Prescription and reality in advanced academic writing

Feng (Kevin) Jiang[1] Ken Hyland[12]

1. School of Foreign Language Education, Jilin University, China
2. School of Education and Lifelong Learning, University of East Anglia, UK

Abstract

The byzantine conventions of advanced academic writing in English are notoriously difficult for graduate students and junior scholars to gain control over. Global higher education and academic publishing have seen a massive expansion in recent years, so that the ability to demonstrate insider disciplinary competence and a grasp of scholarly persuasion has become essential for academic success. With courses in English for Research and Publication Purposes (ERPP) in their infancy and unknown in many countries, writers often turn to the advice of style manuals for guidance. Here they often find recommendations to follow general principles of clarity, brevity and objectivity and are given advice on specific features to achieve this. In this paper we examine three of the most commonly referred to features: demonstrative 'this', existential 'there' and first-person pronouns. With the aid of a diachronic corpus of articles from four disciplines over the past 50 years, we find that expert use of these features has increasingly departed from the advice in the guides. We describe how the use of these features has changed and stress what instructors in ERPP courses might teach.

Keywords: style guide; prescriptive advice; research writing; corpora

1. Introduction

The expansion of international publishing is largely seen as a positive development, both for academics wishing to share their research and for developing nations striving to become part of the "knowledge economy". As a result, the number of researchers around the world has expanded massively, with UNESCO (2017) reporting some 7.8 million full-time equivalent researchers in 2013, an increase of 21% since 2007. There has, therefore, been a corresponding increase in published research, with 33,100 active scholarly peer-reviewed English-language journals publishing some 3 million articles in 2018 (Johnson et al, 2018). This scenario, of course, broadens the opportunities for EAL researchers to participate in a global scholarly dialogue. But gaining control of the byzantine conventions of academic English can be a daunting challenge for writers. Novice researchers, whether native or non-native English speakers, step into unfamiliar territory when crafting academic texts at this level and both PhD students and junior academics, as well as their teachers and mentors, turn to style manuals and writing websites for advice.

In this paper we examine the usefulness of this advice by comparing some of the most common admonishments in style guides to what we found in a diachronic corpus of academic articles. In this way we hope to illuminate the reliability of the advice and, pedagogically, to draw attention to what instructors in English for Research and Publication Purposes (ERPP) courses might teach.

2. Challenges and guidance: Academic writing and style guide advice

Dissertations and research papers are sites where authors require control of sophisticated literacy skills to produce texts which are both comprehensible and plausible to exacting specialist readers. Here, technical issues of grammatical accuracy have less relevance than rhetorical understanding, disciplinary knowledge and an awareness of an authorial self. For most newcomers, and even many seasoned old-hands, however, the conventions of academic writing are an "alien form of literacy"

(Hyland, 2016, p. 246) where formality, abstraction, technicality and objectivity conspire to baffle authors (Bailey, 2015; Hyland, 2019). As a result, new writers often seek refuge in the authority of writing manuals or style guides, where they find advice like this:

> Pronouns confuse readers unless the referent for each pronoun is obvious; readers should not have to search previous text to determine the meaning of the term. Pronouns such as *this*, *that*, *these*, and *those* can be troublesome when they refer to something or someone in a previous sentence. Eliminate ambiguity by writing, for example, *this test*, *that trial*, *these participants*, and *those reports*.
>
> (APA Manual, 2010, p.68)

Cautions against 'unattended pronouns' are just one of the features writers are advised to use or avoid in academic texts, and surveys of style guides list numerous such taboo items (e.g. Bennett 2009; Walkova, 2020). The extent to which this advice represents the pundit's personal preference, subjective impression or critically informed understanding, however, is never certain and empirical studies of style guide advice is rare. Do the prescriptive rules contained in style guides actually reflect the ways expert writers use language in academic texts? The source of the advice is essential as writing guidelines can have a decisive influence on authors' choices (Hagge, 1997; Millar, Budgell, & Fuller, 2013). It is also an important question in view of what features instructors might teach in academic writing classes.

The style guides are helpful to some extent. They point writers to general principles of clarity, brevity and an emphasis on objectivity (Bennett, 2009). The first two principles suggest writers avoid vague and wordy expressions (Belmont & Sharkey, 2011; Mewburn, Firth, & Lehmann, 2019), often advising them to provide clear reference when using pronouns (as in the APA quote above) and to minimise the use of 'dummy it' subjects. Advice on the need for objectivity tends to be less assured than in the past and often depends on the discipline the guide is aimed at. For instance, McMillan and Weyers (2011) warn scientists that: "above all, academic writing is objective, using language techniques that generally maintain an impersonal tone" (p. 106), while Oliver (1996) qualifies this advice by explaining objectivity is common "especially in the natural sciences" (p.5).

Surveys of style guides by Huckin (1993) and Chang and Swales (1999) suggest that anaphoric pronouns (e.g. demonstrative 'this'), expletives (e.g. existential 'there') and first person pronouns (e.g. 'I' and 'we') are among the most frequently mentioned language features in style guides. Taking these three features to represent the goals of clarity, brevity and objectivity in writing manuals, we focus on the advice given on them and compare this with their use in research texts.

3. Demonstrative 'this', existential 'there' and first-person pronouns: Stylistic taboos

3.1 Demonstrative 'this'

For Biber et al. (1999, p. 349), demonstrative 'this' marks "immediate textual reference" as shown in these examples[1].

> (1) <u>The hair root resembles the duodenal crypt in that the mitotic activity is associated with a specific deeper region of the connective tissue.</u> **This** could be taken to imply the existence of tissue-specific types of mesenchymal factors ... (Bio, 1990)

> (2) <u>If the firm will sacrifice "profits" (no matter how measured) for anything else, whether prestige, good public or labor relations, for a quiet life, or liquidity, or what have you, then it is clearly not maximizing profits. And if it is not maximizing profits it must be maximizing "utility", which is simply a more elaborate way of saying that it does what it thinks best.</u> **This** can hardly be untrue, but it is also

[1] All examples are from our diachronic corpus discussed in section 4. Here, demonstrative *this* is bolded, its referent is underlined and attending nominals are italicised.

> not very helpful unless some content can be poured into the empty utility functions. (Soc, 1965)
>
> (3) the categories of information and commensurate assignment would look something like **this**: To fine tune sentence length articulations of textual information using text language… (AL, 1990)

In (1), 'this' refers back to the sentence "the hair root resembles the duodenal crypt in that the mitotic activity is associated with a specific deeper region of the connective tissue," while in (2) "this" points to the whole prior stretch of text of several sentences. In (3), on the other hand, 'this' refers forward to the following proposition. Demonstrative *this* therefore helps writers to link chunks of information and create cohesion in a text, although anaphoric (backward-pointing) reference occurs far more frequently than the cataphoric (forward-pointing) use in academic prose (Biber et al., 1999).

