

## **Points of reference: changing patterns of academic citation**

Academic citation is the mechanism which both demonstrates the advance of knowledge and distributes credit for priority, emphasising that research is embedded in a literature and that writers are linked into wider social networks. It is central to the ways writers are able to rhetorically build on what has gone before and distinguish their work from that of others. The global explosion of journals and the development of online publication in the last 20 years has increased the volume of research and the access scholars have to it, while the emergence of citation indexes has brought another dimension to this knowledge-constructing role. Now that institutions take citation counts as a measure of research impact and justify promotion and funding decisions on these counts, citations can contribute decisively to the professional careers of those cited (Slyder et al., 2011). Citations are therefore now the currency of the scholarly economy and having one's work recognized and referenced by others is an increasingly valued commodity in today's fiercely competitive academic world (Siler, 2012; Hyland, 2015).

In this paper we explore the effect these factors may have had on citation practices over time and offer evidence for a change in the way knowledge is negotiated by academics. By tracking changes in the form and frequency of citations in four disciplines over the past 50 years we present data which provides new insights into English language academic discourse practices.

### **Citation, knowledge construction and diachronic change**

Reporting the views of authorities, as an alternative to direct observation, has been a feature of academic argument since the Late Medieval period (Taavitsainen, 2002). This is what Palmer (1986: 51) refers to as the 'Quotative' mode of knowing, based on "someone said so" and it is now a central plank in academic argument. Acknowledging sources simultaneously assigns credit for originality and allows readers to validate the researcher's argument, thus serving as a key element of what Gilbert and Mulkey (1980) call the "formal accounting processes of science".

Citation is a key way in which claims are integrated into current knowledge, either by situating the new work in the scaffolding of already accredited facts, or by challenging those facts to carve out a novel position. Successful texts, in other words, need to meet two related rhetorical challenges: they must project the author's individual authority and they must position themselves in relation to their topic and their disciplinary community (Hyland, 2012). This tension between originality and rhetorical accommodation contributes to the collaborative construction of knowledge by negotiating novelty against what is already widely accepted. The embedding of claims in networks of references therefore suggests a cumulative progression, but it also reminds us that statements are invariably a response to previous statements and are themselves available for further statements by others (Bakhtin, 1982). They also firmly locate the author in a social matrix. Citations index community membership and indicate a particular stance within that community: how the author wishes to be seen in relation to others who have made statements on the topic under discussion. It is part of the epistemological and social framework for the acceptance of arguments and, as a consequence, its use and expression reflects changes in disciplinary and social contexts (Swales, 1990; Hyland, 1999).

One aspect of this context is that publishing has not remained static. Both the number of articles and the number of journals have grown steadily for the past two centuries, by about 3 per cent and 3.5 per cent per year respectively (Ware and Mabe, 2012). Around 50 million journal articles have been published since the 1650s (Jinha, 2010) and journals have continued to grow—in quantity, size, and influence—with the number of titles doubling about every 24 years (Ware, 2008). Combined with this growth, and partly as a result of it, the number of references in each paper has increased consistently through the 20<sup>th</sup> century (Bazerman, 1988). Today they are heavily concentrated in the early sections of the article (Hu et al, 2013) to accommodate readers' scanning patterns as they rapidly search for relevance and novelty (Berkenkotter & Huckin, 1995).

The move of journals online in the last 20 years, and the accompanying ease of searching and accessing sources through hyperlinked citations, has meant that references tend to be more recent and that citations are to fewer journals and articles (Evans, 2008). Following hyperlinks from one paper to another not only changes academics' reading processes, but also reduces the potential to encounter

unexpected content (Xia, Myers & Wilhoite, 2010). It quickly puts researchers in touch with prevailing opinion and potentially accelerates consensus. Citations are, therefore, becoming more important in a number of ways, both in terms of the scholarly careers of those cited but also, by narrowing the range of findings and ideas built upon, changing the ways research is conducted and knowledge is constructed.

These changes have attracted some interest in the bibliometric literature, so that Siler (2012), for example, found citation networks in the journal *Social Studies of Science* becoming larger, more structurally cohesive and more modular between 1971 and 2008. Maliniak et al. (2013) found that citation follows a pattern of preferential attachment, with new citations referring to papers that are already popular, and Slyder et al (2011) discovered that author-seniority and journal rankings are key factors influencing the citation rate of a given article. We also know that empirical papers tend to attract more citations than theoretical studies (White, 2001) over time and that Open Access availability also has a positive effect (Xia, et al, 2010). While following diachronic change, however, most studies have sought to track researcher networks, citation frequencies of highly cited articles across post-publication lifespans, or the factors influencing heavy citation. They have, moreover, targeted particular, journals, institutions, authors, modes of access or specific areas of study rather than discipline. There have also been diachronic studies of publishing in social science fields (e.g. Atkinson, 1998; Gross, Harmon & Reidy, 2002; Hyland & Jiang, 2016; Salager-Meyer, 1999) but these have only referred to citation incidentally.

In sum, there has been no systematic investigation concerning the ways citation as a rhetorical practice has changed across time in particular disciplines. In this study we address this gap, focusing on high IF journals in contrasting disciplines to answer the following research questions:

1. Has the frequency of citation changed in key disciplines in the past 50 years and to what extent?
2. Has the way citations are presented changed over the past 50 years in these fields?
3. Has the preference for particular categories of reporting verbs changed?

In addressing these questions we hope to show something of how disciplines accommodate new social conditions in their rhetorical practices and so contribute to our understanding English language academic discourse practices.

### **Corpus and method**

To track changes in citation practices in recent times we created three corpora taking research articles from the same five journals in four disciplines spaced at three periods over the past 50 years: 1965, 1985 and 2015. Applied linguistics, sociology, electrical engineering and biology were selected as representative of applied and pure soft fields and applied and pure hard science fields, as well as established and emergent fields. We were, then, interested in how disciplines of very different characteristics have changed their practices. We took six papers at random from each of the five longest-running journals which had achieved the highest ranking in their disciplinary category according to the 5 year impact factor in 2015<sup>1</sup>. That is, 30 articles in total from each discipline from each year. The journals are listed in Appendix 1 and the corpus comprised 360 papers of 2.2 million words (see Table 1).

Table 1: Corpus characteristics

<b>Discipline</b>	<b>1965</b>	<b>1985</b>	<b>2015</b>	<b>Overall</b>
Applied linguistics	110,832	144,859	237,452	493,143
Biology	244,706	263,465	237,998	746,169
Engineering	92,062	97,545,	235,681	425,288
Sociology	149,788	196,232	262,203	608,223
Totals	597,388	702,101	973,334	2,272,823

Making such a selection is, of course, a contentious process as the top ranked journals today may not have existed 50 years ago and those that have survived half a century of publication may not be among the highest ranked now. Journals come and go, they undergo topic fragmentation and specialisation, they rise and fall in popularity, and they are replaced by new ones over 50 years; these are changes which place considerable constraints on compiling a corpus of comparable papers for diachronic research. We admit that our selection may not reflect the major journals in the field now and that there is a certain heterogeneity in the choice of journals as a result of our sampling. But while there are limitations to our choices which might affect our results, we wanted to maintain the internal coherence

in citation conventions by including the same journals. This means that the biology sample, for instance, contains two review journals and the applied linguistics corpus includes journals not currently at the top of the Thompson-Reuters category rankings. Yet despite these limitations, these are all robust periodicals which have had a significant influence at the top of their respective fields with a long history and, we believe, are representative of both the consistency and diversity of disciplinary practices<sup>2</sup>.

