N=58).

at all of hedges or boosters in their abstract. Moreover, articles that use no booster at all are more prevalent among purely descriptive studies, which is consistent with our finding that inferential/mixed studies use more boosters on average.

The following two quotes are taken from the final two sentences of the abstracts with the most hedges (80%, meaning 80 words per 1000) and boosters (111%, i.e., more than one word out of ten). Each hedge is underlined in the first quote, and each boost is underlined in the second.

Results indicated that, in general, contextual information provides more useful information for making expertise judgment of Twitter users than tweets. While the addition of tweets seems to make little difference, or even add nuances to novices’ expertise judgment, experts’ judgments were improved when both content and contextual information were presented.

Our results show both accuracy and response time in color-matching tasks were significantly better with SSMRecolor. This work demonstrates the value of a situation-specific approach to recoloring, and shows that this technique can substantially improve the usability of color displays for users of all types.

We looked at the most common hedges and boosters in our dataset, which are reported in Table 1, for both inferential/mixed and purely descriptive studies. In this table, we can see that there does not seem to be a clear difference between inferential/mixed and purely descriptive studies on the use of hedges. However, the lower part of the table highlights that show, significant and their inflections are omnipresent in inferential/mixed studies are therefore likely to account for the differences we have found between purely descriptive and inferential/mixed papers. To further investigate that possibility, we report in Figure 8 the differences in number of boosters between inferential/mixed and purely descriptive studies for all data and for different subsets where those common boosters have been removed. As previously mentioned (Figure 5), when considering all boosters, our data provides good evidence that inferential/mixed studies use more boosters. When removing all occurrences of show and its inflections, the difference becomes smaller although weak evidence of a difference still remains.

As mentioned above, our coding interface allowed coders to report terms that they thought were hedges or boosters, even if they were not initially highlighted by the dictionary-based automatic classification.

Top hedges in inferential/mixed studies:	
can	14
suggest, suggests, suggesting	10
may	6
could	5
potential	4

Top hedges in purely descriptive studies:	
can	6
indicate, indicated, indicating	4
could	4
primarily	4
about	2

Top boosters in inferential/mixed studies:	
show, showed, shows	41
significantly, significant	19
demonstrate, demonstrates, demonstrated	12
confirm, confirmed	3
proved, prove	3

Top boosters in purely descriptive studies:	
show, shows	10
demonstrate, demonstrating, demonstrates	8
conclude	2
considerable	2
confirmed	1

Table 1: Most common hedges (top) and boosters (bottom) by study type, grouped by lexeme, with their number of occurrences across all 134 papers.

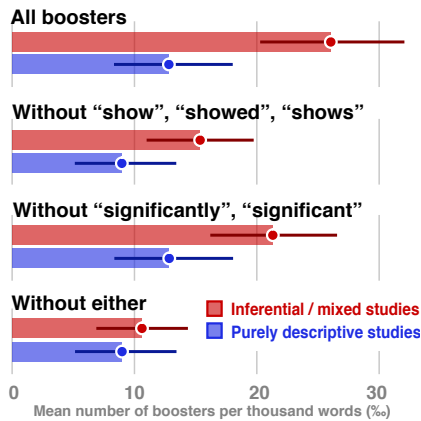


Figure 8: Use of boosters in inferential/mixed vs. purely descriptive studies, for all data and for subsets where common boosters were removed. Error bars are 95% CIs.

The term postulate was reported by all three coders as a hedge in specific instances, and although was reported as a hedge by two. As for boosters, all three coders reported the terms reveal/revealed and validate, and two of them reported a lot, as much as, and highlight. The term illustrate was reported by all three coders as a modal term, but two marked it as a hedge while one marked it as a booster.

Similarly, empirical, empirically and much were reported by two coders, but with disagreement as to whether they were hedges or boosters.

7 DISCUSSION

Our analysis has several limitations, opens avenues for future work, and has implications for HCI researchers. We discuss each of these below.

7.1 Limitations and Future Work

There are 5 main limitations to our work that we will detail and address below. The first four limitations directly concern risks to generalisation of our results.

We see five main limitations to our work, discussed below.

First, our sample size is relatively small, which was necessary to keep coding effort manageable. The 165 papers we analysed represent about 7% of the papers published in the five proceedings included in our study, and although our sample was larger than that used in most analyses of hedges and boosters (e.g., [60, 108, 110]), a larger sample is needed to obtain more precise estimates and to detect subtler trends.

Second, we limited our analysis to abstracts rather than full papers. This was done for several reasons: doing so made the analysis workload manageable; abstracts serve an important role in advertising the paper [32, 47, 63, 92]; they are often the only part of the paper that is read [26, 80]; and they are seldom behind a paywall [9]. Nonetheless, analysis of full articles is an interesting avenue for future work, in particular to place the focus on sections describing results, discussions and conclusions. Focusing on related work sections would also provide interesting findings that could be contrasted with Marshall *et al.*'s results on throwaway citations of prior work at CHI [86].

Third, we limited our analysis to CHI papers, raising questions of how the findings generalise to other HCI research. CHI is the biggest venue of the HCI community and is often considered a prime venue for HCI work, and one that the community is likely to read. These factors have been considered in the past to explain how analysis of CHI papers can generalise to the field of Human-Computer Interaction as a whole [1, 13, 19, 53, 83, 86].

Fourth, our coding method classified papers that contained a study as either inferential/mixed or purely descriptive, with this classification determined by the presence or absence of at least one measure of statistical uncertainty (e.g., *p*-value). While this classification facilitates simple and rapid coding, it provides a coarse division that will fail to reveal nuances in reporting styles.