But while demonstrative 'this' can create cohesive ties, exactly what 'this' refers to may be unclear, as the above examples show. Because the use of such "vague" or "broad reference" (Kolln & Gray, 2010, p. 96) writers run the risk of being unclear or imprecise and as a result style guides often recommend its avoidance. For instance, Strunk and White (2018) advise that "the pronoun 'this', referring to the complete sense of a preceding sentence or clause, can't always carry the load and so may produce an imprecise statement" (p.90). Style guides also recommend writers to eliminate ambiguity by accompanying 'this' with "a summary word" (Swales & Feak, 2012) or "a missing noun" (Kolln & Gray, 2013).

Swales (2005) calls the structure "attended or unattended *this*", that is, whether to attend the demonstrative 'this' with a noun or noun phrase (as in "*this* **uncertainty** is because…") or to leave 'this' unattended (as in "this is because…") and regards it as an important rhetorical choice. Generally, style guides advise writers to attend 'this' with a noun, but give little information regarding how (un)attended 'this' can best be used, what the referential content can best encapsulate.

3.2 Existential 'there'

Existential 'there' affirms the existence or presence of something (or its non-existence or non-presence). It is used with an intransitive verb, typically a copular or linking one, most often in 'there be + indefinite noun group' structures such as the following[2]:

> (4) **there was** *a corresponding increase* in the incidence of failure to correct. (AL, 1990)
>
> (5) **There is** *a final point about goals* which the index constructor should beware. (Soc, 1990)
>
> (6) **there is** *a substantial maternal contribution of dCTCF.* (Bio, 2015)

In these examples, the indefinite noun group is referred to as the notional subject (Biber et al., 1999) in order to be distinguished from 'there', which is the grammatical subject. This pushes the clausal subjects towards the end of the sentence. In addition to this main function, Huckin and Pesante (1988) and Carter-Thomas and Rowley-Jolivet (2001) recognise two additional functions, derived from this basic one. The three functions are: to assert the (non)existence of entities; to mark enumeration; and to summarise given information.

Firstly, existential 'there' asserts or denies the existence of something, often to show an association between two assertions. We can see this in (7), where the author comments on the importance of a previously stated explanation and then announces a practical motivation for this interpretation.

> (7) Since our general aim is the improvement of communication skill, **there must be** *constant emphasis* on the real meaning of the Latin

[2] Existential *there* is bolded and notional subject is italicised

words and the relation of the real meaning to the English-derived words. Many students will find this both illuminating and pleasurable. (AL, 1990)

Here existential 'there' allows "the addressee to prepare for something that is about to be introduced" (Halliday & Matthiessen, 2014, p. 308). By filling the theme position with minimal information, the 'empty' subject 'there' helps to introduce the new message and guide readers to a preferred interpretation. So 'there' in (7), for example, points to the upcoming new information conveyed by the noun group "constant emphasis…", which carries the focus of the message, so bringing the new element to readers' awareness (Celce-Murcia & Larsen-Freeman, 1999).

A second function of existential 'there' is to summarise or reiterate points made earlier in the discourse (Carter-Thomas & Rowley-Jolivet, 2001). In Example (8), the author argues about the disagreement on the distribution of administrative power in British universities, by summarising what has been found in the experiments.

(8) When asked whether most British university departments would be better run by the method of circulating chairmanship than by a permanent head of department, 60 per cent of the natural scientists agreed (including agreement with reservations) that "a serious disadvantage of Redbrick universities is that all too often they are run by a professorial oligarchy". This indicates that **there is** *some discontent* about the way power is distributed. (Soc, 1990)

A third function is to enumerate (Huckin & Pesante, 1988; Celce-Murcia & Larsen-Freeman, 1999; Carter-Thomas & Rowley-Jolivet, 2001). Biber et al. (1999, p. 952), for instance, take this as "a springboard in developing the text when it is used to introduce a series of elements". We can see this prospective signalling of upcoming information in the following examples. In (9), 'there are several challenges' predicts an enumeration, thus introducing a sequence of difficulties rather than a single item.

This enumeration helps to organise the discourse and create a coherent message as it prepares readers for what follows, helping them better understand the upcoming elements:

> (9) **There are** *several challenges* involved in designing zero detection comparator for a micro-power converters. First, the performance of the comparator needs to be very good. If the delay of the comparator is high, the exact timing of the HS control cannot be met. Second, the power consumption of the comparator needs to be small, as it will add to the loss in the boost converter. Finally, the mismatch between the devices in the comparator can result in high offset, which can alter the zero detection.… (EE, 2015)

Existential ‘there’ thus contributes to the organisation of a message, although style guides often warn academic writers to steer well clear of the structure. For example, *The Longman Practical Stylist* (Baker, 2006) tells students to “keep your sentences awake by not putting them into those favourite stretchers of the passivists, *There is..*, *It is…*, and the like” (p.114). This prescription is repeated in other texts, such as Belmont and Sharkey (2011) and Hall and Birkerts (2007). For others, making ‘there’ behave like a grammatical subject by pushing the actual subject further into the sentence makes it “non-referential, dummy and empty” (Celce-Murcia & Larsen-Freeman, 1999, p. 444) because it merely “fills space in a sentence without contributing to its meaning” (Hall & Birkerts, 2007, p. 339). For these reasons writers are typically advised to avoid it.

3.3 First person pronouns

The use or avoidance of first person pronouns (‘I, we, our’, etc.) in academic writing has been a perennial topic of debate among manual writers (Bennett, 2009), sociologists (Gilbert & Mulkay, 1984) and applied linguists (e.g. Hyland, 2001). Generally, however, writers are counselled to use an impersonal style of writing to strengthen the objectivity of their claims. First person pronouns are the most obvious expression of a writer's presence in a text, and style guides generally admonish students to avoid them. This advice is typical:

> The total paper is considered to be the work of the writer. You don't have to say 'I think' or 'My opinion is' in the paper. (. . .) Traditional formal writing does not use *I* or *we* in the paper.
>
> (Spencer & Arbon, 1997, p. 26)

> Typically, in academic writing, one should avoid personalised writing and the use of words related to the first person.
>
> (Decker & Werner, 2016, p. 287)

Admonishments regarding the avoidance of self-mention inform traditional scientific writing and stem from positivist assumptions about rationality and detached reasoning. Reliable descriptions of what the natural and human worlds are actually like are only possible by removing the socially contingent interventions of the recording actor, and these unmediated representations of the world are believed to be the only guarantee of reliable knowledge. Science is perhaps less dogmatic about the reliable bases of 'truth' these days and recognises that reporting occurs in a pragmatic context which includes the theories held by the interpreting researcher (Hyland, 2004). Persuasion, however, still relies heavily on the eradication of the author in order to give prominence to experimental activities and the workings of the natural world. It allows the research to speak directly to the reader in an unmediated way, and many style guides and online academic writing sites continue to advocate the avoidance of self-mention in order to convey an appropriate tone:

> The #1 Writing No-No is to never use 1st or 2nd person. Why? In academic writing, it's important to avoid personal bias. Using "I" or "we" makes the essay about you and your experiences, instead of research and concrete details. (WriteCheck, 2020)

> There is generally a rule in academic writing that the first person should *not* be used. This makes your writing sound more objective and impersonal. (University of Hull, 2020)

It is now recognised, at least in applied linguistics research, that the strategic use of personal pronouns allows writers to emphasise their own contribution to the field and to seek agreement for it (Hyland, 2001). Academic writing advice therefore now seems to be more open to the use of first-person pronouns, for stylistic reasons as much as anything else. *The Manual of Scientific Style* (Rabinowitz & Vogel, 2009), for instance, encourages writers to employ "the active voice whenever possible" (p 433), as does the influential 'How to write and publish a scientific paper':

> I herewith ask all young scientists to renounce the false modesty of previous generations of scientists. Do not be afraid to name the agent of the action in a sentence, even when it is 'I' or 'we'. (Gastel & Day 2016: 202)

While other guides acknowledge that avoidance of self-mention is not always appropriate in the social sciences:

> "disciplines utiliz[ing] the first person in discussing their work, including psychology, philosophy, political theory, nursing and anthropology"
> (Fairbairn & Winch, 2011, p. 177).

This conflicting advice, however, leaves many writers confused about the extent they can reasonably intrude into their discourse to assert their personal involvement (Carciu & Maria, 2009; Hyland, 2001).

In summary, the style guides tend to offer advice which may seem arbitrary and informed more by stereotypical assumptions than actual hard evidence of usage. In order to understand these three features better, and to judge whether there has been any changes in the use of these devices in published writing in recent years, we seek to answer the following questions:

(1) What are the frequency and forms of the three linguistic features in academic writing?

(2) Have there been any changes in recent decades?

(3) To what extent does the use of these features vary across disciplinary fields?

3. Corpora and analysis

To empirically investigate the accuracy of these stylistic prescriptions, and whether they have ever reflected actual writing practices, we created three corpora, taking research articles from the same five journals in four disciplines spaced at 25-year intervals over 50 years: 1965, 1990 and 2015. The fact that journals come and go, that they undergo fragmentation and specialisation, and that they are replaced by new ones over time, places some constraints on diachronic research. However, we sought to select robust journals at the top of their respective fields and with a long history.

We selected journals from disciplines which offered a cross-section of academic practice, representing soft and hard sciences: applied linguistics, sociology, electrical engineering and biology. From each of these four fields we took six papers at random from each of the five journals which had achieved the top ranking according to their 5-year impact factor in 2015. Single and co-authored papers were chosen in equal numbers. Overall, the corpus comprises 30 articles from each discipline from each year, 360 papers of 2.2 million words (Table 1), showing a massive increase in the length of articles over the period:

Table 1 Corpora by discipline and word length

Discipline	1965	1990	2015	Change (%)
applied linguistics	110,832	145,712	237,452	**114.2**
biology	244,706	240,255	237,998	**-2.7**
elec engineering	92,062	124,631	235,681	**156.0**
sociology	149,788	205,238	262,203	**75.0**

Totals	**597,388**	**715,836**	**973,334**	**62.9**

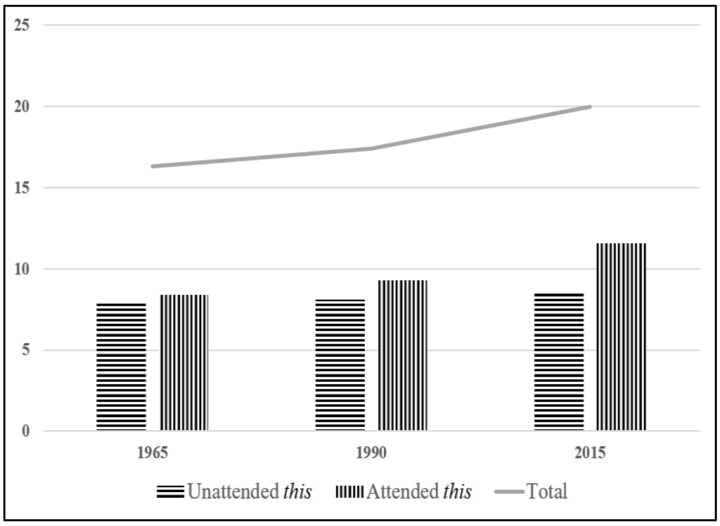
Firstly, to identify demonstrative ‘this’, the corpora were searched for ‘this’ in subject position, the “locus classicus” of this construction (Swales, 2005, p. 2). We then conducted a manual reading of concordance lines to identify each valid instance as either unattended or attended with a noun/noun phrase. Secondly, a search for ‘there’ was followed by a manual examination of every instance to eliminate all cases of the locative adverb there (as in “As can be seen there in text one…”). We considered whether the existential ‘there’ structures we identified were in a basic form (“there is *a silence*”) or whether more information was packaged using modification and adverbial expansion (“there are *widespread new principles of text-making composition*”). Thirdly, we searched for first person pronouns (‘I, my, me, we, our, us’) in the corpora, which indicate authorial intrusion. All cases were examined in context to ensure they were exclusive first person uses and to rule out inclusive uses which refer to both writers and readers.

All the searches were done using *AntConc* (Anthony, 2019) with coding by *MAXQDAplus* (2012). Both authors worked independently on a 10% sample of cases of each search item and achieved an inter-rater agreement of 96% on existential ‘there’, 98% on demonstrative *this*, 98% on first person pronouns after resolving disagreements.

4. Demonstrative ‘this’: attended clarity or unattended persuasion?

Overall, we identified 1,951 cases of demonstrative ‘this’ in the 2015 corpus, or 20 cases per 10,000 words, and 57.8% of them were attended. Figure 1 shows there has been a substantial increase in the use of the structure since 1965, with cases almost doubling over the period to 8.1 cases per article. This remains a significant increase of 22.7% (log likelihood = 27.78, $p < 0.001$) even when adjusted for the longer papers.

