Attribution in novice academic writing

Hilde Hasselgård

University of Oslo

Academic attribution, the direct acknowledgement of external sources, is investigated in two corpora of novice academic English, representing first and second language writing in linguistics. The forms and uses of attribution are analysed in a formal-functional framework. There is an overall underrepresentation of attribution in the learner corpus. However, the corpora have a similar proportional distribution of integral and non-integral attribution, but a difference in subtypes of these. Undated attributions are discussed as a special case. They occur in specific contexts, of which reference to course reading is peculiar to novice writing. Comparisons with expert corpora in Norwegian and English indicate that some, but not all, of the differences between the novice corpora may be linked to influence from the learners' first language and culture.

Keywords: academic English, novice writing, attribution, Norwegian learners of English

1. Introduction

Writing academic texts involves engaging with other academic texts to the extent that "citation is the most overt and most immediately obvious indication that a text is indeed academic" (Swales 2014: 119). At the same time, scholars have observed that novice academic writers often do not use sources according to the conventions of their discipline (Pecorari & Shaw 2012), and that second language writers have additional challenges in this area (e.g. Hyland 2009; Verheijen 2015: 102). Both first and second language novice academics need to learn that "appropriate reference to other texts is an essential feature of most academic writing" (Thompson & Tribble 2001: 91) and that, arguably, "academic writing depends for its success on situating current work in a larger disciplinary narrative" (Hyland 2002: 115).

Academic attribution involves explicit reference to an external text used as a source of information or authority, as illustrated in examples (1) and (2).[1]

(1)   <u>As Eggins (2004) states</u>, this kind of contrastive analysis is useful…
        (BAWE_3127b)

(2)   In general, labels are "unspecific and requires lexical realisation in its
        immediate context, either beforehand or afterwards." <u>(Francis 1994: 162)</u>.
        (VESPA_UIO0265-LIN-01)

The aim of this study is to survey the extent and the nature of academic attribution in two corpora of novice academic writing by native speakers of English and advanced Norwegian learners of English within the discipline of linguistics. This involves investigating how formal-functional types of attribution are distributed in the corpora and what similarities and differences there are between the first language (EL1) and the second language (EL2) writers in their preferred forms of attribution. The study does not include the content of the citation or the writer's purpose of using it. Nor does it aim to uncover errors in the academic attributions of novice academics, but rather to observe the forms of attribution they use and how these are integrated in the text.

To find out whether novices use attribution differently from expert writers in the same discipline and whether the EL2 writers may be influenced by their first language (L1), a corpus of published articles in both English and Norwegian is also consulted, partly drawing on the findings of Fløttum et al. (2006). Despite the existing body of research on attribution, there are no studies of Norwegian EL2 writers. Furthermore, few studies of academic attribution in EL2 writing address the issue of influence from the writers' L1, and the present study is an attempt to fill these gaps.

Previous research on first and second language academic writing has shown that Scandinavian learners of English overuse markers of writer/reader visibility and stance compared to native speakers of English (e.g. Aijmer 2002, Larsson 2017, Hasselgård 2009; 2017; Paquot et al. 2013). Such markers can be indicators of text averral, i.e. statements which express the writer's voice, and for which the writer takes responsibility (Sinclair 1986; Tadros 1993). Academic attribution, by contrast, represents the presence and voice of "others"

---

[1] See Section 3 for a description of the corpora. All examples are quoted verbatim, including any infelicities. Underlining has been added for highlighting.

in the text (Fløttum et al. 2006; Charles 2006a; Hyland 2009). Where stance markers belong to the interpersonal aspect of academic discourse, attribution is more ideational, with its "prominent role in the ways writers seek to construct facts through their communicative practices" (Hyland 2000: 21). The question is whether the Norwegian learners, as well as the EL1 academic novices, understand this part of academic discourse, in which actors other than the writer and reader take precedence.

Academic attribution is expected to be underrepresented in the learner corpus compared to the native speaker corpus. This is mostly because EL2 writers in general have been found to use fewer citations (e.g. Borg 2000; Hyland 2009), but also due to the observed abundance of averral features in texts by Norwegian learners (e.g. Paquot et al. 2013). Furthermore, Fløttum et al. (2006: 231) report that the frequencies of bibliographical references in published linguistics papers – in the authors' respective L1s – vary across languages, "with English first, Norwegian second and French third". Thus, cultural and/or L1-related factors may cause Norwegian learners to use attribution less than their English-speaking peers, and at the same time lead to different preferences as regards the forms of attribution chosen by the two writer groups.

After a review of relevant literature in Section 2, the material and method of the study are introduced in Section 3. Section 4 presents the corpus analysis, Section 5 contains further discussion, and Section 6 offers some concluding remarks.

2. Literature review

According to Swales, reference to previous research is an obligatory component of any academic paper for establishing a territory, providing "a *specification* … of previous findings, an *attribution* to the research workers who published those results, and a *stance* towards the findings themselves" (1990: 148, emphasis in original). Much work on attribution and citation has focused on the research article (e.g. Swales 1990, Thompson & Ye 1991, Hyland 1999, 2000, Fløttum et al. 2006, Hyland & Jiang 2017), but also on other types of academic writing such as academic textbooks (Tadros 1993) and PhD dissertations (e.g. Thompson 2005; Charles 2006a, 2006b) and on spoken academic English (Ädel 2008). Comparisons have been made across disciplines (e.g. Hyland 1999, 2000; Charles 2006a, 2006b; Fløttum et al. 2006; Ädel & Römer 2012; Hyland & Jiang 2017; Nesi 2021) and across levels of writer expertise and/or L1 background (Borg 2000; Pecorari 2006; Petrić 2007; Mansourizadeh & Ahmad 2011; Swales 2014; Verheijen 2015; Wiemeyer 2019). Some studies highlight the challenge of novice academic writers: Groom (2000), for instance, identifies attribution as an important

area of EAP instruction, and Thompson & Tribble (2001: 100) report on problems in students' use of citations. Attribution practice has also been studied in relation to plagiarism, e.g. Pecorari & Shaw (2012), and to the accuracy and the textual integration of direct quotes in student writing (e.g. Docherty 2018; Wiemeyer 2019). Studies of the rhetorical functions of attribution and citations have combined corpora with interview data (e.g. Pecorari 2006; Petrić & Harwood 2013), since corpus data are arguably insufficient sources of such information (Swales 2014: 120).