Finally, it is difficult to compare our results with those from previous studies. While other researchers have, like us, limited their analyses to paper abstracts (e.g., [47, 76, 90]), our focus went further in scrutinising only those sentences that referred to results. Furthermore, previous studies have indicated domain dependence in the use of hedges and boosters (e.g., [40, 60, 108]), complicating comparison across disciplines. Nonetheless, our results are consistent with previous studies suggesting higher assertiveness in quantitative (i.e., inferential/mixed) studies than qualitative (i.e., purely descriptive) [14]. However, the rhetorical strategies achieving this effect may differ between disciplines – while Behnam *et al.* [14] found that

qualitative papers in applied linguistics made greater use of hedges than in quantitative papers, our results suggest that qualitative papers in HCI make lower use of boosters than in quantitative papers.

The observation of higher boosting in quantitative studies is particularly interesting. Inferential statistics necessarily incorporate the concept of uncertainty, so one might expect authors using inferential statistics to carefully hedge their phrasing around statistical outcomes. However our findings indicate the opposite, suggesting that HCI authors have a tendency to use inferential statistics as an uncertainty-eraser – “we found $p < .05$ so we can say that our conclusions are true” – reflecting Gelman’s statement [43]:

Statistics is often sold as a sort of alchemy that transmutes randomness into certainty, an “uncertainty laundering” that begins with data and concludes with success as measured by statistical significance.

The table of most common hedges and boosters (Table 1) shows that most boosting stems from use of the terms *significant* and *show*, and their respective inflections. Removing either of these terms led to a smaller difference in the use of boosters between inferential/mixed and purely descriptive studies, and removing both from the data led to similar usage of boosters between the two types of studies. Since the word *significant* has a specific statistical meaning [41] that differs from its normal use, its categorisation as a booster is subjective and our results need to be analysed in this light.

7.2 Implications for HCI research

We see several *potentially* (hedge) *important* (booster) outcomes and implications from this work. First, adding the terms ‘hedging’ and ‘boosting’ to HCI researchers’ lexicon is in itself a form of contribution. The Sapir-Whorf hypothesis, or Wharfianism [21], proposes that our language influences the ways in which we think about things, and certainly having succinct terms to capture these concepts improves our ability to communicate around them. If authors and reviewers are sensibly alerted to the use of hedges and boosters, results dissemination in our field could improve.

Second, while the quantitative results we present may (or may not) align with the reader’s expectations, there is value in establishing a baseline of current use of these terms in our field. Research disciplines have a tendency to follow fashions and adopt new norms, and as they do so it is easy to imagine authors becoming more categorical in their claims (perhaps in an attempt to appeal to reviewers for acceptance, or to argue for the work’s importance). Again, understanding the role that hedges and boosters play in results communication should assist disciplines in their self awareness regarding changing norms across years.

Third, past research [52] has indicated that readers’ interpretation of uncertainty is influenced by the methods used to display and convey statistical results. This highlights the importance of HCI work on uncertainty visualizations (see [58]), and in particular the relevant work on visualization of inferential statistics (e.g., [30, 52]).

Fourth, our analysis of CHI abstracts revealed a surprisingly wide diversity in the way authors summarise their work, from long abstracts including extensive background motivation, to brief method summaries. We were also surprised that 8% of sampled papers (reporting studies) included no indication of results discussions in the

abstract. Recently, CHI has increasingly encouraged and/or required shorter abstracts, with CHI 2020 imposing a character length limit for the first time, and CHI2021 limiting abstracts to 150 words. One interesting possibility that HCI venues may consider in the future is the use of structured abstracts for papers describing empirical studies. Structured abstracts have become increasingly common in the medical sciences, requiring authors to provide a short statement consisting of around 50 words under each of several categories, typically Introduction, Objectives, Methods, Results, and Conclusion.

Fifth, we have developed and disseminated a suite of tools and procedures that are available as supplementary materials for others to use within and beyond HCI. While developing these tools and procedures involved extensive work and background research, using them to conduct the analysis reported in this paper was reassuringly viable in terms of the coding effort required. We initially reached out to several authors in an attempt to obtain existing data resources from previous work but we were unsuccessful. Our procedures, tools, and vocabularies and corpus may not be perfect, but they are publicly available, can be used for replication or new analyses, and can be further developed by future researchers.

7.3 Hedge/Boost guidelines for authors

When we started our analysis, we hoped that we would identify clear patterns in reporting styles, and that by doing so we could provide useful guidelines for authors. However, while certain patterns have emerged we feel that firm guidelines would be as harmful to some authors as they are beneficial to others. For example, we might consider introducing the guideline “minimise boosting in inferential and mixed-methods studies” to counter the higher observed use of boosting in these studies; but any such guidance might be inappropriate in certain contexts, such as when reporting a robust confirmation of an important and novel finding.

While we abstain from giving firm guidelines, we relay previous recommendations from linguists and statisticians. First, statistical significance relies on a somewhat arbitrary and malleable α level, thus we recommend avoiding the use of “statistically significant” when possible. Instead authors should simply report the associated data, including a measure of central tendency (e.g., mean), and confidence intervals or outcomes of statistical tests (e.g., T- or F-statistic, degrees of freedom, and p value) [5, 6, 29, 45, 46].

Second, *embrace uncertainty* when communicating results [109]. Be wary of using statistics to convey certainty, and instead contemplate and communicate the importance of factors influencing uncertainty, including sample size, noise in the data, effect size, or whether the sample is representative. Consequently, scientists should adapt their use of modal terms in order to reflect as closely as possible and strength of evidence that their experimental setup and results provide, and the uncertainty that they should convey [62].

Third, authors should also be careful to hedge and boost appropriately when citing prior work in order to accurately represent the original work and avoid misrepresenting the (un)certainty around prior findings (see e.g., [65, 69]).