Figure 1 Change of (un)attended ‘this’ in the corpora over time (per 10,000 words)



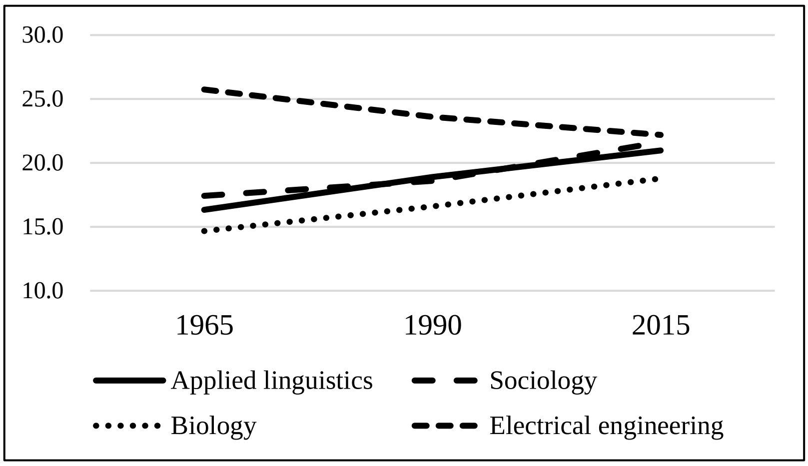
In other words, we observed a different trend than we might expect from the style guides with a substantial increase in the past 50 years. Presumably writers recognise its role in establishing cohesion and clarity in academic writing for an audience which has grown increasingly heterogeneous over the years, with institutional incentives for academics to reach wider popular and corporate readers. Today’s rhetorical context is one in which writers give greater attention to guiding readers through a text and helping them to see connections between ideas (Hyland & Jiang, 2019). Readers with less subject knowledge are likely to require more textual support to see connections and follow the writer’s argument. We can see from Figure 1, however, that style guide advice does correspond to the *proportional* difference in attended ‘this’, which has grown from 3.2% to 15.6% over the period. Writers are taking greater trouble to ensure their readers see clearly where their arguments are going, as we can see from (10).

> (10) We then calculated the rate of drift of these elements and the rates of proximal and distal growth based on the data extracted from the image analyses (Fig 5 and Fig 7). **This** *analysis* showed that drifting speeds range between 0.1409 and 0.5546. (Bio, 2015)

The attended noun helps readers to process the information and gain a clearer understanding of how the writer is organising the discourse.

Turning to disciplinary changes, the increase in the length of papers between 1965 and 2015 has meant an increase in the use of demonstrative ‘this’ to structure the longer texts and accommodate the reading needs of wider audiences. It is surprising, however, to see one discipline, electrical engineering, actually using slightly fewer cases, although it is still the most frequent user. (Fig 2).

Figure 2 Changes in Demonstrative ‘this’ by discipline (per 10,000 words)



As we saw in Table 1, engineering saw the greatest increase in the length of papers in our sample, rising 156% to 7,856 words over 50 years, so we might expect a corresponding rise in devices to help create cohesion. It seems, however, that writers are finding other ways to explicitly structure their texts, especially the formulaic phrases used to link numerical data which are familiar in the quantitative sciences (Kuteeva & McGrath, 2015). Commenting on research writing in engineering, for instance, McNown (1996) stresses the need to “use limited space effectively” and “pack the material densely”, but advises writers not to compress texts using complex structures (p.427). Certainly, demonstrative ‘this’, whether attended or unattended, provides engineers with a succinct way to link up texts compared with other cohesive devices, such as substitution and ellipsis.

We also found that writers in the hard sciences are more likely to use unattended ‘this’ for anaphoric reference than those in the humanities and social sciences. Scientists and engineers feel more confident about readers recovering a referent as they build cohesion far more through reference to specialised craft skills and the "tacit knowledge from their daily work with processes and instruments" (Myers, 1991, p. 6). As we can see in this example, although demonstrative 'this' is not defined by an attending noun, this knowledge can be retrieved by disciplinary familiarity with mechanical procedures (11):

> (11) The first just requires to verify whether (31) holds for some l $\in$ {1, . . . , ns}. The second sub-condition consists of checking the existence of j $\in$ J (k), k $\in$ K and p $\in$ Rq satisfying (32). **This** can be done by determining the values of p that satisfy the first constraint in (32) … (EE, 2015)

Most interestingly, Table 2 shows there has been an increase in the proportional use of unattended 'this' in the two applied disciplines (applied linguistics and electrical engineering) and a decrease in the two pure disciplines (sociology and biology).

Table 2 Changes in the proportion of attended and unattended 'this' by discipline (%)

	applied linguistics			sociology		
	1965	**1990**	**2015**	**1965**	**1990**	**2015**
attended *this*	78.5	68.7	62.9	48.7	52.8	58.2
unattended *this*	21.5	31.2	37.1	51.3	47.2	41.8
	biology			electrical engineering		
	1965	**1990**	**2015**	**1965**	**1990**	**2015**
attended *this*	55.2	56.1	57.3	56.5	53.2	49.7
unattended *this*	44.8	43.9	42.7	43.5	46.8	50.3

Unattended 'this', as Swales (2005) and Jiang and Wang (2018) show, encapsulates prior discourse in a word, turning it into a generalised proposition and indicating how the writer wants the information to be understood. In applied fields in particular, such a generalization may help align the propositional information with the epistemological emphasis of the field, promoting a utilitarian application of knowledge. Our diachronic data testify to a strengthened trend in those disciplines to rely on unattended ‘this’ both to establish textual coherence and the smooth flow of information and, equally importantly, to promote coherence by representing information in a way that readers can clearly understand.

To understand this structure in more detail, we also identified the content referred to by ‘this’ in the immediately preceding discourse and classified it in terms of its length and complexity. We used four categories: simple noun phrases (12), complex noun phrases (13), clause/sentence (14) and extended discourse (15).

> (12) The CEF or CRR is a way of expressing the per-unit reduction of load achieved by operating the DSM alternative in terms of deferred generation. **This** *factor* is used in the evaluation process to calculate dollars per kilowatt-year. (EE, 1990)

> (13) It evolves with the continuing formation of the channel spreading from the source to the drain. With a Schottky junction **this** *process* is suppressed. (EE, 2015)

> (14) Indeed, CRE-based cell tracing of Hes1+ terminal duct cells/CACs in adult mice failed to show any islet progenitor capacity while these cells seemed to contribute to the ductal tree. One explanation for **this** *difference* would be that while zebrafish CACs … (Bio, 2015)

> (15) The second innovation is a task which is not an enviable one for the Alliance but provides a fundamental difficulty to its leaders: the continuous evaluation, not only of fiscal and monetary measures and investment schemes, but also of complex social and political programs. For the former, international and professional agencies have now acquired experience and techniques: for example, in the appraisal of the effectiveness of inflationary or deflationary policy measures, of controls or the abolition of controls of foreign exchange, or of public and private

investments. Their evaluation is relatively easy because their results can be estimated fairly accurately in quantitative terms. But not so with respect to social reforms: their evaluation must be made partly in qualitative terms where value judgments play an influential role. *This* is a great challenge because the redistribution of social, economic, and political privileges formerly ... (Soc, 1965)

Table 3 shows that clause/sentence-level texts are the most common antecedents referred to by demonstrative 'this' across the years, despite a slight fall in the proportion. Although this result aligns with Gray (2010), our diachronic data show that complex noun phrases and extended discourse have increased their proportion of all referents, reaching 34.3% and 10.5% in 2015 respectively. This implies that writers are now inclined to include more information with both attended and unattended 'this', carrying this information into the ongoing discourse.