The most common classification schemes for the surface forms of attribution draw on Swales's distinction between *integral* and *non-integral* citations:

> An integral citation is one in which the name of the researcher occurs in the actual citing sentence as some sentence-element; in a non-integral citation, the researcher occurs either in parenthesis or is referred to elsewhere by a superscript number or via some other device. (Swales 1990: 148)

This distinction is applied in many subsequent studies of attribution, e.g. Hyland (2000), Thompson & Tribble (2001), Fløttum et al. (2006), Petrić (2007), and in the present study (see Section 3). Swales further distinguishes between reporting citations, where the cited author occurs as a grammatical subject accompanied by a reporting verb "to introduce previous researchers and their findings" (1990: 150), and non-reporting citations, which have other types of subjects and verbs.

Verbs accompanying reporting citations have been studied as a means of conveying the writer's evaluation of the citation; see e.g. Thompson & Ye (1991), Hunston (1995) and Charles (2006a, 2006b). Charles (2006b) investigates the phraseology of reporting clauses used in citations. The pattern of **V[erb] that** in reporting clauses comprises four semantic verb groups: ARGUE, THINK, SHOW and FIND (ibid.: 319), with the ARGUE group being most frequent. The most common citation type, integral with human subject, is considered to give prominence to the cited author (2006b: 316). Hyland (2000) identifies disciplinary differences as regards verbs accompanying citations: "philosophy, sociology, marketing and applied linguistics largely favoured discourse activity reporting verbs, and the engineering and science papers displayed a preference for research-type verbs" (2000: 28). The disciplines also differ as to the density of citations, measured per 1,000 words and as an average per text (ibid.: 24). For the present study it is interesting that the applied linguistics discipline has an

above-average citation density. All the disciplines except philosophy have a majority of non-integral citations (ibid.).[2]

Groom (2000) draws on the work of Sinclair (1986) and Tadros (1993) to study the means by which writers express their own textual voice or "acknowledge the presence of an antecedent authorial voice" (2000: 15). Integral and non-integral citations (re-labelled "author-prominent" and "information-prominent"; ibid.; 18) are interpreted pragmatically as tools that writers can use for positioning themselves in relation to the cited authors.[3] Author-prominent citations place the writer in a subordinate position, while information-prominent citations make the writer's textual voice more dominant (ibid.:19).

Thompson & Tribble (2001), building on Swales (1990), present a structural and functional classificatory framework for citations which forms the basis of the present analysis. Their main distinction is between integral and non-integral citations. Integral ones can be *verb-controlling*, where "the citation acts as an agent that controls a verb, in active or passive voice" (2001: 95), similar to Swales's reporting citations. An integral *naming* citation is (part of) a noun phrase with a function other than agent (ibid.: 96), similar to Swales's non-reporting integral citations. Non-integral citations are labelled *source* if they "attribute a proposition to another author" (ibid.: 95), *identification* if the citation "identifies an agent with the sentence it refers to", *reference* if the citation is explicitly signalled by a directive such as *cf.,* and *origin* where a concept or product is attributed to its originator (ibid.). Finally, so-called "non-citations" occur where "there is reference to another writer but the name is given without a year reference" (ibid.: 96). The investigation reveals differences in citation types across different stages of the academic paper, academic disciplines and individual writers.

Both first and second language novice academic writers need to learn the conventions of attribution of their disciplines (Borg 2000; Verheijen 2015; Wiemeyer 2019). Examining citation practice in academic writing by EL2 post-graduates, Pecorari (2006) finds that students often do not master the conventions of academic citation, the relationship between the cited and the citing text, and the appropriate use of quotation marks to delimit the quoted matter. Interview data reveal that these aspects of academic writing were "in a blind spot" for both students and supervisors (Pecorari 2006: 24).

---

[2] Hyland and Jiang (2019) discovered a growing number of citations and an increasing reliance on non-integral citations in research papers over the last 50 years.

[3] Groom follows Thompson & Ye (1991: 366) in using "'writer' to refer to the person who is reporting and 'author' to refer to the person who is being reported".

Studies of academic attribution in second language contexts have pointed to both cross-linguistic and cross-cultural differences. As Carter-Thomas & Rowley-Jolivet (2013) find, this may concern everything from the linguistic form of reporting structures to differences regarding writer- and reader-responsibility in French vs. English writing cultures.

Fløttum (2003) and Fløttum et al. (2006) are important points of reference for the present study as they investigate attribution across three languages (English, French, and Norwegian) and three disciplines (economics, linguistics, and medicine). Quantitative differences are greater across disciplines than across languages, with linguistics having more attributions than economics but fewer than medicine (Fløttum et al 2006: 220). English and Norwegian generally use attribution with equal frequency (ibid.: 221), but in linguistics, English has a significantly greater number. The functional interpretation of integral attribution is that the floor is given to the cited author(s) (ibid.: 230). Linguistics favours integral attribution with indirect speech and non-integral references that mention publication year (ibid.: 231). The types of verb that accompany cited material display considerable lexical variation across the subcorpora. English and Norwegian linguistics mainly use verbs that denote argumentation (e.g. *say, argue, note*), while other disciplines use more factive verbs (e.g. *find, show*) (ibid.: 236 f).

Far from being exhaustive, the above review of previous research contains items that have informed the present study in various ways, for example as to the classification of attributions and how to retrieve them in a corpus. Studies of EL2 novice academic writing and cross-linguistic comparisons of English and Norwegian expert academic writing are sources of hypotheses about of attribution in the EL2 corpus than in the EL1 corpus. However, studies of EL1 novice writing also demonstrate that the general (and variously attributed) idea that "nobody is native speaker of academic English" pertains also to the use of attribution.

3. Material and method

3.1 Corpora and software

In line with Granger's (2015) revised model of contrastive interlanguage analysis, this study compares a Norwegian-based interlanguage variety of English to a comparable L1 reference corpus quantitatively, identifying the frequencies of attribution, and qualitatively, examining the nature of the attributions. Both corpora contain novice academic writing in English in the form of student assignments from various academic disciplines, namely the Varieties of English for Specific Purposes dAtabase (VESPA) and the British Academic Written English

corpus (BAWE). For this study I use only linguistics texts by students whose first language is Norwegian (from VESPA) and English (from BAWE). The choice is partly pragmatic, since VESPA in fact only contains (English) literature and linguistics in any usable quantity, and partly motivated by the opportunity to compare results with the cross-linguistic study of Fløttum et al. (2006). To enhance corpus comparability, only texts above 1,000 words in length were considered for the present study, thus excluding 123 texts from VESPA and three from BAWE. Furthermore, texts which did not contain any instances of attribution were excluded, thus reducing the material with another 31 texts from VESPA and one from BAWE. Table 1 shows the size and composition of the resulting subsets of the corpora.