Analysis was conducted by first part-of-speech tagging the three time corpora using *TreeTagger*, and then searching each of them individually for numbers through the regular expression “\w+\_CD” using the concordance software AntConc (Anthony, 2011). Next, we examined concordance lines to ensure we included only canonical citational forms such as a name or date in brackets (1), a number in squared brackets (2) and superscript references (3). We then scanned the corpora for latinate references to other citations (4).

(1) Such models arise naturally in applications of linear networked systems, e.g. for cyclic pursuit (Marshall, Broucke & Francis, 2004). (Electrical Eng)

(2) This is due to the fact that ILS has a higher probability of occurring if the time between consecutive speciation events is short [29,30]. (Biology)

(3) As Hughes put it, “Montreal is the port of entry from which English influence and the industrial revolution radiate into the remote French-Canadian world.”<sup>3</sup> (Sociology)

(4) Those questions designed to probe the encoding function comprise the 'selection of correct lexical item for several types of context' (ibid.: 383). (App Linguistics)

We followed this up with a manual sweep of the corpus for generalised terms for agents, such as ‘these researchers’ and the general referential mention of a source which ascribed propositions using noun phrases such as ‘Durkheim’s theory’ or ‘Davidson’s argument’. Next we distinguished whether authors were referred to in a syntactically integrated way or not (i.e. in the reporting clause or in parenthesis) and how citations were incorporated into the article, as quotation, summary or generalisation.

Finally we checked all the citations we had identified to quantify the use of reporting verbs which, using the co-text of the items to make judgements about the apparent intention of authors, we coded according to Hyland's (1999) categorization of reporting verbs. This is based on the activity the verbs referred to and the type of evaluation they carried. First, we distinguished three kinds of activity:

1. Research Acts, which refer to things in the real-world, often occurring in statements of findings (*observe, discover, notice, show*) or procedures (eg. *analyse, calculate, assay, explore*).
2. Cognition Acts, which are concerned with mental processes (*believe, suspect, view*).
3. Discourse Acts, which involve verbal expression (*ascribe, discuss, hypothesize, state*).

Next we coded for the evaluative meaning of the verb and whether writers represented the reported information positively (*acknowledge, point out, establish*), negatively (*fail, overlook, exaggerate, ignore*) or neutrally, giving no clear signal either way. This last option allows the writer to ascribe a view to the source author, reporting him or her as attributed positive (*advocate, argue, see*), attributed neutral (*address, comment, look at*), attributed tentative (*allude to, believe, suggest*), or attributed critical (*attack, object, refute*). Working independently, both authors of this paper then coded a 10% sample, about 400 examples, refining our agreement through successive passes to achieve an inter-rater reliability of 91%.

We discuss the results in the following sections. In our discussion we follow Hyland (1999) and Thompson and Ye's (1991) useful convention of referring to the person citing as the "writer" and the cited person as the "author". We should point out that the accounts we give for our findings are based on the data and our experience of working with academic discourse, but we are aware that these are simply plausible interpretations rather than definitive explanations.

### **Changing patterns of citation: a quantitative overview**

Overall we found nearly 13,500 citations in the 2015 corpus, representing 137.8 cases per 10,000 words of text. Table 2 shows this was a substantial increase since 1965, with raw figures increasing by 230% and even doubling when adjusted for the large rise in the length of papers (*log Likelihood =*

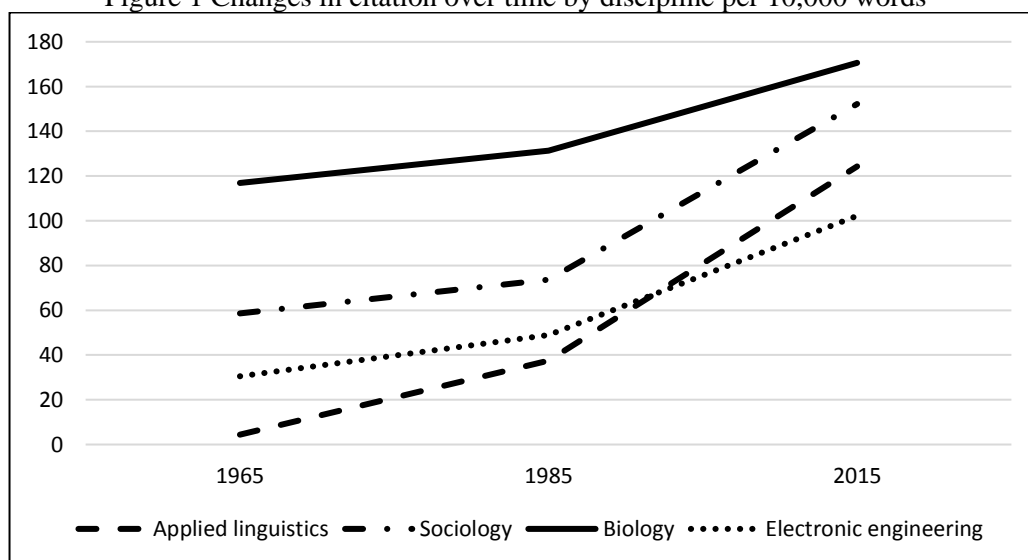
1734.42,  $p < 0.001$ ). Interestingly, this is a finding which seems to have escaped bibliometric research.

Table 2: Total citations over time

	1965	1985	2015	% change
total citations	4068	5926	13411	229.7
per 10,000 words	68.1	84.4	137.8	102.0
Total reporting verb structures	1265	1041	1655	30.8
Reporting as % of citations	31.1	17.6	12.3	-60.5

We also found a significant change in the way that citations are expressed with a 20% decline in structures containing a reporting verb, when adjusted for the length of papers. The table shows that the use of reporting verbs dropped substantially between 1965 and 1985 before picking up again, although this structure now comprises only 12% of all citations. This is a novel, and somewhat surprising, finding which seems to have gone unnoticed in the literature. The increases in citing behaviour over the period have, then, overwhelmingly been in non-integral forms, with a consequent reduction in the scope for evaluating cited material. Importantly, however, the rate of growth of citations during this period has not been uniform across disciplines with the biggest increases in the soft knowledge disciplines, and particularly in applied linguistics ( $LL = 1873.30$ ,  $p < 0.001$ ) (see Figure 1).