Table 3 Changes in the categories of referent antecedents across time (%)

	1965	**1990**	**2015**
simple noun phrases	18.6	17.2	15.6
complex noun phrases	33.5	34.1	34.3
clause/sentences	41.8	40.5	39.6
extended discourse	6.1	8.2	10.5

Table 4 shows that simple and complex noun phrases are both preferred and have gained an increased proportional use in the hard disciplines.

Table 4 Changes in the proportion of referent antecedents across disciplines (%)

	applied linguistics			sociology		
	1965	**1990**	**2015**	**1965**	**1990**	**2015**
simple noun phrases	32.7	30.5	28.6	31.9	29.3	26.8
complex noun phrases	33.3	29.8	28.3	34.5	31.7	29.1
clause/sentences	29.1	35.5	39.3	27.8	34.0	39.9
extended discourse	4.9	4.2	3.8	5.8	5.0	4.2
	biology			electrical engineering		
	1965	**1990**	**2015**	**1965**	**1990**	**2015**
simple noun phrases	29.4	29.8	28.6	33.2	33.1	31.3
complex noun phrases	34.0	35.2	39.0	33.8	36.7	40.2
clause/sentences	33.5	33.0	30.2	30.1	28.4	26.6
extended discourse	3.1	2.0	2.2	2.9	1.8	1.9

Phrasal structures figure prominently in scientific writing and have increased over time (Biber and Gray, 2016). Certainly, in our corpus, writers in the hard sciences mainly use phrases to convey information and then weave them into the ongoing argument using demonstrative ‘this’. Example (16) is typical in this respect.

> (16) Values of q within s-1 led to <u>a significant increase in time-averaged information</u> (Table 1 and Fig 4b; Friedman ANOVA, global effect on all considered q values: $p < 0.001$). **This** *effect* was robust early after the feedback.… (Bio, 2015)

By contrast, soft knowledge fields tend to use ‘this’ to refer to, and comment on, information expressed at a clause/sentence level as in (17).

> (17) <u>While not technically a control, we also find that ethnic identity mobilizes respondents independent of repression experiences (Model 2) and even for those who have not experienced repression</u>

personally (Model 3). **This** *finding* is in line with traditional movement theory… (Soc, 2015)

5. Existential 'there': structural verbosity or cogent interaction?

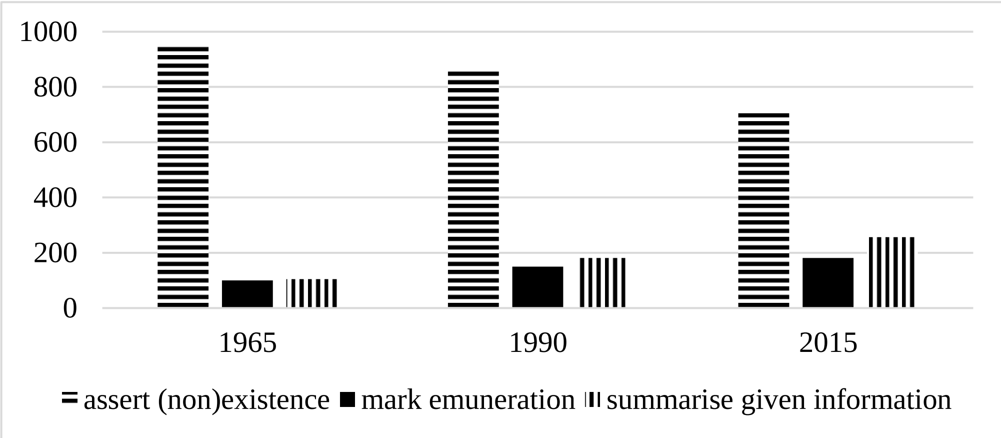
We identified 1143 cases of existential 'there' in the 2015 corpus, averaging 9.5 cases per article and 11.7 cases per 10,000 words. However, there has been a 39.3% drop from 19.4 cases per 10,000 words in 1965. But while there are fewer cases, the structure is still widely used in research writing and some functions have actually increased. This is because existential 'there' offers writers a useful means of introducing a new topic and strengthening the coherence of their text. In the examples below, we see writers evaluating a theoretical hypothesis (18) and smoothly transitioning between negative and positive features of corpora before addressing further arguments (19).

> (18) Might not such selection eventually result in the production of a protective and nurturing compartment—a cell (note, there is no protocell stage in this model)? **There are** *many problems* with this hypothesis. … (Bio, 2015)
>
> (19) Although **there are** *drawbacks*, **there are** *at least three main advantages* of using corpora in comparison with direct observations.
>
> (AL, 2015)

We should note that the notional subjects in the above examples, 'problems', 'drawback' and 'advantage', project judgements about something, indicating that the so-called 'empty' subject of 'there' actually offers a powerful mechanism for expressing authorial stance. Existential 'there' "relieves the writer of any responsibility for the viewpoint", presenting 'problems', 'drawback' and 'advantage' as "a given existential statement" (Davies, 1988, p. 197). In addition, in the case of 'there are at least three main advantages' in (19), 'advantage' constructs a frame for readers' understanding of the upcoming discussion, guiding their view to the author's positive evaluation of using corpora.

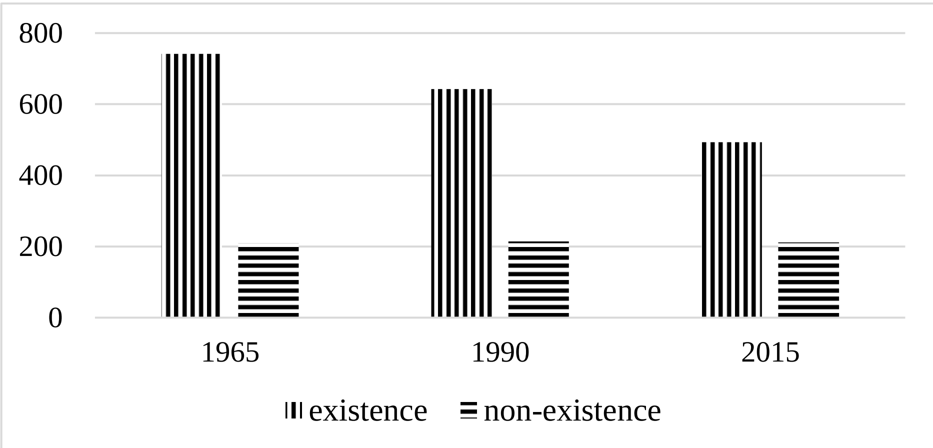
In terms of the rhetorical functions, Figure 3 shows that existential 'there' is increasingly used to summarise information and mark enumeration while asserting the (non) existence of entities continues to be the dominant function across the years. The increase of article length and the corresponding expansion of information packed in each article requires writers to provide additional textual support to assist readers' in processing texts. Existential 'there' creates a coherent flow moving between given and new information.