8 CONCLUSION







Hedging and boosting are rhetorical means that can change the strength of a statement – hedging tones statements down, while


























boosting increases their strength. Understanding the ways in which these terms are used can provide insight into the reporting norms of different research disciplines and sub-disciplines, including how they evolve over time. However, among the difficulties of analysing hedging and boosting strategies are the lack of established vocabularies, the lack of suitable tools for analysis, the lack of well documented methods for analysis, and the lack of baseline data.




The work reported in this paper eases each of these paucities: our supplementary materials include vocabularies, tools, and methods for future analysis of hedging and boosting in any corpus; and our results provide a baseline of quantitative measures for Human Computer Interaction research, including the suggestion that CHI studies reporting inferential statistics are particularly likely to boost more than they hedge. As HCI transitions (hopefully) to an era of greater transparency and openness in its research methods, it is also (hopefully) likely that fewer papers will report boosted categorical findings on the basis of a single study. Better understanding and awareness of how we use hedges and boosters can only positively contribute to this transition. Definitely (maybe).

REFERENCES

- [1] Jacob Abbott, Haley MacLeod, Novia Nurain, Gustave Ekobe, and Sameer Patil. 2019. Local Standards for Anonymization Practices in Health, Wellness, Accessibility, and Aging Research at CHI. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–14. <https://doi.org/10.1145/3290605.3300692>  **Open Access version available.**
- [2] Shashank Agarwal and Hong Yu. 2010. Detecting Hedge Cues and Their Scope in Biomedical Text with Conditional Random Fields. *J. of Biomedical Informatics* 43, 6 (Dec. 2010), 953–961. <https://doi.org/10.1016/j.jbi.2010.08.003>
- [3] Erdem Akbas and Jan Hardman. 2018. Strengthening or Weakening Claims in Academic Knowledge Construction: A Comparative Study of Hedges and Boosters in Postgraduate Academic Writing. *Educational Sciences: Theory and Practice* 18, 4 (2018), 831–859.  **Open Access version available.**
- [4] David B. Allison, Josep Bassaganya-Riera, Barbara Burlingame, Andrew W. Brown, Johannes le Coutre, Suzanne L. Dickson, Willem van Eden, Johan Garssen, Raquel Hontecillas, Chor San H. Khoo, Dietrich Knorr, Martin Kussmann, Pierre J. Magistretti, Tapan Mehta, Adrian Meule, Michael Rychlik, and Claus Vögele. 2015. Goals in Nutrition Science 2015–2020. *Frontiers in Nutrition* 2 (2015), 26. <https://doi.org/10.3389/fnut.2015.00026>  **Open Access version available.**
- [5] Valentin Amrhein, Sander Greenland, and Blake McShane. 2019. Scientists rise up against statistical significance. *Nature* 567, 7748 (2019), 305–307. <https://doi.org/10.1038/d41586-019-00857-9>  **Open Access version available.**
- [6] Valentin Amrhein, Fränzi Korner-Nievergelt, and Tobias Roth. 2017. The earth is flat ($p>0.05$): Significance thresholds and the crisis of unreplicable research. *PeerJ Preprints* 5 (June 2017), e2921v2. <https://doi.org/10.7287/peerj.preprints.2921v2>  **Open Access version available.**
- [7] Valentin Amrhein, David Trafimow, and Sander Greenland. 2018. Inferential Statistics as Descriptive Statistics: There is No Replication Crisis if We Don't Expect Replication. *The American Statistician* (2018). <https://doi.org/10.7287/peerj.preprints.26857v4>  **Open Access version available.**
- [8] Stephen Kwadwo Antwi and Kasim Hamza. 2015. Qualitative and quantitative research paradigms in business research: A philosophical reflection. *European journal of business and management* 7, 3 (2015), 217–225.  **Open Access version available.**
- [9] Yusuf Assem, Sam Adie, Jason Tang, and Ian A. Harris. 2017. The overrepresentation of significant p values in abstracts compared to corresponding full texts: A systematic review of surgical randomized trials. *Contemporary Clinical Trials Communications* 7 (2017), 194–199. <https://doi.org/10.1016/j.conctc.2017.07.007>  **Open Access version available.**
- [10] Laura Aull. 2015. Connecting writing and language in assessment: Examining style, tone, and argument in the U.S. Common Core standards and in exemplary student writing. *Assessing Writing* 24 (2015), 59–73. <https://doi.org/10.1016/j.asw.2015.03.002>
- [11] J. Austin, C. Smith, K. Natarajan, M. Som, C. Wayant, and M. Vassar. 2019. Evaluation of spin within abstracts in obesity randomized clinical trials: A cross-sectional review. *Clinical Obesity* 9, 2 (2019), e12292. <https://doi.org/10.1111/cob.12292>  **Open Access version available.**