Figure 1 Changes in citation over time by discipline per 10,000 words



From an extremely low base, citations in applied linguistics have shot up over 270% (per 10,000 words), reflecting the growth of the discipline itself from an emerging sub-field in the 1960s with a limited literature to become a substantial discipline today with nearly 600 journals listed on the SCImago rankings<sup>3</sup>. Lin and Luyt (2012) found a similar situation when discussing the evolution of an obscure zoology journal, noting that the earliest issues contained no citations at all. Sociology with a 160% increase and electrical engineering, with 235%, also show huge rises in the use of citation ( $LL = 787.89, p < 0.001$ ;  $LL = 501.25, p < 0.001$ ). For biology, which is an established natural science, there is also a significant rise ( $LL = 243.84, p < 0.001$ ). The initial level here is much higher, but this appears to be a particularly heavily citing discipline (cf Adel & Garretson, 2006; Samraj 2008; Hyland, 2004). Swales (2014: 137), for example, states that citation in his A grade undergraduate biology papers “is denser and more variously patterned than most of the other 15 disciplines collected in MICUSP”.

For all these disciplines the biggest rises have, once again, been in the last 30 years, reflecting both the increase in research needing to be referenced and the emergence of the internet and electronic publishing which makes source material easier to access. This only partially accounts for the increase, however, as we may also be witnessing changing patterns of argumentation in these fields. We will discuss these issues further below as we unpack the changing practices our corpora reveal, beginning by looking at citation patterns in terms of writers preferences for particular surface forms.

### **Representation and agency: reporting source material**

One aspect of reporting which has interested researchers is how source material is used in the writer’s argument (Hyland, 2004; Ädel & Garretson, 2006) and the ways cited authors are represented in a text (Swales, 1990; Clugston, 2008).

Clearly the ways writers choose to incorporate others’ work into their own, ranging from extended discussion to casual acknowledgement, can influence how they position themselves towards the reported material and collaboratively construct an effective argument. Choices here concern how far the reporting duplicates or replaces the original language event. Writers have the choice of short direct



quotes (three or more words) as in (5), extensive use of original wording set as indented blocks as in (6), a summary from a single source as in (7), or generalisation where material is ascribed to two or more authors as in (8).

(5) One way of examining validity is by correlating the scores obtained on the dictation test “with other tests designed to assess the same attribute” (Thorndike 1982:187). (AL)

(6) ...As McRae (1982:167) comments on the relationship between social authority and medical profession,

The refusal of state medical boards to protect the public from licensees untrained in acupuncture may rest less on negligence than on an illogical, unproved, and even arrogant assumption that Western medical training is adequate preparation for acupuncture practice.

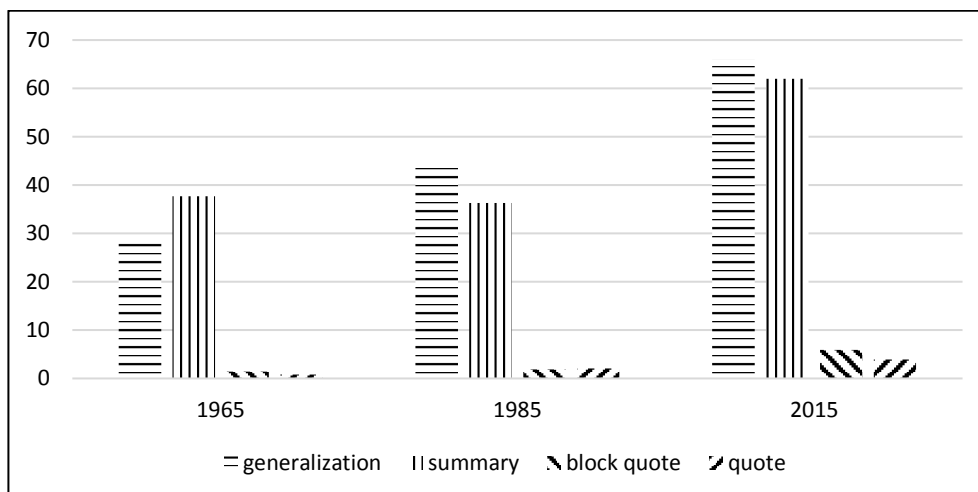
(Soc)

(7) Properly defined links provide a clear picture of the underlying structure of the ensemble of the related data sets [147]. (EE)

(8) Conversely, inhibition or depletion of aurora B kinase, which over-stabilises kMT attachments (by reducing q), causes errors in chromosome alignment and segregation [7], [27], [30], [34], [35]. (Bio)

We can see from Figure 2 that writers overwhelmingly favour putting others' words into their own. This both facilitates greater succinctness, eliminating redundancy and stylistic quirks, while enabling writers to slant the original material closer to their own views and better incorporate it into their argument (cf Swales, 1990). Generalisation and summary thus give the writer greater flexibility to emphasise and interpret what they are citing, reserving quotes for when they consider the original author's words to be the most vivid and effective way of presenting their case. As a result both have increased since 1965 ( $LL = 1113.84, p < 0.001$ ;  $LL = 436.37, p < 0.001$ ).

Fig 2 Proportional distributions of how cited work is presented (per 10,000 words)



The figure, however, also shows a relative decline in the proportional use of summary (by 18.6%), as writers have moved from condensing the position of a single author to generalisation, attributing a claim to several authors, as here:

(9) These institutional changes go hand in hand with the postmodernist intellectual critique of oppressive Western civilization, and internal and external cultural, political and economic colonialism (see Said, 1978; Hechter, 1975). (Soc)

(10) Previous work has demonstrated that, alliinase is an abundant protein in almost all tissues of garlic plants (Nock and Mazelis, 1986; jansen et al., 1989; Lindsluiter et al., 1994). (Bio)

The reasons for this are unclear but are perhaps related to the massive increase in published papers in all disciplines over the past 50 years and the consequent additional research which might be acknowledged. Longer papers also offer writers more space to support their statements with more referencing. The results also show a huge increase in quotation, with both short and block quotes doubling over the period, although not much can be made of this as the normed frequency of 7.1 cases per 10,000 words comprises only 7% of all references. This is, however, almost entirely due to a three-fold increase in sociology and a ten-fold growth in applied linguistics. One might speculate that the discursal turn in the social sciences has affected the treatment of secondary sources<sup>4</sup>.

Table 3 shows that the soft knowledge fields evinced the most pronounced changes in representing cited text. Applied linguistics, for instance, reversed its 1965 preference for summary so that generalisation increased to 54% of all its citations. Sociology on the other hand, strengthened its use of

direct quotes to 13% of all references, ensuring a vivid way of interacting with the views of other researchers. Generalisation also increased as a proportion of citations in both biology and engineering, although summarizing from a single source remained the predominant presentation format in the hard knowledge fields. Thus while there is an increasing need to acknowledge more sources with the growth of published research over recent years, there is still greater engagement with a single line of work in the sciences, with authors collected into a single group, common in the linear development of knowledge characteristic of the sciences. Quotation, however, remains low in these fields with no examples in our 2015 engineering corpus at all.