Table 1 Size and composition of the corpora used in the present study[4]

Corpus	Texts	Mean text length	S.D.	Words
VESPA	111	1963.9	778.3	217,993
BAWE	72	2331.3	1094.6	167,856

VESPA and BAWE are comparable in many respects: both comprise disciplinary texts produced by BA and MA students as part of their course work, with a majority coming from the undergraduate level (see Nesi & Gardner 2012 and Paquot et al. 2013). However, the percentage of MA-level texts is higher in VESPA than in BAWE: 45.9% vs. 19.4%. This is because the texts removed from VESPA largely came from the undergraduate level. All the texts were produced in untimed conditions with access to reference works. The writers in BAWE are students of applied linguistics (in the UK) and those in VESPA of English linguistics (in Norway). Skills-based, introductory courses are not represented. However, BAWE texts are assignments "with grades of at least 60 per cent (or equivalent)" (Nesi & Gardner 2012: 7). This is not the case in VESPA because most of the texts are (non-final) course assignments, which are not graded in Norwegian universities. All the texts are specifically disciplinary, and the exclusion of texts without attribution should ensure that the writers to some extent relate their discussion to literature in the field. The genre classification of BAWE texts is more detailed and systematic than the one found in VESPA, but in both

---

[4] Word counts were performed with WordSmith after excluding material tagged as e.g. reference lists and linguistic examples (Ebeling & Heuboeck 2007).

corpora, the majority of texts have been assigned to the category of "essay".[5] Both VESPA and BAWE have functional annotation (see Ebeling & Heuboeck 2007), so that material not originally produced by the writer (e.g. quotations and reference lists) can be ignored in corpus searches. The corpora were searched by means of WordSmith Tools 6.0 (Scott 2012).

In addition, the KIAP corpus (Fløttum et al. 2006)[6] will be consulted as a potential source of explanation for differences between EL1 and EL2 writing. This corpus is multidisciplinary and multilingual, containing published articles on linguistics, economics, and medicine in English, French, and Norwegian. Only the English and Norwegian linguistics parts of KIAP are relevant for this study. Each contains 50 texts comprising 437,798 words in KIAP-Eng and 269,913 words in KIAP-Nor (Fløttum et al. 2006: 7). Research articles and student assignments obviously differ in important respects: "scholars and students write for different audiences, have different writing goals, and use different genres, all of which could affect their citation use" (Petrić 2007: 239). Nevertheless, the KIAP corpus is a valuable source of information about expert L1 usage in the relevant discipline and in both languages.

3.2 Identification and classification of attributions

An instance of attribution is "a research report which has a specific reference point that is clearly identifiable" Charles (2006b: 314). Such reference points can constitute the basis for corpus searches. Following the practice of e.g. Thompson & Tribble (2001) and Fløttum et al. (2006), I first searched for attributions through the year of publication (four-digit numbers starting in 19 and 20 and the abbreviation *ibid. – op.cit.* was not attested). The resulting concordances were sifted manually to delete irrelevant hits, e.g. when the number was not a publication year. In a second step, lexical collocates of the year references within a span of +/- 5 words were identified with WordSmith and used as search terms to identify instances of attribution without a publication year. Such collocates include *cf.* and *according (to)*, report verbs (e.g. *argue, state*), and nouns referring to texts (*article, book*). The attributions identified through the latter procedure were included unless they duplicated those identified by publication year. Unlike Thompson & Tribble (2001) and Hyland (2000), but similarly to

---

[5] The vast majority of VESPA texts have been labelled "essay", except a few term papers. Of the BAWE texts, 62.5% are essays. The other text types represented with more than five texts are critique (12.5%) and methodology recount (11.1%).

[6] KIAP is the Norwegian acronym for Cultural Identity in Academic Prose (Fløttum et al. 2006). The corpus was accessed from http://clarino.uib.no/korpuskel/page.

Fløttum et al. (2006), I did not search for author names. References given in captions, titles and footnotes were excluded from this investigation (as in Fløttum et al. 2006).[7]

Many of the corpus texts, being linguistics papers, discuss other texts/corpora as the data for analysis. References to such material are excluded from this investigation, as they do not manifest "the presence of researchers other than the author(s)" (Fløttum et al. 2006: 215). Secondary attribution, as in (3), was counted as a single instance of attribution; likewise multiple sources for the same content, as in (4).

(3)   Similarly women have been assigned the stereotype of asking lots of questions (Brouwer et al 1979 as in Coates 2004: 93)… (BAWE_6026a)

(4)   There appears to be a degree of instability in genre definition and classification (Bhatia, 1993; Fairclough, 2003; Paltridge, 1996)... (BAWE_3127b)

The attributions were classified according to an adapted version of Thompson & Tribble's (2001) framework (see Figure 1). No distinction was made between attributions with direct and indirect quotes (unlike Fløttum et al. 2006), as formal signals of quotation in novice academic writing have been found to be unreliable (Pecorari 2006; Wiemeyer 2019). Thompson & Tribble's (2001) two types of non-integral citations *source* and *identification* (see Section 2) may be hard to distinguish – especially in student writing – because the "agent" in an identification may also be identical to the author of the cited paper; hence they were merged and labelled 'source'. Furthermore, 'non-citation' was renamed 'undated attribution' to avoid the pejorative connotations of Thompson & Tribble's term and because dated and undated attributions alike make reference to external texts. The classificatory framework is outlined in Figure 1. Examples of the different types are given in (5)-(10).