- [12] Nida Ul Habib Bajwa, Cornelius J. König, and Octavia S. V. Harrison. 2016. Toward Evidence-Based Writing Advice: Using Applied Linguistics to Understand Reviewers' Expectations. *Academy of Management Learning & Education* 15, 3 (2016), 419–434. <https://doi.org/10.5465/amle.2015.0002>  **Open Access version available.**
- [13] Louise Barkhuus and Jennifer A. Rode. 2007. From Mice to Men - 24 Years of Evaluation in CHI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '07). ACM, New York, NY, USA, Article 1. <https://doi.org/10.1145/1240624.2180963>  **Open Access version available.**
- [14] Biook Behnam, Amin Naeimi, and Ali Darvishzade. 2012. A comparative genre analysis of hedging expressions in research articles: is fuzziness forever wicked? *English Language and Literature Studies* 2, 2 (2012), 20. <https://doi.org/10.5539/ells.v2n2p20>  **Open Access version available.**
- [15] Lonni Besançon and Pierre Dragicevic. 2019. The Continued Prevalence of Dichotomous Inferences at CHI. In *CHI '19 - Proceedings of CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, Glasgow, United Kingdom, 1–11. <https://doi.org/10.1145/3290607.3310432>  **Open Access version available.**
- [16] Douglas Biber. 2006. Stance in spoken and written university registers. *Journal of English for Academic Purposes* 5, 2 (2006), 97–116. <https://doi.org/10.1016/j.jjeap.2006.05.001>
- [17] Julia Brannen. 2017. *Mixing methods: Qualitative and quantitative research*. Routledge, Abingdon-on-Thames, UK. <https://doi.org/10.4324/9781315248813>
- [18] David V. Budescu, Han-Hui Por, and Stephen B. Broomell. 2012. Effective communication of uncertainty in the IPCC reports. *Climatic Change* 113, 2 (01 Jul 2012), 181–200. <https://doi.org/10.1007/s10584-011-0330-3>
- [19] Kelly Caine. 2016. Local Standards for Sample Size at CHI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) (CHI '16). ACM, New York, NY, USA, 981–992. <https://doi.org/10.1145/2858036.2858498>  **Open Access version available.**
- [20] Paul Cairns. 2007. HCI... Not As It Should Be: Inferential Statistics in HCI Research. In *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI... But Not As We Know It - Volume 1* (University of Lancaster, United Kingdom) (BCS-HCI '07). British Computer Society, Swinton, UK, UK, 195–201. <http://dl.acm.org/citation.cfm?id1531294.1531321>  **Open Access version available.**
- [21] John B. Carroll. 1956. *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf*. The MIT Press.
- [22] J Channel, J Sinclair, and R Carter. 1994. Vague Language (Describing English Language).
- [23] Chaomei Chen and Min Song. 2017. *Representing scientific knowledge*. Springer. <https://doi.org/10.1007/978-3-319-62543-0>  **Open Access version available.**
- [24] Chaomei Chen and Min Song. 2017. *The Uncertainty of Science: Navigating Through the Unknown*. Springer International Publishing, Cham, 1–35. <https://doi.org/10.1007/978-3-319-62543-0>
- [25] Chenghui Chen and Lawrence Jun Zhang. 2017. An intercultural analysis of the use of hedging by Chinese and Anglophone academic English writers. *Applied Linguistics Review* 8, 1 (2017), 1–34. <https://doi.org/10.1515/applirev-2016-2009>  **Open Access version available.**
- [26] Kellia Chiu, Quinn Grundy, and Lisa Bero. 2017. 'Spin' in published biomedical literature: A methodological systematic review. *PLOS Biology* 15, 9 (09 2017), 1–16. <https://doi.org/10.1371/journal.pbio.2002173>
- [27] Soonja Choi. 2006. Acquisition of modality. *The expression of modality* (2006), 141–171.  **Open Access version available.**
- [28] Gudrun Clemen. 1997. The concept of hedging: Origins, approaches and definitions. *Hedging and discourse: Approaches to the analysis of a pragmatic phenomenon in academic texts* (1997), 235–248.  **Open Access version available.**
- [29] Andy Cockburn, Pierre Dragicevic, Lonni Besançon, and Carl Gutwin. 2020. Threats of a Replication Crisis in Empirical Computer Science. *Commun. ACM* 63, 8 (July 2020), 70–79. <https://doi.org/10.1145/3360311>
- [30] Michael Correll and Michael Gleicher. 2014. Error Bars Considered Harmful: Exploring Alternate Encodings for Mean and Error. *IEEE Transactions on Visualization and Computer Graphics* 20, 12 (2014), 2142–2151. <https://doi.org/10.1109/TVCG.2014.2346298>  **Open Access version available.**
- [31] Geoff Cumming. 2014. The new statistics: Why and how. *Psychological Science* 25, 1 (Jan. 2014), 7–29. <https://doi.org/10.1177/0956797613504966>  **Open Access version available.**
- [32] Trine Dahl. 2004. Some characteristics of argumentative abstracts. *Akademisk Prosa* 2, 1 (2004), 49–67.  **Open Access version available.**
- [33] R. Dal-Ré, M.V. Castell, and J. García-Puig. 2015. If the results of an article are noteworthy, read the entire article; do not rely on the abstract alone. *Revista Clínica Española (English Edition)* 215, 8 (2015), 454–457. <https://doi.org/10.1016/j.rceng.2015.08.004>