Table 3 Changes in surface patterns by discipline (per 10,000 words)

Features	Applied linguistics			Sociology			Biology			Engineering		
	1965	1985	2015	1965	1985	2015	1965	1985	2015	1965	1985	2015
<b>presentation</b>												
generalization	1.5	22.0	67.0	30.2	35.3	70.0	45.8	73.6	79.4	10.4	14.8	46.7
summary	2.1	11.4	41.7	23.2	29.1	62.9	69.5	56.6	87.8	19.6	33.3	55.5
block quote	0.7	2.3	7.6	2.1	3.6	12.3	1.6	0.8	2.9	0.5	0.8	0.0
quote	0.1	1.8	8.0	3.1	5.6	7.0	0.0	0.4	0.4	0.0	0.0	0.0

In terms of citation patterns, our study confirms the overwhelming preference for non-integral citation which has been observed in both research articles (Hyland, 1999; Clugston, 2008) and advanced student writing (Ädel & Garretson, 2006; Swales, 2014). As mentioned above, an integral citation includes the cited author within the sentence structure, either as subject (11), as agent (12), as part of a noun phrase (13) or as an adjunct (14), while in a non-integral form the cited authors occur in parenthesis, as in examples 9-10 above (Swales, 1990).

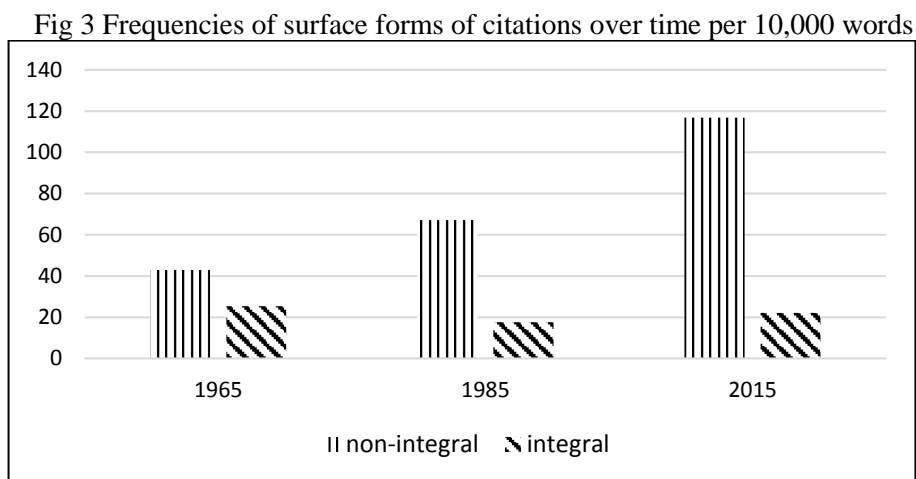
(11) Overdale (2014) argues that this is not an option for the poor. (Soc)

(12) Northern blots were hybridized as described by Van Damme et al. (1992b). (Bio)

(13) Matthei's equations [17, 19] were first used as a starting point in the scale model. (EE)

(14) According to Hirvela (1989), there is no clear pedagogical framework for computer-assisted instruction. (AL)

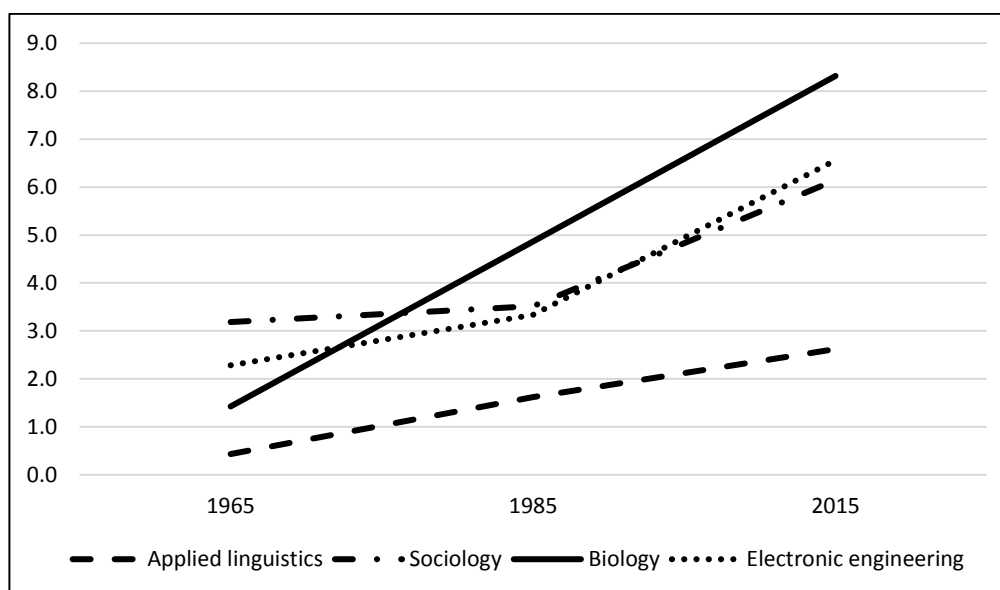
Integral citations are typically associated with a focus on the actions of researchers, while the non-integral ones emphasize the research itself. As we can see from Fig 3, the partiality academics have for non-integral forms is strengthening, increasing by over a third during the last 50 years ( $LL = 2540.81$ ,  $p < 0.001$ ) and now representing 85% of all citations in these four fields.



Clearly writers are moving towards a rhetorical style which gives less prominence to other authors. This may reflect an increasing ‘scientism’ in the social science disciplines, but overall it allows writers to appropriate authors’ contributions with more cursory acknowledgement of their originators than in the past. Cited authors are increasingly relegated to peripheral spaces in the text over time thus allowing writers to introduce material which supports their ideas but does not foreground the originators.

Figure 4 shows that non-integral structures have increased as a proportion of all citations in each discipline since 1965, with the most dramatic shift being in biology, which has risen from 58% to 90% of all forms being non-integral. Biology was also the only discipline to actually reduce its use of integral citations overall, and substantially too, perhaps reflecting a change in its early proclivity for a disciplinary ethos which emphasised a proprietary right to claims; a preference which gave greater weight to *who* originally stated the prior work, rather than the traditional conventions of impersonalisation (Halloran, 1984; Judson, 1995). The split of 90% non-integral to 11% integral for biology in 2015 corresponds exactly to Hyland’s (2004) corpus of articles.

Figure 4 Ratio of non-integral to integral citations



The figures for applied linguistics are also worth mentioning here. While starting from a much lower base in 1965, this discipline has also seen a significant rise in its use of non-integral citations with these forms rising from 29% of all citations to 73% in 2015. During these years, applied linguistics grew from a branch of linguistics to establish itself as an independent field on the borders of the humanities and social sciences and taking on the rhetorical style of more established, and more ‘scientific’ academic communities. Ratios in favour of non-integral citations also increased by 10% in sociology and 17% in engineering which suggests that sociology, like applied linguistics, is adopting the more cursory reporting style of more scientific fields.