---
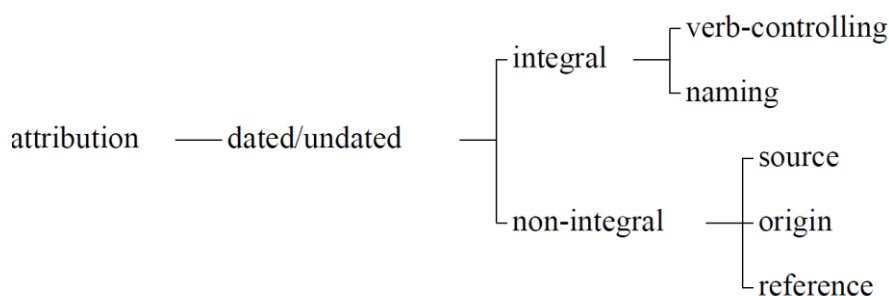
[7] Footnotes in BAWE and VESPA have a mark-up which makes it possible to retrieve them; however, they appear after the running text of the paragraph they belong to, and are thus difficult to interpret from concordance lines (Heuboeck et al. 2010: 31 f).

Figure 1 The classification of attribution types

(5)   INTEGRAL, VERB-CONTROLLING: <u>Peters (2004:400) argues</u> that ought to's place in English is shrinking,... (VESPA_UIO0030-LIN-01)

(6)   INTEGRAL, NAMING: I based this format on <u>Lixian and Cortezzi's (1996) table</u> comparing cultural and academic attitudes... (BAWE_3118b)

(7)   NON-INTEGRAL, SOURCE: …where it follows its lexicalization, it will be called a retrospective label. <u>(Francis 1994: 157)</u> (VESPA_UIO0265-LIN-01)

(8)   NON-INTEGRAL, ORIGIN: Briscoe et al (1998) used <u>the Bus story (Renfrew, 1991)</u> to assess narrative skills… (BAWE_6206b)

(9)   NON-INTEGRAL, REFERENCE: The unit following the conjunct is seen in relation to the one which precedes it <u>(cf. Quirkal. 1985: 637)</u>. (VESPA_UIO0204-LIN-03)

(10)   INTEGRAL, VERB-CONTROLLING (UNDATED): <u>Dam-Jensen and Zethsen mention</u> the Danish word *stinke* in their article, which is a cognate of the English *stink*. (VESPA_UIO0025-LIN-03)

For an attribution to qualify as verb-controlling, the agent named in the running text must be identical with the cited author. The attribution in (11) is thus non-integral (source) because the agent is Wernicke, but the cited author is Caplan.

(11)   Wernicke emphasises that patients do not mispronounce phonemes <u>(Caplan 1987: 50)</u>… (BAWE_6206c)

Attributions written in unorthodox ways have been grouped with the most similar standard form. For example, (12) is considered non-integral even if the author's name appears outside the brackets because it is not syntactically integrated in the clause.

(12)   These words are commonly mixed up by foreign learners because of their similarity in form. <u>Johansson (2008: 132)</u> (VESPA_UIO0042-LIN-02)

Verb-controlling attributions contain a report verb or a research verb. This distinction was originally made by Thompson & Ye (1991), simplified by Hyland (1999; 2000) and applied in even simpler form here. My category of report verbs includes both verbal and mental process

verbs (Halliday 1994: 107), e.g. *claim, say, suggest,* and *believe*, thus conflating Hyland's categories of 'cognition' and 'discourse' (2000: 28). Typical representatives of research verbs (Thompson & Ye 1991: 370) are *conduct, demonstrate,* and *identify*. The verb classification was made from individual concordance lines; thus e.g. *find* can be a report verb or a research verb depending on whether it means 'think' or 'locate'. Like Thompson & Ye (1991: 336) and many subsequent studies, I use the term 'writer' to refer to the originator of the texts under study and 'author' to refer to the originator of the cited works.

4. Corpus analysis

4.1 Frequency and dispersion of attribution

The frequency of attributions is almost twice as high in BAWE as in VESPA at group level (103.18 vs. 54.36 per 10,000 words). However, both corpora show considerable variation across texts, ranging from 0.4 to 14.9 attributions per 1,000 words in VESPA (S.D. = 3.7) and from 0.6 to 27.1 in BAWE (S.D. = 9.9). The text dispersion is shown in Figure 2. The interquartile range is greater in BAWE than in VESPA, in addition to a considerably higher median and mean frequency and a less dense clustering of data points at the lower end of the box. The 95% confidence intervals are non-overlapping (Brezina 2018: 20).
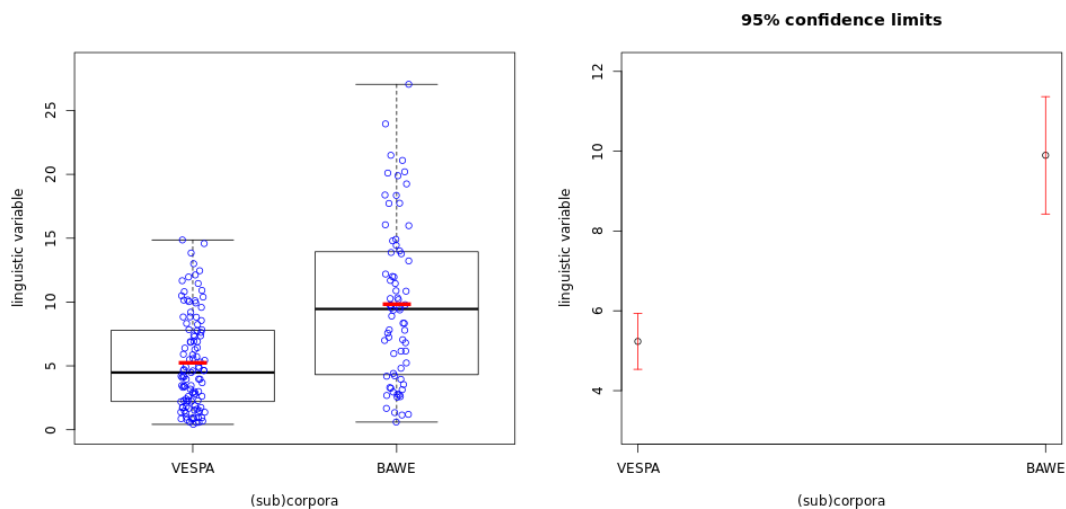
Figure 2 Dispersion of attribution across corpus texts (freq. per 1,000 words)

The quantitative differences between the two writer groups are hard to explain. To some extent they may reflect variables that have not been (or could not be) controlled for, such as the aforementioned issue of high-graded vs. ungraded papers, which is due to

different university regulations between the two countries where the corpora were compiled. However, the difference between VESPA and BAWE echoes Fløttum et al.'s (2006: 221) finding that attributions are significantly less frequent in published linguistics papers in Norwegian L1 than in English L1, and may thus indicate cultural differences in academic practice.