- [34] Anita de Waard and Henk Pander Maat. 2012. Epistemic Modality and Knowledge Attribution in Scientific Discourse: A Taxonomy of Types and Overview of Features. In *Proceedings of the Workshop on Detecting Structure in Scholarly Discourse*. Association for Computational Linguistics, Jeju Island, Korea, 47–55. <https://www.aclweb.org/anthology/W12-4306>  **Open Access version available.**
- [35] Jasmine M. DeJesus, Maureen A. Callanan, Graciela Solis, and Susan A. Gelman. 2019. Generic language in scientific communication. *Proceedings of the National Academy of Sciences* 116, 37 (2019), 18370–18377. <https://doi.org/10.1073/pnas.1817706116> arXiv:<https://www.pnas.org/content/116/37/18370.full.pdf>
- [36] Chrysanne Di Marco, Frederick W. Kroon, and Robert E. Mercer. 2006. *Using Hedges to Classify Citations in Scientific Articles*. Springer Netherlands, Dordrecht, 247–263. https://doi.org/10.1007/1-4020-4102-0_19
- [37] Leila Dobakhti. 2013. Expressing Certainty in Discussion Sections of Qualitative and Quantitative Research Articles. *Journal of Pan-Pacific Association of Applied Linguistics* 17, 1 (2013), 57–77.  **Open Access version available.**
- [38] Louise Doyle, Anne-Marie Brady, and Gobnait Byrne. 2016. An overview of mixed methods research – revisited. *Journal of Research in Nursing* 21, 8 (2016), 623–635. <https://doi.org/10.1177/1744987116674257>
- [39] Pierre Dragicevic. 2016. Fair Statistical Communication in HCI. In *Modern Statistical Methods for HCI*, Judy Robertson and Maurits Kaptein (Eds.). Springer International Publishing, Cham, Switzerland, Chapter 13, 291–330. https://doi.org/10.1007/978-3-319-26633-6_13  **Open Access version available.**
- [40] Farahman Farrokhi and Safoora Emami. 2008. Hedges and boosters in academic writing: native vs. non-native research articles in applied linguistics and engineering. *Journal of English Language Pedagogy and Practice* 1, 2 (2008), 62–98.  **Open Access version available.**
- [41] Sue Finch, Geoff Cumming, and Neil Thomason. 2001. Colloquium on Effect Sizes: the Roles of Editors, Textbook Authors, and the Publication Manual: Reporting of Statistical Inference in the Journal of Applied Psychology: Little Evidence of Reform. *Educational and Psychological Measurement* 61, 2 (2001), 181–210. <https://doi.org/10.1177/0013164401612001>
- [42] Bruce Fraser. 2010. *Pragmatic Competence: The Case of Hedging*. Brill, Leiden, The Netherlands, 15–34. https://doi.org/10.1163/9789004253247_003  **Open Access version available.**
- [43] Andrew Gelman. 2016. The problems with p-values are not just with p-values. *The American Statistician* 70, supplemental material to the ASA statement on p-values and statistical significance (2016), 10. http://www.stat.columbia.edu/~gelman/research/published/asa_pvalues.pdf  **Open Access version available.**
- [44] Gerd Gigerenzer. 1993. The Superego, the Ego, and the Id in Statistical Reasoning. In *A handbook for data analysis in the behavioural sciences: Methodological issues*. In G. Keren & C. Lewis (Ed.), Hillsdale, NJ: Erlbaum., 311–339. <https://doi.org/10.1093/acprof:oso/9780195153729.003.0013>
- [45] Gerd Gigerenzer. 2004. Mindless statistics. *The Journal of Socio-Economics* 33, 5 (2004), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>
- [46] Gerd Gigerenzer. 2018. Statistical Rituals: The Replication Delusion and How We Got There. *Advances in Methods and Practices in Psychological Science* (2018), 2515245918771329. <https://doi.org/10.1177%2F2515245918771329>  **Open Access version available.**
- [47] Paul Gilllaerts and Freek Van de Velde. 2010. Interactional metadiscourse in research article abstracts. *Journal of English for Academic Purposes* 9, 2 (2010), 128–139. <https://doi.org/10.1016/j.jeap.2010.02.004>
- [48] Roger Giner-Sorolla. 2012. Science or art? How aesthetic standards grease the way through the publication bottleneck but undermine science. *Perspectives on Psychological Science* 7, 6 (2012), 562–571.
- [49] Kilem L. Gwet. 2014. *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.
- [50] Mark Hallahan and Robert Rosenthal. 2000. Interpreting and Reporting Results. In *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, Howard Tinsley Steven Brown (Ed.). Academic Press, San Diego, 125–149. <https://doi.org/10.1016/B978-012691360-6/50006-9>
- [51] Elham Hariri and Abbas Zare-Ee. 2015. Variation in Hedging Across Psychology, Medicine and chemical Engineering Research Articles AND Across different Rhetorical Sections of the RAs. *International Journal of Applied Linguistics* 1 (12 2015), 39 – 46.  **Open Access version available.**
- [52] J. Helske, S. Helske, M. Cooper, A. Ynnerman, and L. Besançon. 2021. Can visualization alleviate dichotomous thinking? Effects of visual representations on the cliff effect. *IEEE Transactions on Visualization & Computer Graphics* 27, 8 (apr 2021), 3397–3409. <https://doi.org/10.1109/TVCG.2021.3073466>  **Open Access version available.**
- [53] Julia Himmelsbach, Stephanie Schwarz, Cornelia Gerdenitsch, Beatrix Wais-Zechmann, Jan Bobeth, and Manfred Tscheligi. 2019. Do We Care About Diversity in Human Computer Interaction: A Comprehensive Content Analysis on Diversity Dimensions in Research. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland UK) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–16. <https://doi.org/10.1145/3290605.3300720>  **Open Access version available.**