#### Activity and evaluation: changes in reporting verbs

Overall the preferred reporting verbs have changed only a little over the past 50 years. *Show*, *find* and *discuss* have slipped in the preferences but remain in the top 10 although *consider*, *study* and *demonstrate* have dropped out of it altogether. *Propose* and *describe* have moved up the top 10 and *argue*, *develop*, *observe* and *suggest* have entered it (Table 4). These broad figures, however, overlook the importance of particular verbs in some disciplines, such as *report* in applied linguistics, *note* in sociology, *use* and *review* in biology and *present* and *define* in engineering.

Table 4: The most frequent reporting verbs across time

1965	1985	2015
show	report	argue
find	find	propose
describe	propose	describe
point out	show	introduce
suggest	describe	find
discuss	suggest	show
study	identify	develop
propose	note	observe
consider	point out	discuss
demonstrate	present	suggest

Perhaps more importantly, the categories these verbs represent have changed considerably over time, with a marked trend towards verbs relating to verbal activities, towards expression in the present tense and towards evaluatively neutral forms. Table 5 presents the distributions of reporting verbs by action, tense and evaluation over the period.

Table 5 Changes in reporting frequencies by discipline (per 10,000 words)

Features	Applied linguistics			Sociology			Biology			Engineering		
	1965	1985	2015	1965	1985	2015	1965	1985	2015	1965	1985	2015
<b>reporting</b>	<b>1.9</b>	<b>9.7</b>	<b>27.0</b>	<b>18.2</b>	<b>17.4</b>	<b>15.1</b>	<b>37.2</b>	<b>18.6</b>	<b>10.5</b>	<b>6.6</b>	<b>7.2</b>	<b>15.6</b>
research	1.0	4.7	9.7	4.4	5.6	6.5	22.9	9.5	5.9	4.7	5.2	9.4
discourse	0.9	4.6	15.2	10.9	10.3	8.2	11.4	7.6	4.0	1.6	1.7	5.7
cognition	0.0	0.4	2.1	2.9	1.5	0.4	2.9	1.5	0.6	0.3	0.3	0.5
<b>Tense</b>												
perfect	0.5	1.9	2.9	3.9	3.5	1.5	11.0	4.6	1.7	2.8	3.2	3.8
past	0.4	4.8	14.7	8.5	8.2	5.1	22.9	11.0	7.1	2.7	3.1	4.9
present	1.0	3.0	9.3	5.8	5.7	8.5	3.3	3.0	1.7	1.1	0.9	6.9
<b>Evaluation</b>	<b>1.9</b>	<b>9.7</b>	<b>27.0</b>	<b>18.2</b>	<b>17.4</b>	<b>15.1</b>	<b>37.2</b>	<b>18.6</b>	<b>10.5</b>	<b>6.6</b>	<b>7.2</b>	<b>15.6</b>
neutral	0.9	4.3	14.8	7.2	8.2	8.6	10.6	7.2	3.8	1.4	3.1	6.8
critical	0.0	0.0	0.0	1.2	1.0	0.4	0.2	0.1	0.2	0.3	0.2	0.2
positive	1.0	5.3	12.2	9.8	8.2	6.1	26.6	11.4	6.7	4.9	3.9	8.6

The table shows statistically significant increases in the use of reporting verbs in applied linguistics and engineering ( $LL = 352.20$ ,  $p < 0.001$ ;  $LL = 46.75$ ;  $p < 0.001$ ), with discourse and research verbs dominating the rises, and a much higher *proportion* of discourse verbs used to report others work

now in all fields. It also shows a shift in preferences from past to present tenses in sociology and engineering, rises in all tenses in applied linguistics and a fall in all in biology, with both these latter fields dominated by the past. Finally, it shows a significant rise in neutral and positive reporting in applied linguistics ( $LL = 199.65; p < 0.001; LL = 152.96; p < 0.001$ ) and engineering ( $LL = 46.74; p < 0.001; LL = 13.26; p < 0.001$ ) while biologists still prefer positive reporting despite a 300% fall in this category.

It would, of course, be perilous to propose a single cause for these changes, but we seem to be witnessing an, admittedly very gradual, change in reporting practices. Writers in applied linguistics and electrical engineering are attempting to integrate their work with that of the discipline more clearly and make it more immediate and plausible to audiences beyond the local disciplinary subgroup. Those in sociology are seeking to be more currently relevant and perhaps more empirically-oriented than in the past, while biologists seem to be reducing their dependence on a disciplinary literature and approaching it more positively. We briefly expand these comments below.

### ***(1) Reporting activities***

The first section of the table distinguishes three categories of activity referred to by the reporting verb, described in the methods section above. These groups are not absolutely watertight, but all represent the involvement of the writer in attributing some action to the author. Both integral and non-integral structures may or may not contain a reporting verb that shows what the cited author did, wrote, or thought, as here, using research (15), discourse (16) and cognition (17) acts respectively:

(15) A growing number of recent studies indicate that ACC oxidase can also be regulated (Kim and Yang, 1994; Barry et al., 1996; Kathiresan et al., 1996). (Bio)

(16) Oxhorn describes the dilemma in persuasive fashion. (Soc)

(17) Raith (2014), for example, considers the network to be "ein sprachkologisches Substrat der Sprachgemeinschaft" (p. 204). (AL)

By foregrounding how the imported text is referred to, such choices direct attention to its authors by ascribing a particular process to them. Hyland (1999) observed that writers in different fields draw on

almost completely different sets of items to refer to their literature, so to summarise, engineers *show*, philosophers *argue*, biologists *describe* and linguists *suggest*. These preferences seem to reflect broad disciplinary purposes. Thus the soft knowledge fields largely use verbs which refer to discourse activities, like *discuss*, *hypothesize*, *suggest* and *argue* which involve the expression of arguments and allow writers to discursively explore issues. Engineers and scientists, on the other hand, prefer verbs which point to the research itself like *observe*, *discover*, *show*, *analyse*, and *calculate*. These represent real world actions and so represent knowledge as proceeding from impersonal lab activities rather than from the interpretations of researchers.