Within both VESPA and BAWE the average number of attributions appears to increase with study level (as in Nesi 2021: 14), but the tendency is not consistent – there are BA-level texts with above-average numbers of attributions and MA-texts with below-average numbers in both corpora. There may also be individual variation as well as intra-writer variation. For example, BAWE writer 6010 has two essays of the same length, both from year 1, of which one contains 18.4 attributions per 1,000 words and the other 9.8. Another potentially relevant factor is text type. However, the BAWE texts, which have the most detailed genre classification, do not show any consistent pattern of variation for the number of attributions per 1,000 words in e.g. essays and critiques; for example, writer 6020 has almost twice as many attributions in an essay as in a critique while writer 6038 uses more attributions in a critique than in an essay (13.2 vs 9.8 per 1,000 words in texts 6038c and 6038b). The number of texts is too small for reliable quantitative analysis across all these variables. On the whole, it will be more useful to study the ways in which students attribute their sources, so the emphasis in the remainder of this study will be qualitative, albeit with some quantitative backing.

4.2 Categories of attribution in the corpora

The major categories of attribution, integral and non-integral, occur in equal proportions in both VESPA and BAWE, as shown in Table 2.

Table 2 Frequencies of attribution types

		VESPA			BAWE		
		N	%	%	N	%	%
integral	verb-contr	392	33.1	50.0	696	40.2	50.9
	naming	201	17.0		185	10.7	
non-integral	source	544	45.9	50.0	837	48.3	49.1
	origin	6	0.5		11	0.6	
	reference	42	3.5		3	0.2	

Total		1185	100		1732	100	

There are both similarities and differences in preferred subtypes of integral and non-integral attribution. Non-integral source is the most common type of attribution in both corpora, followed by integral verb-controlling. VESPA has higher proportions than BAWE of integral naming and non-integral reference. The following subsections will look further into different types of integral attribution (4.2.1 and 4.2.2) and non-integral attribution (4.2.3). While the present section does not distinguish between dated and undated attribution, the use of non-dated attribution is considered specifically in section 4.3 because such attributions are less well described in previous research.

4.2.1 *Verb-controlling attribution*
Verb-controlling attribution is frequent in both corpora. The cited author is most commonly the subject of the controlled verb, but can also be the agent in a passive construction. Report verbs outnumber research verbs in both active and passive verb-controlling attributions, as shown in Table 3.

Table 3 Features of the verb in verb-controlling attributions

		VESPA			BAWE		
		N	%		N	%	
report verb	active	261	66.6	78.8%	491	70.5	76.0%
	passive	48	12.2		38	5.5	
research verb	active	67	17.1	21.2%	135	19.4	24.0%
	passive	16	4.1		32	4.6	
Total		392	100		696	100	

The proportions of report and research verbs are fairly similar in VESPA and BAWE. The picture is comparable to that of the 'soft disciplines' in Hyland (2000: 28). However, VESPA has more passives than BAWE. The most frequent passive verb lexemes in VESPA are *mention* and *make*. The latter, exemplified in (13), occurs repeatedly in almost identical patterns, i.e. *claims made by Tottie and Algeo* (or just one of these authors), and evidently echoes the wording of a particular writing task. In BAWE the most common passive verbs are *describe* and *report*.

(13)   …so I will compare my findings with the <u>claims made by both Tottie and
        Algeo</u>. (VESPA_UIO0070-LIN-03)

As in Hyland (2000) and Fløttum et al. (2006), the controlled verbs display a high degree of
lexical variation. The type-token ratios are practically identical in VESPA and BAWE, at
19.9% and 19.0%, respectively, with most of the types occurring only once or twice. The
most frequent report verb lexemes are shown in Table 4 and the research verbs in Table 5.

Table 4 The ten most common report verb lexemes in verb-controlling attributions.
Percentages of report verbs in each corpus and distribution across texts.

	VESPA (N=309)			BAWE (N=529)			
LEXEME	N	%	N of texts (of 111)	**LEXEME**	N	%	N of texts (of 72)
define	38	12.3	27	suggest	84	15.9	22
mention	35	11.3	19	state	64	12.1	29
suggest	26	8.4	18	write	41	7.8	6
say	20	6.5	14	point out	33	6.2	24
claim	20	6.5	14	describe	30	5.7	20
state	17	5.5	14	claim	23	4.3	14
point out	16	5.2	11	report	23	4.3	6
write	15	4.9	11	say	23	4.3	14
explain	15	4.9	9	define	20	3.8	10
argue	13	4.2	10	note	16	3.0	8

The verb preferences in the two corpora overlap to a great extent: of the top ten lexemes,
seven (in shaded cells) are shared although they appear at different ranks. The top three items
in VESPA are also the most widely dispersed ones, but even *define* only occurs in 24% of the
texts. By contrast, the most widely dispersed item in BAWE, *state*, occurs in 40% of the texts,
while *write* is boosted by 33 occurrences in two texts by the same student. The second most
frequent report verb in VESPA, *mention,* occurs only five times in BAWE (rank 19) and only
twice as a report verb in KIAP-Eng. It is possible that its popularity in VESPA is due to L1
influence; its closest Norwegian equivalent, *nevne,* occurs 17 times as a report verb in KIAP-
Nor (6.3 per 100,000 words), and is thus a viable option, though less frequent than *mention* in
VESPA (16.1 per 100,000 words). The top item in BAWE, *suggest,* ranks third in VESPA,

but its proportion is twice as great in BAWE. It is noteworthy that *suggest* does not have an obvious counterpart in Norwegian but rather corresponds to a range of different words such as *foreslå* 'put forward' and *antyde* 'indicate', whose combined frequency as report verbs (dominated by *foreslå*) is lower than that of *nevne* in KIAP-Nor.

Table 5 The five most common research verb lexemes in verb-controlling attributions. Percentages of research verbs in each corpus and distribution across texts.

	VESPA (N=83)				BAWE (N=167)		
LEXEME	N	%	N of texts	**LEXEME**	N	%	N of texts
call	12	14.5	7	find	12	7.2	6
distinguish	7	8.4	5	use	10	6.0	7
give	7	8.4	5	demonstrate	9	5.4	8
divide	5	6.0	5	identify	9	5.4	7
use	5	6.0	4	provide	8	4.8	6

The research verbs used by the linguistics students reflect the business of classifying, naming, and finding out about things. Because of their low degrees of recurrence and dispersion only the five most frequent research verb lexemes are shown in Table 5. Only *use* occurs in both lists, typically concerning the use of terms, data, and methods. The favourite in VESPA is *call*, illustrated by (14), while BAWE prefers *find*, as in (15). Both examples show typical uses of these lexemes.