- [54] Victor Ho and Cissy Li. 2018. The use of metadiscourse and persuasion: An analysis of first year university students’ timed argumentative essays. *Journal of English for Academic Purposes* 33 (2018), 53 – 68. <https://doi.org/10.1016/j.jeap.2018.02.001>
- [55] Janet Holmes. 1988. Doubt and Certainty in ESL Textbooks. *Applied Linguistics* 9, 1 (03 1988), 21–44. <https://doi.org/10.1093/applin/9.1.21> arXiv:<http://oup.prod.sis.lan/applij/article-pdf/9/1/21/9741799/21.pdf>  **Open Access version available.**
- [56] Kelly Horn. 2001. The Consequences of Citing Hedged Statements in Scientific Research Articles: When scientists cite and paraphrase the conclusions of past research, they often change the hedges that describe the uncertainty of the conclusions, which in turn can change the uncertainty of past results. *BioScience* 51, 12 (12 2001), 1086–1093. [https://doi.org/10.1641/0006-3568\(2001\)051\[1086:TCHS2.0.CO;2](https://doi.org/10.1641/0006-3568(2001)051[1086:TCHS2.0.CO;2) arXiv:<http://oup.prod.sis.lan/bioscience/article-pdf/51/12/1086/26890902/51-12-1086.pdf>
- [57] Katarzyna Hryniuk. 2018. Expert-Like Use of Hedges and Boosters in Research Articles Written by Polish and English Native-Speaker Writers. *Research in Language* 16, 3 (2018), 263–280. <https://doi.org/10.2478/rela-2018-0013>  **Open Access version available.**
- [58] Jessica Hullman, Xiaoli Qiao, Michael Correll, Alex Kale, and Matthew Kay. 2019. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (2019), 903–913. <https://doi.org/10.1109/TVCG.2018.2864889>  **Open Access version available.**
- [59] Ken Hyland. 1996. Writing Without Conviction? Hedging in Science Research Articles. *Applied Linguistics* 17, 4 (12 1996), 433–454. <https://doi.org/10.1093/applin/17.4.433>  **Open Access version available.**
- [60] Ken Hyland. 1998. Boosting, hedging and the negotiation of academic knowledge. *Text-Interdisciplinary Journal for the Study of Discourse* 18, 3 (1998), 349–382. <https://doi.org/10.1515/text.1.1998.18.3.349>
- [61] Ken Hyland. 1998. Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics* 30, 4 (1998), 437 – 455. [https://doi.org/10.1016/S0378-2166\(98\)0009-5](https://doi.org/10.1016/S0378-2166(98)0009-5)
- [62] Ken Hyland. 2000. Hedges, Boosters and Lexical Invisibility: Noticing Modifiers in Academic Texts. *Language Awareness* 9, 4 (2000), 179–197. <https://doi.org/10.1080/09658410008667145>
- [63] Ken Hyland. 2004. *Disciplinary discourses, Michigan classics ed.: Social interactions in academic writing*. University of Michigan Press.
- [64] Ken Hyland and John Milton. 1997. Qualification and certainty in L1 and L2 students’ writing. *Journal of Second Language Writing* 6, 2 (1997), 183 – 205. [https://doi.org/10.1016/S1060-3743\(97\)90033-3](https://doi.org/10.1016/S1060-3743(97)90033-3)  **Open Access version available.**
- [65] John PA Ioannidis. 2018. Massive citations to misleading methods and research tools: Matthew effect, quotation error and citation copying. *European Journal of Epidemiology* 33, 11 (2018), 1021–1023. <https://doi.org/10.1007/s10654-018-0449-x>  **Open Access version available.**
- [66] Pierre-Antoine Jean, Sébastien Harispe, Sylvie Ranwez, Patrice Bellot, and Jacky Montmain. 2016. Uncertainty Detection in Natural Language: A Probabilistic Model. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics (Nîmes, France) (WIMS '16)*. ACM, New York, NY, USA, Article 10, 10 pages. <https://doi.org/10.1145/2912845.2912873>  **Open Access version available.**
- [67] T Jefferson, C Di Pietrantonj, M G Debalini, A Rivetti, and V Demicheli. 2009. Relation of study quality, concordance, take home message, funding, and impact in studies of influenza vaccines: systematic review. *BMJ* 338 (2009). <https://doi.org/10.1136/bmj.b354>  **Open Access version available.**
- [68] Samuel Jellison, Will Roberts, Aaron Bowers, Tyler Combs, Jason Beaman, Cole Wayant, and Matt Vassar. 2020. Evaluation of spin in abstracts of papers in psychiatry and psychology journals. *BMJ Evidence-Based Medicine* 25, 5 (2020), 178–181. <https://doi.org/10.1136/bmjebm-2019-111176> arXiv:<https://ebm.bmj.com/content/25/5/178.full.pdf>  **Open Access version available.**
- [69] Hannah Jergas and Christopher Baethge. 2015. Quotation accuracy in medical journal articles—a systematic review and meta-analysis. *PeerJ* 3 (2015), e1364. <https://doi.org/10.7717/peerj.1364>  **Open Access version available.**
- [70] Bixi Jin. 2017. Rhetorical differences in research article discussion sections of high- and low-impact articles in the field of chemical engineering. *IEEE Transactions on Professional Communication* 61, 1 (2017), 65–76. <https://doi.org/10.1109/TPC.2017.2747358>
- [71] Branden B. Johnson and Paul Slovic. 2015. Fearing or fearsome Ebola communication? Keeping the public in the dark about possible post-21-day symptoms and infectiousness could backfire. *Health, Risk & Society* 17, 5-6 (2015), 458–471. <https://doi.org/10.1080/13698575.2015.1113237>  **Open Access version available.**
- [72] Hüseyin Kafes. 2018. Stance In Academic Writing. *European Journal of Education Studies* 0, 0 (2018). <https://oapub.org/edu/index.php/ejes/article/>