These general characterisations remain broadly true today, especially in the case of applied linguistics which Table 5 shows has strengthened its use of discourse verbs substantially, so that these now comprise well over half of all activity types. The science disciplines continue to favour research verbs, although these have fallen as a proportion of all verbs, and sociology has increased its use of research verbs to almost 45% of all citing types. This moves sociology further towards the reporting practices of the physical sciences, indicating a growing preference for supporting claims with reference to research rather than appealing to verbal arguments:

(18) Thompson shows links to deconstructionism by demonstrating the essential mutability of all language and taking it seriously in its own right.. (Soc)

(19) While Trumbach has studied the pedigree of the modern homosexual man, Lillian Faderman has mapped the pedigree of the lesbian. (Soc)

(20) Here, I am following the notion of disposition as found in the work of Bourdieu (1991) and Wittgenstein. (Soc)

Perhaps surprisingly, while engineers retain a preference for presenting others' work by focusing on their research, they are now much more likely to use a discourse verb than in the past. This corresponds with other changes in electrical engineering which indicate an authorial repositioning, now favouring the present tense and a greater proportion of neutrally evaluative verbs. These changes suggest that writers are doing more to promote both the immediate relevance of their



research and choosing to build on a neutral assessment of what it builds on. These examples suggest the effect of these choices:

(21) It is stated in the cited article that if the ratio of two time constants is small, e.g. there is only a factor of 2 or 3, then they cannot be distinguished. (EE)

(22) Zarebski and Grecki [6] describe the technical details of the thermal transient measurements. (EE)

(23) These authors observe that their cutoff frequencies virtually overlap apart for a narrow range near cutoff. (EE)

Decisions to refer to what others said rather than what they did may be related to the increasing imperative in the applied sciences to address audiences beyond an immediate group of informed insiders. This can also be seen in the wider context of writers tending to use fewer asides and references to shared knowledge over this period (Hyland & Jiang, 2016). Speculatively, these changes may suggest that writers have less confidence in the ability of readers to recover insider allusions and recognise a greater need to guide readers who may be unfamiliar with their specialised practices or knowledge.

## 2) Reporting tenses

The second section of Table 5 concerns the tense of the verb and shows a dramatic shift towards the past (particularly in applied linguistics) and present tenses (especially sociology and engineering).

While the APA manual (2010) advises writers to use the past or present perfect when describing earlier research, MLA (2009) recommends using present tense arguing it is a “signal phrase”, or one that signals either a quote or a paraphrase will follow. The option of past or present tense also relates to the rhetorical embedding of current research in the disciplinary literature. Swales and Feak (2014: 83) believe that the present is often used to emphasize the state of current knowledge while the past tends to refer to single studies. The difference is illustrated here:

(24) The data presented by Parker et al. (1995) and Graham and Bowling (1995) indicate that, for young people, having used a drug is a far from unusual experience.

(Soc)

(25) Druesne (1994) indicated that of a total of 1,897 students who participated in the Lingua Program in 1990-1991, 32 percent of students from the United Kingdom chose to go to Spain and 22 percent opted for France. (AL)

We can, however, see disciplinary preferences for different tenses. Biology and applied linguistics remain strongly wedded to past tense reporting, with 70% of all cases in the former and 55% in the latter, corresponding, perhaps, to a greater reporting of particular studies. Swales and Feak (2014:255) point out that it is more or less obligatory to refer to what a researcher *did*, what we are calling *research verbs*, in the past tense while what the researcher *wrote* or *thought* (discourse and cognitive verbs) offer writers more opportunities to use the present. This flexibility may explain why the present has come to be the preferred tense for sociologists and engineers over the last 50 years. The rhetorical immediacy it conveys helps emphasise the current relevance of reported material to the writer's viewpoint or current state of knowledge (Swales, 1990):

(26) Philip Ferguson (1987) describes and explains the failure of the disability rights movement, and of critical social scientists, to effectively address the disadvantages and oppressions experienced by very cognitively disabled people (Soc)

(27) Kumar et al. [159] propose a Bayesian approach that can identify and eliminate spurious data from a sensor. (EE)

(28) In accounting for how deceit occurs, Simmel (2010; 2012 [1916]: 244–46) suggests life is lived within and outside social and cultural forms. (Soc)

As these examples suggest, present tense options allow the writer to imply a general truth and wider relevance of the others' work to their own. This is a very useful rhetorical strategy in an increasingly competitive academic environment where readers are frequently searching for bottom line applications of research to their own work and writers get rewarded increasingly on the basis of their visibility through citation.

### 3) Reporting evaluation

The final section of Table 5 refers to the use of reporting verbs to convey the writer's evaluation of the imported information, commenting on its evidential status. So writers are able to represent

reported material as correct (as in 29), as mistaken (30) or neutrally (31). These examples give a flavour of the differences between the main categories:

(29) Edson et al (1993) showed processes were induced only after the cells were treated. (Bio)

(30) Connell's theoretical framework fails to explain the complex hierarchies of power which exist within 'subordinated' groups themselves. (Soc)

(31) Ballard and Clanchy (1991) argue that cultural attitudes toward knowledge range on a continuum from valuing the conservation of knowledge to valuing the extension of knowledge. (AL)

It seems that criticality tends not to be expressed through reporting verbs in any discipline and writers in the hard sciences generally employ slightly more verbs indicating their belief in the truth of a report while those in the soft fields withhold judgement. Table 5 shows that only biologists have reduced their use of neutral verbs, and even then the *proportion* increased from 28% to 36% of all cases. Both applied linguists and engineers have increased their use of positive signals but these increases have been exceeded by a rise in neutral verbs which averages about 10 cases per 10,000 words. Overall, then, writers are referring to prior work as a way of supporting their own research rather than critiquing it to establish a gap.

The presentation of material without giving an explicit indication of the writer's perspective, is nevertheless a useful evaluative option as it allows a position to be attributed to the author.

Despite very low normed frequencies the changes we observe in our data reflect a shift in how disciplines, through the behaviour of its members, adjusts its knowledge-making machinery by changes in expressing a position towards prior work and cited authors. Generally, and in line with their attempts to use cited material in support of their own work, writers in all disciplines credit the cited author as taking a positive view towards what they had discussed. This attribution of a positive attitude has risen in all fields except biology over the past 50 years, so that attributions of positive positions account for 56% of all verb choices. This is partly, perhaps, because writers want to subtly recruit the author's position in support of their own

claims. These examples suggest something of the way such positive attributions can contribute to the writer's subsequent argument:

(32) Acton (1984) sees preparing students psychologically as a necessary correlate to improving their pronunciation. Phonology, he says, has both "inside-out" and "outside-in" dimensions which function in a kind of loop: speakers can control their nerves or inner states by speaking properly. This is the basic tenet of successful programs in voice training and public speaking. (AL)

(33) Barry & Diamond (1984) and Dainty (1985) have argued that water relation parameters are considerably underestimated if concentration-polarization effects occur within the root tissue. However, in contrast to the root pressure probe data it was found that the crr values decreased with increasing concentration of the osmoticum in the external solution. (Bio)

(34) Craib (1995) claims that this form of constructionism leans towards a subjectivist cognitivism. But what is also a problem here is the way that the object of study is confused at various points. (Soc)

Although there is a preference for positive attributions to the author, ascribing a neutral view to authors makes up most of the remaining neutral verbs choices. Their growth since 1965 is due largely to the flexibility they offer the writer to build on or refute the statements made. Representing the author as taking a neutral stance provides the writer with a point of departure for their own claims: a clean slate from which to launch an argument which builds on the cited work without necessarily supporting or criticising it:

(35) McClintock (1984) looks at the mobilizing impact of public policy, and Midlarsky and Roberts' (1985) discuss the impact of state-elite relations. (Soc)

(36) Coupland's focus is more firmly grounded in social criticism in that he maps a novel territory - that of ageism. (AL)

(37) Much of the modeling research in this area focuses on the evaluation of alternative storage policies in SIR systems with respect to expected transaction service/waiting times (Malmborg 1996). (Eng)

(37) MacArthur (2012) addresses the specific question of whether Iraq intended or planned to invade Saudi Arabia. (Soc)

Thus the growth in neutral verbs across all fields but biology enables writers to attribute positive or neutral attitudes to the author concerning the research presented, making it possible to build on earlier work in the way they think best without taking a specific stance towards it. The writer is free to simply recognise what has gone before, nodding to the understandings of the community and acknowledging past work, and quickly moving on to their own claim.