> (14)   The last kind of reference is what Thompson (2004) *calls* comparative.
>   (VESPA_UIO0201-LIN-02)
> (15)   However, Bartok, Rutter and Cox (1975) have *found* this not to be the case…
>   (BAWE_6206e)

4.2.2 *Naming attribution*

In naming attribution, the name of the cited author occurs as part of a clause constituent, but not as the head of an NP functioning as agentive subject (active) or PP complement (passive). The NP containing the author's name can be the subject or object, or the complement of a preposition. PPs introduced by *according to* are singled out as a separate type. The structural types of naming are exemplified in (16)-(19).

(16)   SUBJECT: <u>A study by Brown (1957)</u> demonstrated this. (BAWE_6067e)

(17)   OBJECT: This supports <u>Renouf and Sinclair's claim</u> about frameworks, and indeed many collocations, "lying somewhere between word and phrase" (1991: 129). (VESPA_UIO0033-LIN-05)

(18)   *ACCORDING TO:* <u>According to Pinker and Prince (1988)</u> this is a serious shortcoming… (BAWE_6038e)

(19)   OTHER PP: <u>In Fairclough (2001)</u>, the notion of causality is stressed as being at the root of representing power in discourse... (VESPA_UIO0265-LIN-01)

Naming attributions in subject and object NPs typically attach the author name to nouns such as *study, finding(s), argument*, and *claim*, i.e. products or activities of the cited authors, as in (16) and (17). The type shown in (19) resembles verb-controlling attribution, but the preposition *in* signals that the reference is to the work rather than the author (Li & Panther 2014: 233).
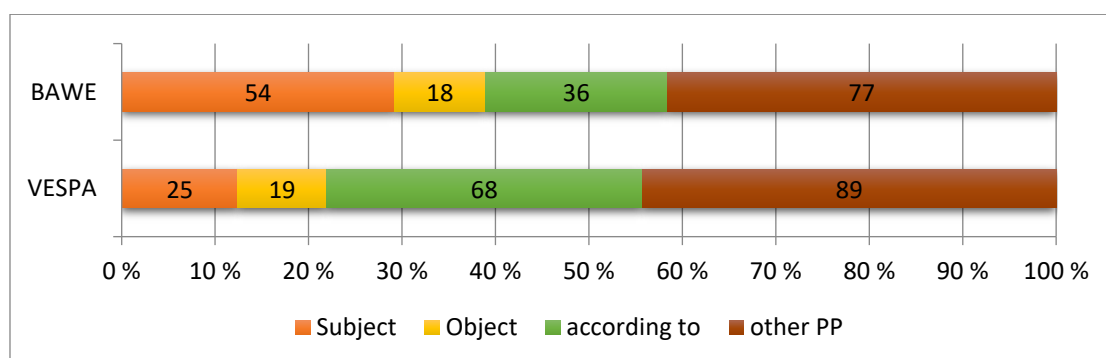
Figure 3 Structural types of naming. The numbers in the figure show raw frequencies.

Figure 3 shows that the structural types of naming are distributed differently in the corpora. The difference concerns mainly the proportions of subjects vs. prepositional phrases with *according to*. VESPA writers use naming in subject NPs much less than the BAWE writers, clearly preferring PPs with *according to* as well as other prepositions. However, a subject NP may be an alternative to a clause-initial PP, as demonstrated by (20), which paraphrases (19) above. The low frequency of subject NPs with naming attribution in VESPA, as in (16) and (20), may be due to the fact that they involve syntactically more complex structures than the PP in (19) and are thus challenging for EL2 writers (Parkinson & Musgrave 2014).

(20) <u>Fairclough's (2001) notion of causality</u> is stressed as being at the root of representing power in discourse...

Thompson & Tribble (2001: 100) regard the repeated use of *according to* in novice academic texts as a symptom of a "lack of variety of citation types". The popularity of *according to* in VESPA may be interpreted in this light.[8] Furthermore, *according to* is more frequent in BAWE than in KIAP-Eng, at 21.4 vs. 13 occurrences per 100,000 words, while VESPA has 31.2. *According to* is also more widely distributed in VESPA, occurring in 38% of the texts compared to 25% of the BAWE texts; in both corpora with frequencies from 1 to 6 per text. Apart from the novice factor, the frequency of *according to* in VESPA may be helped along by the closest Norwegian counterpart of *according to*, *ifølge,* which occurs in attributions 15.2 times per 100,000 words in KIAP-Nor, making it slightly more frequent than *according to* in KIAP-Eng.

4.2.3 *Non-integral attribution*

The proportion of non-integral attribution (compared to integral attribution) is similar between VESPA and BAWE, but the distribution of subtypes of non-integral attribution differs considerably (see Table 2). In particular, reference – attribution with a directive – is much more frequent in VESPA, with 42 occurrences where BAWE has only three, i.e. 16.1 and 1.8 occurrences, respectively, per 100,000 words.

Searches in KIAP for *cf., cp.* and the corresponding Norwegian *jf., jfr., kfr.* retrieve 25.6 hits per 100,000 words in KIAP-Eng and 151.5 in KIAP-Nor. Not all of these are intertextual attributions; some are for example metadiscursive (as in *cf. Section 2.2*). But judging from the first 60 hits in each corpus, about 2/3 of the occurrences of both *cf.* and *jf.* are part of references. Thus, reference appears to be common in expert academic writing, and more so in Norwegian than in English. Norwegian expert writers seem to use reference (in their L1) in contexts where their English-speaking colleagues might have preferred to omit the directive. This may be illustrated by (21), where the reference with a page number appears as the writer's source of information. This use occurs in KIAP-Eng, too, but the directive *cf.* seems more common with general pointers to sources that make similar claims as the present writer, as in (22), or represent more incidental information. Thus, it appears that English and

---

[8] Wiemeyer (2019: 146) found very few instances of *according to* in German learner English, but this might be because she looked primarily at direct quotations.

Norwegian (L1) academic writing have different conventions for the use of references with directives so that the frequency difference between the novice corpora may reflect English and Norwegian attribution practices.