- view/1423  **Open Access version available.**
- [73] Kris N Kirby and Daniel Gerlanc. 2013. BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior research methods* 45, 4 (2013), 905–927.
- [74] Nikolai Klitzing, Rink Hoekstra, and Jan-Willem Strijbos. 2019. Literature practices: Processes leading up to a citation. *Journal of Documentation* (2019).  **Open Access version available.**
- [75] Lisa Koeman. 2018. How many participants do researchers recruit? A look at 678 UX/HCI studies. Online. Last visited 06 January 2019. <https://lisakoeman.nl/blog/how-many-participants-do-researchers-recruit-a-look-at-678-ux-hci-studies>
- [76] Wellman Kondowe. 2014. Hedging and boosting as interactional metadiscourse in literature doctoral dissertation abstracts. *International Journal of Language Learning and Applied Linguistics World* 5, 3 (2014), 214–221. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.429.1732&rep=1&typepdf>  **Open Access version available.**
- [77] William Labov. 1984. Intensity. *Meaning, form, and use in context: Linguistic applications* 70 (1984).
- [78] George Lakoff. 1973. Hedges: A Study in Meaning Criteria and the Logic of Fuzzy Concepts. *Journal of Philosophical Logic* 2, 4 (1973), 458–508. <http://www.jstor.org/stable/30226076>  **Open Access version available.**
- [79] Bruno Latour and Steve Woolgar. 2013. *Laboratory life: The construction of scientific facts*. Princeton University Press.
- [80] Jeff A Lehmen, Rachel M Deering, Andrew K Simpson, Charles S Carrier, and Christopher M Bono. 2014. Inconsistencies between abstracts and manuscripts in published studies about lumbar spine surgery. *Spine* 39, 10 (2014), 841–845. <https://doi.org/10.1097/BRS.0000000000000290>
- [81] Ting Li and Sue Wharton. 2012. Metadiscourse repertoire of L1 Mandarin undergraduates writing in English: A cross-contextual, cross-disciplinary study. *Journal of English for Academic Purposes* 11, 4 (2012), 345 – 356. <https://doi.org/10.1016/j.jeap.2012.07.004>
- [82] Marc Light, Xin Ying Qiu, and Padmini Srinivasan. 2004. The language of bioscience: Facts, speculations, and statements in between. In *HLT-NAACL 2004 Workshop: Linking Biological Literature, Ontologies and Databases*. Association for Computational Linguistics, Boston, Massachusetts, USA, 17–24.  **Open Access version available.**
- [83] Yong Liu, Jorge Goncalves, Denzil Ferreira, Bei Xiao, Simo Hosio, and Vassilis Kostakos. 2014. CHI 1994-2013: Mapping Two Decades of Intellectual Progress through Co-Word Analysis. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) (CHI '14). Association for Computing Machinery, New York, NY, USA, 3553–3562. <https://doi.org/10.1145/2556288.2556969>  **Open Access version available.**
- [84] Chek Kim Loi and Jason Miin-Hwa Lim. 2019. Hedging in the discussion sections of English and Malay Educational Research Articles. *GEMA Online@ Journal of Language Studies* 19, 1 (2019). <https://doi.org/10.17576/gema-2019-1901-03>  **Open Access version available.**
- [85] Chek-Kim Loi, Jason Miin-Hwa Lim, and Sue Wharton. 2016. Expressing an evaluative stance in English and Malay research article conclusions: International publications versus local publications. *Journal of English for Academic Purposes* 21 (2016), 1 – 16. <https://doi.org/10.1016/j.jeap.2015.08.004>
- [86] Joe Marshall, Conor Linehan, Jocelyn Spence, and Stefan Rennick Egglestone. 2017. Throwaway Citation of Prior Work Creates Risk of Bad HCI Research. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (Denver, Colorado, USA) (CHI EA '17). ACM, New York, NY, USA, 827–836. <https://doi.org/10.1145/3027063.3052751>  **Open Access version available.**
- [87] Anna Mauranen. 1997. *Hedging in language revisers' hands*. Berlin and New York: Walter de Gruyter. 115–133 pages.
- [88] Elisabeth Mayweg-Paus and Regina Jucks. 2015. Evident or doubtful? How lexical hints in written information influence laypersons' understanding of influenza. *Psychology, Health & Medicine* 20, 8 (2015), 989–996. <https://doi.org/10.1080/13548506.2014.986139> PMID: 25482286.
- [89] Fatemeh Mirzapour and Mohammad Rasekh Mahand. 2012. Hedges and boosters in native and non-native library and information and computer science research articles. *3L: Language, Linguistics, Literature* 18, 2 (2012).  **Open Access version available.**
- [90] Congjun Mu, Lawrence Jun Zhang, John Ehrlich, and Huaqing Hong. 2015. The use of metadiscourse for knowledge construction in Chinese and English research articles. *Journal of English for Academic Purposes* 20 (2015), 135–148. <https://doi.org/10.1016/j.jeap.2015.09.003>  **Open Access version available.**
- [91] Greg Myers. 1989. The pragmatics of politeness in scientific articles. *Applied Linguistics* 10, 1 (03 1989), 1–35. <https://doi.org/10.1093/applin/10.1.1> arXiv:<http://oup.prod.sis.lan/applij/article-pdf/10/1/1/9740121/1.pdf>
- [92] David Nicholas, Peter Williams, Ian Rowlands, and Hamid R. Jamali. 2010. Researchers' e-journal use and information seeking behaviour. *Journal of Information Science* 36, 4 (2010), 494–516. <https://doi.org/10.1177/0165551510371883>
- [93] David Nicholas, Peter Williams, Ian Rowlands, and Hamid R. Jamali. 2010. Researchers' e-journal use and information seeking behaviour. *Journal of Information Science* 36, 4 (2010), 494–516. <https://doi.org/10.1177/0165551510371883>  **Open Access version available.**
- [94] Mihaela Onofrei Plămadă, Diana Trandabăt, and Daniela Gifu. 2019. Towards Identifying Author Confidence in Biomedical Articles. *Data* 4, 1 (2019). <https://doi.org/10.3390/data4010018>  **Open Access version available.**