Figure 3 Frequency of the rhetorical functions of existential 'there' across years



As we might expect, the construction largely affirms the existence of something, with 74.3% of cases throughout the period. Figure 4 presents the changes in the proportion of existence and non-existence affirmed by existential 'there'.

Figure 4 Frequency of affirming by existential 'there' across years



English academic prose tends to privilege negative assertions. This is because writers often either refute a hypothesis (20) or problematise previous studies to move closer to their own position (21):

> (20) **There is** *no epistemological reason* not to integrate this rich, uninitiated data in social analysis. (Soc, 2015)
>
> (21) Numerous attempts had been made to explain its nature, but **there was** *no general agreement* among these interpretations. (Bio, 1965)

It is also worth mentioning that there has been a slight rise in the use of existential ‘there’ among hard science writers as they have increasingly sought to strengthen the presentation of their results (Hyland & Jiang, 2019). Conversely, the more discursive soft fields have reduced their use, perhaps with a growing awareness of the potential kickback by readers against making strong claims. However, both sociologists and, particularly, applied linguists remain the heaviest users of the structure. Table 5 summarises the changing frequency of existential ‘there’ across disciplines.

Table 5 Changing frequency of existential ‘there’ by discipline (per 10,000 words)

applied linguistics			**sociology**		
1965	1990	2015	1965	1990	2015
45.1	31.6	17.2	25.7	18.2	12.2
electronic engineering			**biology**		
1965	1990	2015	1965	1990	2015
7.9	12.0	8.5	8.1	8.5	9.0

In order to explore how writers use existential ‘there’ to achieve disciplinary persuasion, we are also interested in its discoursal expansion in constructing arguments. This can be achieved either through pre-modification (22), post-modification (23), or adverbials (24). Alternatively, modification and adverbials can be combined as in (25).

(22) **there have been** *no industrially acceptable methods*. (EE, 2015)

(23) **there is** *evidence* that stubborn group-related dispositions inform sonic styles and sonic evaluation… (Soc, 2015)

(24) In particular **there may be** *scope* for future studies which deliberately aim to collect data from classrooms … (AL, 2015)

(25) **there was** *a strong curvilinear relationship between the foliar chlorophyll content and SPAD values*. (Bio, 2015)

Our results show that authors in the hard domains tend to expand the head nouns in their work more than those in the soft fields. In so doing, they also package more information into each structure (3.2 expansions per head noun in hard disciplines vs 2.1 in soft fields on average across the years).

Table 6 Expansion of head noun in existential 'there' (% of expanded nouns)

	AL			Soc			EE			Bio		
	1965	1990	2015	1965	1990	2015	1965	1990	2015	1965	1990	2015
Discoursal expansion	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
modifications	**76.5**	**75.8**	**74.8**	**75.5**	**76.1**	**77.4**	**68.9**	**68.3**	**68.2**	**71.9**	**71.7**	**71.5**
(1) post-modification	52.4	51.8	51.3	49.5	50.4	51.7	23.8	23.4	23.3	20.7	21.1	22.1
phrase	27.0	25.8	24.5	26.6	27.2	27.8	5.8	5.5	5.3	10.8	11.2	12.3
clause	25.4	26.0	26.8	22.9	23.2	23.9	18.0	17.9	18.0	9.9	9.9	9.8
(2) pre-modification	24.1	24.0	23.5	26.0	25.7	25.7	45.1	44.9	44.9	51.2	50.6	49.4
adverbials	**23.5**	**24.2**	**25.2**	**24.5**	**23.9**	**22.6**	**31.1**	**31.7**	**31.8**	**28.1**	**28.3**	**28.5**
(1) logic	20.3	20.8	21.3	22.5	21.6	19.9	17.1	18.1	19.2	20.2	21.2	21.7
(2) time/place	3.2	3.4	3.9	2.0	2.3	2.7	14.0	13.6	12.6	7.9	7.1	6.8

We can see from Table 6 that typically, the discoursal expansion in hard discipline texts is largely achieved through pre-modification and adverbials, as in these examples:

> (26) **there is** *a non-zero steady-state prediction error*. (EE, 2015)
>
> (27) **there is** *a logical path from surface protoplasm to protocells via selection for the ability to colonize distant mineral surfaces*. (EE, 2015)

Writers in the soft disciplines, in contrast, tend to expand head nouns through post-modification.

> (28) **there is** *an increasing tendency* to generate new, and very different, mutant forms of a social contagion with each new transmission. (Soc, 2015)
>
> (29) **there was** *no reference* to the student's research process, and barely *any reference* to the student's work, at all. (AL, 2015)

In comparing these examples, we can see that writers in hard domains typically expand head nouns to add precision to their descriptions, while those in soft fields add discoursal elaboration to ensure a more diverse readership will find their arguments coherent.

6. First person pronouns: objectivity or presence?

The use of first person pronouns has increased dramatically in academic writing over the last 50 years, rising from 223.1 cases per 10,000 words in 1965 to 284.0 cases in 2015. This growth of 27.3% suggests that the style guides which advise against personal forms of language use are wildly out of date. Table 7 shows that frequencies have increased a massive 163% in biology, 63% in electrical engineering and 38% in sociology, although they have declined 27% in applied linguistics. The huge rises in the hard sciences are perhaps the most surprising as science writers are conventionally seen to represent meanings in an objective way, subordinating the authority of the individual to the authority of the text (Hyland, 2001, 2004a). These trends undermine those assumptions.

Table 7 Changes in first person pronouns by discipline (per 10,000 words)

	applied linguistics			sociology			electrical engineering			biology		
	1965	1990	2015	1965	1990	2015	1965	1990	2015	1965	1990	2015
we	42.8	41.7	29.5	33.6	31.9	40.2	47.1	66.7	71.0	10.1	15.1	37.9
I	22.1	18.2	29.5	12.8	13.3	28.7	0.1	0.2	0.5	2.2	1.6	1.8
our	16.2	19.0	8.9	12.8	14.4	12.9	3.6	4.2	6.9	5.8	6.0	10.1
us	5.7	6.0	3.2	3.9	3.3	4.0	1.8	4.5	7.2	1.2	1.2	1.8
Total	**86.8**	**83.9**	**64.3**	**64.0**	**64.9**	**81.9**	**52.8**	**65.1**	**85.7**	**19.2**	**19.8**	**51.7**

This increase, however, is confined to plural forms, which allow authors to create more distance between themselves and their reporting than first person singular, and so temper a more invasive stance. Partly, of course, it reflects the growing trend, and career pressures, towards multiple authorship over the last 70 years with the average number of authors per paper in Scopus growing from 3.2 in 1996 to 4.4 in 2015 (Economist, 2016) and almost two-thirds of global articles having authors from multiple institutions (Hyland, 2015).