(21)   Syntaktisk kan de stå alene, og de kalles på bakgrunn av dette for setningsekvivalenter (<u>jf Lie 1991: 25</u>). (KIAP-Nor)
        'Syntactically they can stand alone, and because of this they are called sentence equivalents (cf. Lie 1991: 25).'

(22)   But an independent linear precedence effect favors Fillers (here nominal heads) before their gaps or resumptive pronouns, i.e. Fillers First (<u>cf. Hawkins 1999b</u> for supporting evidence). (KIAP-Eng)
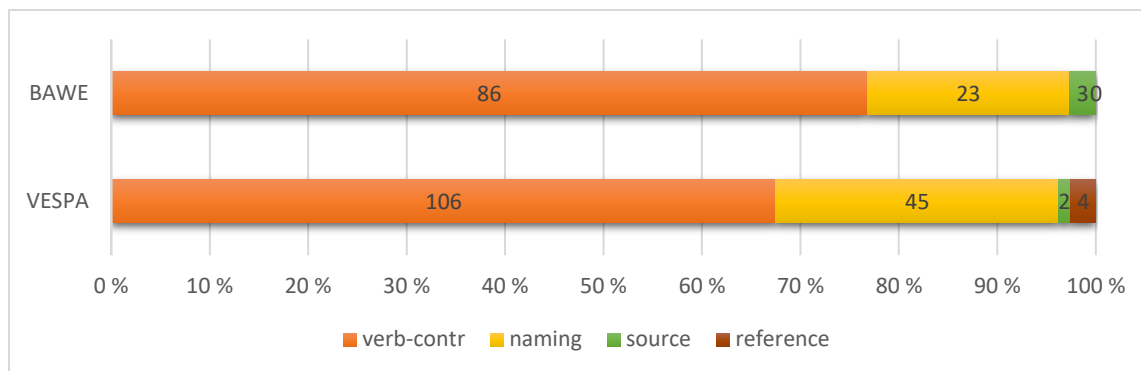
4.3 The use of undated attribution

Table 6 shows the frequencies and proportions of dated and undated attribution in VESPA and BAWE. Compared to dated attribution (containing publication year or *ibid.*), undated attribution is an infrequent option, though with a higher proportion in VESPA than in BAWE, as shown in Table 6. The difference is significant (χ^2= 37.86, p<0.0001, φ = 0.1151).

Table 6 Frequencies of attribution in VESPA and BAWE

	VESPA		BAWE	
	N	%	N	%
dated	1028	86.75	1620	93.53
undated	157	13.25	112	6.47
Total	1185	100	1732	100

Figure 4 shows the distribution of the types of undated attribution. Both writer groups prefer integral, verb-controlling attribution, as in (23), followed by naming attribution. This may be partly a reflection of the search terms for retrieving undated attributions (see Section 3.2). Non-integral attributions are rare, with 'origin' missing from both corpora.

(23)   Salkie writes that repetition of *function words* and *content words* can make a text more coherent. (VESPA_UIO0053-LIN-01)

Figure 4 Types of undated attribution. The numbers in the figure show raw frequencies.

Individual texts in both corpora may contain both dated and undated attributions, or only one type. Of the 111 VESPA files with attribution, 12.6% have only undated instances, 46.8% have only dated ones, and 40.5% contain both types. The corresponding percentages for the 72 BAWE files with attribution are 2.8%, 45.8% and 51.4%. This suggests that undated attributions are mostly used as a supplement to dated ones.

Undated attributions should not be written off as sloppy or incompetent citation practice, as they typically alternate with dated ones in certain contexts. As Thompson & Tribble (2001: 96) observe, undated attribution "is most commonly used when the reference has been supplied earlier in the text and the writer does not want to repeat it". Both VESPA and BAWE have numerous examples of this, as illustrated by (24) and (25), where undated attribution occurs in the vicinity of a dated attribution to the same source. A similar example from KIAP is given in (26).

(24) …I have consulted Hilary Hillier's "Analysing real texts" (2004). As suggested in this book, I have divided the articles into clauses… (VESPA_UIO0265-LIN-01)

(25) Graddol and Swann (1989) discuss empirical studies undertaken by Zimmerman and West (1975) which focussed on interruptions made in conversations. […] They argue that interruptions are a way of controlling conversation... (BAWE_6010e)

(26) According to Swales (1990, p. 58), a genre is "…a class of communicative events, the members of which share some set of communicative purposes". Further, he argues that genres have certain structural characteristics… (KIAP-Eng)

Undated non-integral attribution may be signalled by reference to the text, as in (24), or by pronominal reference to the author(s), as in (25) and (26), or by the author's name, as in (29) below. Undated attribution is also common when the cited source is a reference work such as a dictionary, and this practice is found with the novices as well as the experts, as illustrated by (27) and (28).

> (27) <u>Oxford English Dictionary's (OED)</u> defines phraseology as a 'selection or arrangement of words and phrases in the expression of ideas… (VESPA_UIO0175-LIN-01)
>
> (28) …as defined explicitly in any one lexical entry for V <u>in the American Heritage Dictionary</u>? (KIAP-Eng)

However, one use of undated attribution is unique to novice writing: EL1 and EL2 students alike sometimes omit dates when referring to course reading, in which case they probably presume that the reader (their teacher) is familiar enough with the source. This seems to be the case in (29), which is one out of 20 undated references to Ingram in BAWE. A similar use is found in a number of VESPA papers that answer the aforementioned assignment on "claims made by Tottie and Algeo" (see example (13) above) where two specific sources are given in the prompt so that it may be superfluous to give the full reference every time they are mentioned.

> (29) <u>Ingram writes</u> that "consonants tend to be voiced when preceding a vowel… (BAWE_6062e)

Sometimes the reference point includes the name of the work cited, as in (30), which for all its detail lacks a publication date for the textbook and thus indicates insufficient knowledge of attribution conventions.

> (30) As Stig Johansson writes <u>in the compendium, *Contrastive analysis and learner language - A corpus based approach, chapter 8.7*</u>, this is not unexpected… (VESPA_HIOF0007-LIN-02)

To estimate the extent of undated attribution in published academic papers, the most frequent controlled verbs in KIAP were investigated, English *argue* and Norwegian *si* ('say') (Fløttum

et al. 2006: 233). Of the 151 attributions with *argue*, 35% were undated, as were 23% of the 104 with *si*. If these verbs are representative, the percentage of undated verb-controlling attributions in VESPA is not particularly high (27%); rather, the one in BAWE is very low (12.4%). The present data thus do not support Thompson & Tribble's (2001: 100) claim that student writers overuse undated attribution. The only type of undated attribution that seems peculiar to the novice corpora concerns course reading, in which case the familiarity of the cited source makes some writers depart from the conventions of academic attribution.