- [95] Matthew Peacock. 2006. A cross-disciplinary comparison of boosting in research articles. *Corpora* 1, 1 (2006), 61–84. <https://doi.org/10.3366/cor.2006.1.1.61>  **Open Access version available.**
- [96] Keith F Punch. 2013. *Introduction to social research: Quantitative and qualitative approaches*. sage.
- [97] Randolph Quirk. 2010. *A comprehensive grammar of the English language*. Pearson Education India.
- [98] Ghalab Rabab'ah and Ronza Abu Rumman. 2015. Hedging in Political Discourse: Evidence from the Speeches of King Abdullah II of Jordan. *Prague Journal of English Studies* 4, 1 (2015), 157 – 185. <https://doi.org/10.1515/pjes-2015-0009>  **Open Access version available.**
- [99] Ole Bjørn Rekdal. 2014. Academic urban legends. *Social Studies of Science* 44, 4 (2014), 638–654. <https://doi.org/10.1177/0306312714535679>  **Open Access version available.**
- [100] Vassiliki Rizomilioti. 2006. *Exploring Epistemic Modality in Academic Discourse Using Corpora*. Springer US, Boston, MA, 53–71. https://doi.org/10.1007/978-0-387-28624-2_4
- [101] Françoise Salager-Meyer. 1994. Hedges and textual communicative function in medical English written discourse. *English for Specific Purposes* 13, 2 (1994), 149 – 170. [https://doi.org/10.1016/0889-4906\(94\)90013-2](https://doi.org/10.1016/0889-4906(94)90013-2)  **Open Access version available.**
- [102] Françoise Salager-Meyer. 1997. I think that perhaps you should: A study of hedges in written scientific discourse. *Functional approaches to written text: Classroom applications* 1 (1997), 127–143. https://www.tesol-france.org/uploaded_files/files/TESOL%20V2N2%20C8%20I%20think%20that%20Perhaps.pdf  **Open Access version available.**
- [103] Sara Salar and Behzad Ghonsooly. 2016. A comparative analysis of metadiscourse features in knowledge management research articles written in English and Persian. *International Journal of Research Studies in Language Learning* 5 (2016). <https://doi.org/10.5861/ijrsl.2015.1154>  **Open Access version available.**
- [104] Joanna EM Sale, Lynne H Lohfeld, and Kevin Brazil. 2002. Revisiting the quantitative-qualitative debate: Implications for mixed-methods research. *Quality and quantity* 36, 1 (2002), 43–53. <https://doi.org/10.1023/A:1014301607592>  **Open Access version available.**
- [105] Shana M. Shaw, Stephen M. Walls, Breana Sylvester Dacy, Joel R. Levin, and Daniel H. Robinson. 2010. A follow-up note on prescriptive statements in nonintervention research studies. *Journal of Educational Psychology* 102, 4 (2010), 982–988. <https://doi.org/10.1037/a0020924>  **Open Access version available.**
- [106] Elena Sheldon. 2018. Dialogic spaces of knowledge construction in research article Conclusion sections written by English L1, English L2 and Spanish L1 writers. *Ibérica: Revista de la Asociación Europea de Lenguas para Fines Específicos (AELFE)* 35 (2018), 13–40.  **Open Access version available.**
- [107] John Skelton. 1988. The care and maintenance of hedges. *ELT journal* 42, 1 (1988), 37–43. <https://doi.org/10.1093/elt/42.1.37>  **Open Access version available.**
- [108] Masahiro Takimoto. 2015. A corpus-based analysis of hedges and boosters in English academic articles. *Indonesian Journal of Applied Linguistics* 5, 1 (2015), 95–105. <https://doi.org/10.17509/ijal.v5i1.836>  **Open Access version available.**
- [109] Kees van Deemter. 2010. *Not Exactly: in Praise of Vagueness*. Oxford University Press.
- [110] Teppo Varttala. 2001. *Hedging in scientifically oriented discourse. Exploring variation according to discipline and intended audience*. Tampere University Press.  **Open Access version available.**
- [111] Irena Vassileva. 2001. Commitment and detachment in English and Bulgarian academic writing. *English for Specific Purposes* 20, 1 (2001), 83 – 102. [https://doi.org/10.1016/S0889-4906\(99\)0029-0](https://doi.org/10.1016/S0889-4906(99)0029-0)  **Open Access version available.**
- [112] Ignacio Vázquez Orta and Diana Giner. 2008. Beyond mood and modality: epistemic modality markers as hedges in research articles. A cross-disciplinary study. In *Revistas - Revista Alicantina de Estudios Ingleses*. Vol. 21. Universidad de Alicante. Departamento de Filología Inglesa. <https://doi.org/10.14198/raei.2008.21.10>  **Open Access version available.**
- [113] Lavi Wolf. 2015. It's probably certain. *Proceedings of the Israel Association of Theoretical Linguistics* 30 (2015).  **Open Access version available.**
- [114] Oktay Yagiz and Cuneys Demir. 2015. A comparative study of boosting in academic texts: A contrastive rhetoric. *International Journal of English Linguistics* 5, 4 (2015), 12. <https://doi.org/10.5539/ijel.v5n4p12>  **Open Access version available.**

- [115] Tatyana Yakhontova. 2002. Selling or telling? The issue of cultural variation in research genres. *Academic discourse* 10, 1 (2002), 216–232.  **Open Access version available.**
- [116] Amélie Yavchitz, Isabelle Boutron, Aida Bafeta, Ibrahim Marroun, Pierre Charles, Jean Mantz, and Philippe Ravaud. 2012. Misrepresentation of Randomized Controlled Trials in Press Releases and News Coverage: A Cohort Study. *PLOS Medicine* 9, 9 (09 2012), 1–11. <https://doi.org/10.1371/journal.pmed.1001308>  **Open Access version available.**
- [117] Chrysoula Zerva, Riza Batista-Navarro, Philip Day, and Sophia Ananiadou. 2017. Using uncertainty to link and rank evidence from biomedical literature for model curation. *Bioinformatics* 33, 23 (07 2017), 3784–3792. <https://doi.org/10.1093/bioinformatics/btx466>  **Open Access version available.**