It is nevertheless a significant trend towards greater authorial presence in scientific texts. One reason for this may be the need for scientists to respond to the imperative of "impact" as a measure of their research contributions in annual performance reviews and career assessments. Taking their work beyond the confines of a specialised group of insiders involves interpersonal adjustments, making their research less impersonal and more accessible to audiences in the commercial world. A more likely explanation, however, is that greater visibility is now essential: creating a presence as an acting, decision-making individual is a way of ensuring that their interpretations and claims do not go unnoticed by university human resource panels where applications for jobs, tenure and promotion are judged. A more personal voice garners more visibility, more citations and more professional credit. Something of this claim for recognition can be seen in these examples:

(30) **We** highlight where data for certain fragile Y predictions are lacking and propose research avenues that would further elucidate particular dynamics of Y-chromosome loss. (Bio, 2015)

(31) In this work **we** develop a theory of system approximation for timed systems by quantifying the timing differences between corresponding system events. (EE, 2015)

Here we see personal reference makes a clear claim for credit and an explicit indication of the perspective from which a statement should be interpreted, enabling writers to emphasise their own contribution and to seek agreement for it.

Sociology has followed a similar path with authors increasing their presence in their texts through first person pronouns, although this is perhaps more understandable than in the sciences. Here research is generally more explicitly interpretative and less abstract, with less 'exact' data collection procedures and less control of variables. Readers expect that authors will not write with positivist detachment but will craft a convincing argument from a personal point of view:

(32) **we** argue that when errors occur individual humans may attempt to correct them, essentially adding in words or actions that seem appropriate in the available context. Thus **our** overall theory applies broadly to transferable, endogenously duplicable... (Soc, 2015)

(33) **I** attribute this to the contingent nature of competition ... (Soc, 2015)

Overall, therefore, avoidance of the first person to convey impersonality now seems less rigidly adhered to than in the past, reflecting the ambivalence about its use in the style guides. The patterns in applied linguistics, however, are more curious, with a substantial fall in personal forms over the last 50 years. It possibly reflects the growing maturity of the discipline and an increase in empirically-oriented studies, as opposed to more personal accounts of teaching practices in earlier times. The discipline has increasingly taken on the trapping of the social sciences with a corresponding moderation of self-mention and increased reliance on replicability, falsifiability and more rigorous independent support of claims. An alternative explanation may be the influence on the corpus of growing numbers of second language writers, most notably from China and the Middle East, who have been schooled in the virtues of eliminating explicit agency from academic writing (e.g. Hyland, 2015). More likely, perhaps, is that this development simply represents a heightened awareness, and growing self-consciousness, among rhetorically sensitive applied linguists keen to replace an ego-centric voice with a more collectivist one.

These explanations, however, are rather fanciful. Their weakness is emphasised, moreover, by the fact that both applied linguistics and sociology showed an increase in the use of singular first person over the period, with it doubling in the former since 1990 and increasing threefold in the latter. Quite clearly, the ability of authors to position themselves as reflective agents with important ideas to contribute (34) or insights to convey (35) remains very important in these fields:

> (34) **I** broaden their observations and argue that highly discretionary programs are themselves a common feature of American governance. **I** furthermore point to how assumptions about the sexual division of labor ordered a federal and locally applied discretionary program. Framed through a feminist-Weberian lens, **I** argue that the American state's locally… (Soc, 2015)
>
> (35) **I** am distinctly aware that **my** approach stems from the SLA periphery. **My** objective in advancing such a perspective, however, is not to discount the earlier SLA work on language anxiety. Rather, **I** seek to extend the borders of future research on language anxiety in SLA ... (AL, 2015)

While 'I' is increasing in the soft knowledge fields, we can see from Table 7 that *we* dominates the frequencies in all disciplines, comprising 58% of all forms in 2015. This has nearly doubled in electrical engineering and increased by almost four times in biology.

7. Final observations and conclusions

The globalisation of academia has created a highly competitive and assessment-oriented research culture in which individuals are judged by the length of their publication lists. Publication has come to equal 'productivity' and is used as a crude measure of worth, with institutions conferring posts, promotions, tenure and sometimes bonus payments on those who are successful. So not only has the need to publish never been greater, but this has to be done in an arcane form of English which requires both native and non-native English writers to acquire special literacy skills. With *English for Research and Publication Purposes* courses in their infancy, many writers turn to style guides and publication manuals for advice. What they often find in them are admonishments to observe principles of clarity, brevity and objectivity in their writing, with advice to avoid demonstrative 'this', existential 'there' and first person pronouns. Our detailed examination of these features in a diachronic corpus of research articles suggests that this advice does not accurately reflect current academic writing practices.

We have observed remarkable increases in the use of demonstrative 'this' and first person pronouns in research writing over the past 50 years, although existential 'there' has fallen. The guides seem to have it partly right. While demonstrative 'this' has seen a significant rise, in defiance of style guide admonishments, much of this has been in cases with an attended noun/noun phrase, which reflects the advice provided in the guides to ensure readers can recover the intended referent. The significant drop in the use of existential 'there' largely corresponds to advice given to academic writers, although it remains an important means of introducing topics and strengthening textual coherence. It has come increasingly, moreover, to mark enumeration and summarise information, especially in the hard science fields, to ensure readers get the writer's take-home message. Style guides have been less successful in accurately reflecting the use of first person in professional research writing. The sciences in particular are now significantly more likely to express arguments in the active voice using inclusive *we* to claim agency for their decisions and credit for their work.

These authorial practices are a rhetorical response to the dramatic changes in academic life in recent years which demand more publication, greater collaboration and expanded outreach to new audiences. The commercial repositioning of disciplinary knowledge means that research is often read by a more diverse audience who can make practical use of it, and authors have responded by making themselves more accessible and their work more reader friendly.

Our research also has implications for graduate writing classes and the burgeoning field of English for Research and Publication Purposes. At these levels of advanced academic writing, novices need advice based on empirical research of what occurs rather than outdated impressions of conventions. While style guides and publication manuals can be useful, it is more important for students to treat their advice with scepticism and to learn from the rhetorical practices embodied in the pages of the journals in their discipline. Teachers have an important role to play here, guiding students away from tame compliance with discrete stylistic rules and towards an understanding of academic discourse as a coherent means of achieving practical ends. The decisions writers make are not effective because they comply with extant rules, but because they meet readers' expectations of what is familiar, plausible and persuasive. The kind of corpus studies we have reported here can not only assist novice research writers but should inform the work of language teachers. Such an approach encourages us to be critical in our work and to prioritise functions and understand what the writer is trying to do, rather than adhere slavishly to stylistic prescriptions of forms.

References

Anthony, L. (2019). AntConc (Version 3.5.8). Tokyo: Waseda University. Retrieved from http://www.antlab.sci.waseda.ac.jp/

APA. (2010). *Publication manual of the American Psychological Association* (6th ed.). Washington: American Psychological Association.

Bailey, S. (2015). *The essentials of academic writing for international students*. London: Routledge.

Baker, S. W. (2006). *The Longman practical stylist*. New York: Pearson.