### **Some observations and conclusions**

How writers choose to present reported information provides important insights into the context of academic persuasion. The diachronic changes discussed in this study indicate several general trends over the past 50 years which suggest changing rhetorical practices in response to a more crowded, competitive and evaluative environment with greater consequences for writers. Here we answer our three research questions and seek to explain some of the major changes.

First, there has been a massive increase in citations, particularly over the past 30 years, with references to prior work doubling per 10,000 words of text across our four corpora since 1965, and increases in applied linguistics and sociology particularly marked. We have noted that among the possible reasons for this are that there is much more research to report and that recognising previous work is much easier now as a result of electronic access to source texts. Another, and equally important reason, however, is the enhanced role that citation is required to play to ensure that current work is appropriately embedded in significant research issues and disciplinary understandings of those issues (e.g. Bazerman, 1988: 164). With greater competition and topic specialisation it is now more crucial than ever for writers to carve out a distinctive niche and define a specific contribution as their own. Citation is a key way of both establishing the relevance of their work to current concerns and, perhaps more importantly, its significance as an individual contribution.

Second, the period has seen a relative decline in the use of reporting verbs, an increase in non-integral structures and generalisation replacing summary, the last of these most noticeably in the soft fields. These trends suggest a movement towards suppressing human agency in knowledge-making and placing greater emphasis on the reported studies rather than those who conducted them. This change has occurred in all fields, with sociology, electrical engineering and biology now comprising between 86% - 89% of non-integral citations. While non-integral forms have shot up over 30% in biology, applied linguistics has seen a massive 7000% rise, reaching 71% of all cases in that field. The main reason for this large rise is obviously the emergence of applied linguistics as a scientific interdisciplinary field of inquiry from an incipient body of work largely concerned with second language teaching and first language literacy. The growth and increasing specialization of the field with ever smaller niches, each with a body of literature, has been accompanied by a movement towards greater technicality or scientism in some areas with an associated trend towards scientific reporting styles. This is a trend occurring across all fields we studied.

Third, we have documented changes in the types of reporting, with the most significant results being an overall decline in structures which contain reporting verbs and a shift to a preference for generalisation, neutral stances, discursive representation and present tense citing. It is difficult to see any overall rationale behind these changes, especially given the disciplinary variations in the figures. Both applied linguistics and electrical engineering are resisting the overall trend and have increased their use of citations and reporting verbs to offer positive evaluations of prior work, to represent authors in terms of their research rather than their discussions of it, and to present all this in the past tense. Thus despite an overall movement towards less explicit recognition of authors and a prioritizing of objectivity in citations, the more applied disciplines in our study have tended to interpret this trend by engaging with authors by showing more explicitly how earlier research supports their own and to highlight specific studies in the past. Overall, however, it appears that disciplinary differences are slowly narrowing in terms of how citations are used in argument patterns.

We need to repeat here that the journals in our corpus represent disciplines with different degrees of internal coherence, at different stages of maturity and which take very different epistemological

positions, so it is not easy to make overall comparisons or offer a general explanation for them. It is, moreover, difficult to identify motives for rhetorical changes and, indeed, to read off citation functions from purely textual evidence. Harwood (2009), for instance, found that writers are rarely able to give clear purposes for their citing decisions and Willett (2013) shows that there is often little correspondence between the reasons writers and readers give for citing a particular source or using a particular form in a given reference. But while individual purposes for many citations might remain elusive, investigating a corpus aggregates these distinct choices its overall patterns reveal a certain automaticity by writers as a result of their familiarity with the routine practices of a discipline. By offering evidence of actors' orientations to scholarly communities and the ways they stake out individual positions through repeated citation practices, corpora show how decisions are regularly performed and recognised as legitimate.

As we hope to have shown here, corpora not only offer evidence of mutual recognition and community-preferred patterns of alignment, they also indicate changes in these practices. The social conditions under which research is conducted and rhetorically negotiated with colleagues are no more static than the epistemologies which inform disciplinary beliefs and research. In particular, the growth of managerialism in academia and the rise of an accountability culture places ever-greater demands on academics to ensure that they reach wider audiences and take credit for their work for career purposes.

It is no longer sufficient for a writer to simply fill a research gap; that gap has to be shown to matter to the community (and often to those beyond it) so that the writer has to more clearly 'own' the space that he or she carves out. By showing how their work both builds on and can be distinguished from what has gone before, citation helps writers to assert both novelty and a certain proprietorial relationship to ideas. Thus, by increasing reference to prior work, representing that work in their own words, placing it in the present tense, signifying it was what an author said rather than did, reporting it using neutral verbs and relegating authors to parenthetical asides, writers are able to give less prominence to the work they build on while focusing on their own.

Overall, these changes appear to have proceeded by stealth and have gone largely unnoticed in the literature, but while our findings should be corroborated by further research in other fields, we believe we can see here a gradual shift in persuasive practices. Citation remains and, indeed, is perhaps more important than ever as an element of the scholars' rhetorical armoury, but as more research is being included to support new claims, writers are increasingly reporting it more impersonally and with greater emphasis on its contribution to their own work. With disciplines becoming ever more specialised, and its research required to gain impact outside of the immediate academic discipline, we might expect these trends to continue as writers seek to ensure that their work is securely anchored in antecedents that are also accessible to non-specialists.

### Notes

1. Two journals, *TESOL Quarterly* and *Foreign Language Annals*, began only in 1967, and so article were chosen from issues in that year. *College Composition and Communication (CCC)* is the only one not indexed in SSCI, but it is an influential, long-running journal listed in the language and linguistics subcategory of Arts and Humanities in the Thomson Reuters citation index.

2. We are aware that aggregating numbers from journals in four disciplines may not represent patterns over the entire range of academic activity. While such extrapolations are common in corpus studies, any temptation to generalise to the entire population should only be done with extreme caution.

3. The SCImago journal lists can be found at <http://www.scimagojr.com/journalrank.php>

4. We are indebted to an anonymous reviewer for this intriguing observation.

### References

Ädel A. & Garretson G. (2006). Citation practices across the disciplines: The case of proficient

student writing. In Perez-Llantada C., Plo R., Neumann C.-P. (Eds.), *Academic and professional communication in the 21st century: Genres and rhetoric in the construction of disciplinary knowledge* (pp. 271-280). Zaragoza, Spain: Prensas Universitarias.