Belmont, W., & Sharkey, M. (2011). *The easy writer: Formal writing for academic purposes* (3rd ed.). Frenchs Forest: Pearson.

Bennett, K. (2009). English academic style manuals: A survey. *Journal of English for Academic Purposes*, *8*(1), 43–54.

Biber, D., & Gray, B. (2016). *Grammatical complexity in academic English: Linguistic change in writing*. Cambridge: Cambridge University Press.

Biber, D., Johansson, S., Leech, G. Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The Grammar book: An ESL/EFL teacher's course* (2nd ed.). Boston: Heinle & Heinle.

Chang, Y. Y., & Swales, J. M. (1999). Informal elements in English academic writing: threats or opportunities for advanced non-native speakers. In C. N. Candlin & K. Hyland (Eds.), *Writing: Texts, Processes and Practices* (pp. 145–167). Harlow: Longman.

Davies, F. (1988). Reading between the lines: thematic choice as a device for presenting writer viewpoint in academic discourse. *The ESPecialist*, *9*(1/2), 173–200.

Decker, C., & Werner, R. (2016). *Academic research and writing: A concise introduction*. Frankfurt am Main: iCADEMICUS.

Fairbairn, G., & Winch, C. (2011). *Reading, writing, and reasoning: A guide for students* (3rd ed.). Maidenhead: Open University Press.

Carciu, & Maria, O. (2009). An intercultural study of first-person plural references in biomedical writing. *Ibérica*, *18*, 71–92.

Carter-Thomas, S., & Rowley-Jolivet, E. (2001). Syntactic differences in oral and written scientific discourse: the role of information structure. *ASp*. (31-33), 19–37.

Economist, The. (2016). Why Research Papers Have so Many Authors. *The Economist*, https://www.economist.com/science-and-technology/2016/11/24/whyresearch-papers-have-so-many-authors.

Gastel, B. & Day, R. (2016). *How to write and publish a scientific paper* (8th ed.). Santa Barbara: Greenwood.

Gilbert, G. N., & Mulkay, M. (1984). *Opening Pandora's box: A sociological analysis of scientists' discourse*. Cambridge: Cambridge University Press.

Gray, B. (2010). On the use of demonstrative pronouns and determiners as cohesive devices: a focus on sentence-initial *this*/*these* in academic prose. *Journal of English for Academic Purposes*, 9 (3), 167-183.

Hagge, J. (1997). Disciplinary style manuals as reliable guides to scientific discourse norms. *Technical Communication*, *44*(2), 129–141.

Hall, D. & Birkerts, S. (2007). *Writing well* (9th ed). New York: Longman.

Halliday, M. A. K., & Matthiessen, Christian M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). London: Routledge.

Huckin, T. N. (1993). Stylistic prescriptivism vs. expert practice. *Technostyle*, *11*(2), 1–17.

Huckin, T. N., & Pesante, L. H. (1988). Existential *there*. *Written communication*, *5*(3), 368–391.

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles. *English for Specific Purposes*, *20*(3), 207–226.

Hyland, K. (2004). *Disciplinary discourses: Social interactions in academic writing*. Ann Arbor: University of Michigan Press.

Hyland, K. (2015). *Academic publishing: Issues and challenges in the construction of knowledge*. Oxford: Oxford University Press.

Hyland, K. (2016). Writing with attitude: conveying a stance in academic texts. In E. Hinkel (Ed.), *Teaching English grammar to speakers of other languages* (pp. 246–265). London: Routledge.

Hyland, K. (2019). *Second language writing* (2nd edition). Cambridge: Cambridge University Press.

Hyland, K., & Jiang, F. K. (2019). *Academic discourse and global publishing: Disciplinary persuasion in changing times*. London: Routledge.

Jiang, F. K., & Wang, F. (2018). 'This is because … ': authorial practice of (un)attending *this* in academic prose across disciplines. *Australian Journal of Linguistics*, *38*(2), 162–182.

Johnson, R. Watkinson, A. & Mabe, M. (2018). *The STM report: An overview of scientific and scholarly publishing* (5th ed.). Hague: International Association of Scientific, Technical and Medical Publishers

Kolln, M., & Gray, L. S. (2013). *Rhetorical grammar: Grammatical choices, rhetorical effects* (7th ed.). Boston: Pearson.

Kuteeva, M & McGrath, L. (2015). The theoretical research article as a reflection of disciplinary practices: the case of pure mathematics. *Applied Linguistics*, 36, 215-235.

McMillan, K., & Weyers, J. D. B. (2011). *How to write essays & assignments*. Harlow: Pearson.

McNown, J. (1996). Guest editorial. *Journal of Hydraulic Engineering, 122*(8), 427.

Mewburn, I., Firth, K., & Lehmann, S. (2019). *How to fix your academic writing trouble: A practical guide*. Maidenhead: Open University Press.

Millán, E. L. (2010). "Extending this claim, we propose …": The writer's presence in research articles from different disciplines. *Ibérica*, *20*, 35–56.

Millar, N., Budgell, B., & Fuller, K. (2013). 'Use the active voice whenever possible': the impact of style guidelines in medical journals. *Applied linguistics*, *34*(4), 393–414.

Myers, G. (1991). Lexical cohesion and specialized knowledge in science and popular science texts. *Discourse Processes, 14*(1), 1-26.

Oliver, P. (1996). *Writing essays & reports*: *Teach yourself*. London: Hodder & Stoughton.

Rabinowitz, H. & Vogel, S. (2009) *The manual of scientific style*. Burlington: Academic Press.

Spencer, C. M., & Arbon, B. (1997). *Foundations of writing: Developing research and academic writing skills*. Lincolnwood: National Textbook Co.

Strunk, W. & White, E.B. (2018). *The elements of style.* Vancouver: Spectrum Inc.

Swales, J. M. (2005). Attended and unattended "*this*" in academic writing: a long and unfinished story. *ESP Malaysia*, *11*(1), 1–15.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). Ann Arbor: University of Michigan Press.

UNESCO. (2017). *Science Report.* Retrieved from https://en.unesco.org/unesco_science_report/figures

University of Hull. (2020). Academic Writing style. Retrieved from https://canvas.hull.ac.uk/courses/371/pages/academic-writing-style

WriteCheck. (2020). Writing no-no # 1. Retrieved from http://en.writecheck.com/blog/2012/11/30/writing-no-no-1-video

Ken Hyland is Professor of Applied Linguistics in Education at the UEA. He is best known for his research into writing and academic discourse, having published over 240 articles and 28 books on these topics with 51,000 citations on Google Scholar. A collection of his work was published by Bloomsbury in 2018.

Feng (Kevin) Jiang is Kuang Yaming Distinguished Professor in applied linguistics in the School of Foreign Language Education at Jilin University, China and gained his PhD under the supervision of Professor Ken Hyland at the University of Hong Kong. His research interests include disciplinary discourse, corpus studies and academic writing, and his publications have appeared in most major applied linguistics journals.