Anthony, L. (2011). *AntConc 3.4.3*. <http://www.laurenceanthony.net/software.html>

Atkinson, D. (1999). *Scientific discourse in sociohistorical context: The philosophical transactions of the Royal Society of London, 1675-1975*. Mahwah, NJ: Lawrence Erlbaum.

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



- Bakhtin, M. (1982). *The Dialogic Imagination*. Austin: University of Texas Press.
- Bazerman, C. 1988. *Shaping written knowledge*. Madison: University of Wisconsin Press.
- Berkenkotter, C. & Huckin, T. 1995 *Genre knowledge in disciplinary communication*. Hillsdale, New Jersey: Lawrence Erlbaum.
- Clugston M. (2008). An analysis of citation forms in health science journals. *Journal of Academic Language and Learning*, 2, 11-22
- Evans, JA (2008). Electronic Publication and the Narrowing of Science and Scholarship. *Science*. Vol. 321, Issue 5887, pp. 395-399
- Gibaldi, J. (2009). *MLA handbook for writers of research papers*. New York: The Modern Language Association of America.
- Gilbert, G. N. and M. Mulkay. (1984). *Opening Pandora 's box: a sociological analysis of scientific discourse*. Cambridge: CUP.
- Gross, A. G., Harmon, J. E., & Reidy, M. S. (2002). *Communicating science: The scientific article from the 17th century to the present*. Oxford University Press.
- Halloran, S. 1984. The birth of molecular biology: an essay in the rhetorical criticism of scientific discourse. *Rhetoric Review*, 3: 70-83.
- Harwood, N. (2009). An interview-based study of the functions of citations in academic writing across two disciplines. *Journal of Pragmatics*, 41, 497-518.
- Hu, Z., Chen, C. & Liu, Z. (2013) Where are citations located in the body of scientific articles? A study of the distributions of citation locations. *Journal of Informetrics*. 7: 887-896.
- Hyland, K. (1999). Academic attribution: citation and the construction of disciplinary knowledge. *Applied Linguistics*. 20 (3): 341-267
- Hyland, K. (2004). Patterns of engagement: dialogic features and L2 student writing. In Ravelli, L. & Ellis, R. (Eds.) *'Analyzing academic writing: contextualized frameworks'*. London: Continuum. pp 5-23.
- Hyland, K. (2012) *Disciplinary Identities*. Cambridge: Cambridge University Press
- Hyland, K. (2015). *Academic publishing: issues and challenges in the construction of knowledge*. Oxford: OUP

- Hyland, K., & Jiang, F. (2016). Change of attitude? A diachronic study of stance. *Written Communication*, 33(3), 251-274.
- Jinha, A. E. (2010). Article 50 million: An estimate of the number of scholarly articles in existence. *Learned Publishing*, 23/3, 258–263.
- Judson, H. 1995. *The eighth day of creation: the makers of the revolution in biology*. Harmondsworth: Penguin Books.
- Lin, H., & Luyt, B. (2012). The evolution of scientific journal articles in the periphery: a case study and analysis of the Raffles Bulletin of Zoology from 1928 to 2008. *Journal of Information Science*, 38(5), 407-422.
- Maliniak D., Powers, R. and Walter, B. (2013). The Gender Citation Gap in International Relations. *International Organization*. doi:10.1017/S0020818313000209
- Palmer, F.R. (1986) *Mood and modality* (Cambridge: Cambridge University Press).
- Salager-Meyer, F., (1999). Referential behaviour in scientific writing: a diachronic study 1810–1995. *English for Specific Purposes* 18 (3), 279–305.
- Samraj, B. (2008). A discourse analysis of master's theses across disciplines with a focus on introductions. *Journal of English for Academic Purposes*, 7, 55-67.
- Siler, K. (2012). Citation choice and innovation in science studies. *Scientometrics*, 95(1), 385–415.
- Slyder, J. B., Stein, B. R., Sams, B. S., Walker, D. M., Jacob Beale, B., Feldhaus, J. J., & Copenheaver, C. A. (2011). Citation pattern and lifespan: A comparison of discipline, institution, and individual. *Scientometrics*, 89(3), 955–966.
- Swales, J. (1990). *Genre Analysis: English in Academic and Research Settings*. Cambridge: CUP.
- Swales, J. (2004). *Research genres*. Cambridge: Cambridge University Press.
- Swales, JM (2014) Variation in Citational Practice in a Corpus of Student Biology Papers: From Parenthetical Plonking to Intertextual Storytelling. *Written Communication* 2014 31: 118
- Swales, J.M. and Feak, C.B. (2012) *Academic Writing for Graduate Students: A Course for Nonnative Speakers of English 3<sup>rd</sup> ed.* Ann Arbor, MI: University of Michigan Press.
- Taavitsainen, I. (2002). Historical discourse analysis: Scientific language and changing thought-styles. In T. Fanego, B. Méndez-Naya, & E. Seoane (Eds.), *Sounds, words, texts and change* (pp. 201-226). Amsterdam: John Benjamins.

- Thompson, G. & Ye, Y. 1991. Evaluation of the reporting verbs used in academic papers. *Applied Linguistics*, 12: 365-82.
- Thompson, P. (2012). Achieving a voice of authority in PhD theses. IN Hyland, K. & Guinda, C. (eds.). *Stance and voice in written academic genres*. London: Palgrave.
- Ware, M. (2008). Peer review: Benefits, perceptions and alternatives. *Publishing research consortium*. Retrieved from <http://www.publishingresearch.net/documents/PRCsummary4Warefinal.pdf>
- Ware, M., & Mabe, M. (2012). *The STM report: An overview of scientific and scholarly publishing, third edition*. Oxford: International Association of Scientific, Technical and Medical Publishers.
- White, H. D. (2001). Authors as citers over time. *Journal of the American Society for Information Science and Technology*, 52(2), 87–108.
- Willett, P. (2013). Readers' perceptions of authors' citing behavior. *Journal of Documentation*, 69, 145-156.
- Xia, J., Lynette Myers, R., & Kay Wilhoite, S. (2010). Multiple open access availability and citation impact. *Journal of Information Science*, 37(1), 19–28.

## **Appendix 1: Journal list**

### **Applied Linguistics**

TESOL Quarterly (1967- )

Language Learning (1948- )

Foreign Language Annals (1967- )

Modern Language Journal (1916- )

College Composition and Communication (1950- )

### **Sociology**

American Journal of Sociology (1895- )

Social problems (1953- )

The British Journal of Sociology (1950- )

American Journal of Economics and Sociology (1941- )

The Sociological Quarterly (1960- )

### **Biology**

The Quarterly Review of Biology (1926- )

Biological Reviews (1923- )

Radiation Research (1954- )

BioScience (1964- )

The Journal of Experimental Biology (1923 - )

### **Electrical Engineering**

Proceedings of the IEEE (1963 - )

Automatica (1963 - )

IEEE Transactions on Automatic Control (1963 - )

IEEE Journal of Solid-State Circuits (1966 - )

IEEE Transactions on Information Theory (1963 - )