5. Discussion

5.1 The frequency of attribution in novice texts

This study has uncovered both similarities and differences in attribution practice between the two novice corpora VESPA and BAWE. A conspicuous difference is quantitative: attribution is considerably less frequent in VESPA than in BAWE (cf. Section 4.1). This frequency difference was expected, as e.g. Borg (2000) found that EL1 students produced significantly more citations than EL2 students in response to the same task, and Hyland (2009) reports similar findings. The underrepresentation of attributions in EL2 writing compared to EL1 writing may have many reasons. One is the EL2 factor, which could make academic writing (even) more difficult than for EL1 students. As noted by Carter-Thomas & Rowley-Jolivet (2013: 112), there is a double challenge in using a second language while acquiring the discourse norms of academic writing. However, there are also significantly fewer attributions in Norwegian than in English in the published linguistics articles in KIAP (Fløttum et al. 2006: 219), so the learners may be influenced by their native language and culture.

The frequency of attributions in the BAWE material may be boosted by an apparent tendency to "over-attribute". There seem to be more instances in BAWE than in VESPA of the type shown in (31) and (32), where the same material is attributed twice:[9] in (31) to the same source and in (32) to different ones, probably because the writer took the Schiffrin reference from Cameron. Even if this is most likely a secondary attribution, it is presented as having two sources.

> (31) After this same period, which may also be seen to coincide with the one-word
>   stage (Trask, 1995), the rate at which a child acquires new vocabulary seems to

---

[9] Wiemeyer (2019: 145) found similar examples of "over-citation" in academic writing in English by German learners.

increase more and more rapidly, as well as the length of his/her utterances; he/she moves on to the two-word stage, then three word stage, etc. (Trask, 1995). (BAWE_6050a)

(32)   One of the most integral parts of the approach, according to Schiffrin (1994) is placing the discourse within a cultural and social context, which enables the 'big picture' to be seen (Cameron 2004: 52). (BAWE_6020c)

5.2 Attribution types

The proportional distribution of integral and non-integral attributions is relatively similar between the corpora (Table 2). Moreover, the proportions of report and research verbs are similar (Table 3). Thus, as in Borg's study (2000: 39), the non-native writers use sources in a similar fashion to native writers in terms of the form and function of attributions.[10] However, the choice of subtypes of integral and non-integral references differs: within integral attribution, VESPA has more naming (possibly linked to L1 influence) while BAWE has more verb-controlling attributions. VESPA uses more non-integral attributions of the reference type (with a directive), which may reflect both linguistic and cultural issues in EL2 academic writing, as this attribution type was also used more in KIAP-Nor than in KIAP-Eng.

The use of integral vs. non-integral attribution is a way for writers to choose a role with respect to the cited source. Integral attributions are author-prominent, placing the cited author in a dominant position, while non-integral attributions are information-prominent, and place the writer in the dominant position (Groom 2000). The fact that both groups of novice writers employ integral and non-integral attribution in equal measure shows their willingness to share propositional responsibility (ibid.: 22) with the cited authors for the attributed material. Groom suggests that "the interplay between these two citation foci provides a major resource for the hierarchical positioning of textual and intertextual voices that all successful argumentative academic texts must achieve" (ibid.: 18). As the proportions of integral and non-integral attributions are relatively similar between the novice and the expert corpora, too (see Fløttum 2003: 119), the novices appear to have grasped the required interplay.

Some studies highlight the evaluative potential of report verbs in verb-controlling attributions, e.g. Thompson & Ye (1991), Hunston (1995) and Charles (2006a), i.e., the choice of report verb can reveal whether the writer agrees or disagrees with the cited content or remains neutral to it (Thompson & Ye 1991: 372). The ten most frequent verbs in verb-

---

[10] However, Ädel & Römer (2012: 13) found that US linguistics students greatly preferred integral attribution.

controlling attributions in VESPA and BAWE (Tables 4 and 5) are either neutral or positive as regards the writer's alignment with the cited author, as also observed in Thompson (2005: 41). Only two of the most frequent verbs, *claim* and *argue,* are susceptible to signalling disagreement, as in (33), but they can also occur in contexts where the writer gives a neutral report of the cited matter, as in (34); see Hunston (1995: 156). On the whole, the novice writers' use of verb-controlling attribution seems to indicate trust in the cited sources or a neutral attitude to them; thus the students bypass this opportunity to express their stance.

(33)   <u>Lakoff claims</u> that the two key differences between women's and men's language are the use of hedges […] and boosters […]. *However* these two features might in actual fact be the language of an unconfident speaker of either sex. (BAWE_6126d)

(34)   <u>Hunston argues</u> that the downside of DDL is that the teacher has "very little control over what happens" (2002: 171). (VESPA_UIO0038-LIN-02)

Undated attributions occur in both VESPA than in BAWE, but they are more common in VESPA (Section 4.3). However, as undated attribution occurs in published academic writing, too, the use of undated attribution *per se* cannot be considered a weakness in novice academic texts. Although some instances of undated attribution reveal that the writer is unfamiliar with academic conventions (as in example (30)), most of the undated attributions in the novice corpora appear in contexts where they are perfectly acceptable, particularly in the immediate vicinity of a dated attribution to the same source. Another use, also found in KIAP, was undated attribution to a reference work. A practice that is peculiar to the novice corpora, however, is that of omitting the date when referring to (what appears to be) course reading; this can be observed in both VESPA and BAWE.[11]

BAWE was found to contain more attributions than VESPA that involve either serial sources or secondary citation. These phenomena have not been thoroughly analysed here, but they may reflect an important function of attribution in student writing, namely that of "knowledge display" (Petrić 2007: 239). It is not immediately obvious why the learner corpus should be any different, but it is possible that both the nature of the course reading and the grading of the assignment may play a role.

---

[11] The different contexts of undated attribution were not quantified, partly because the location of a dated reference to the same source was generally outside the concordance window, and partly because it was difficult to identify a source as course reading.

Carter-Thomas & Rowley-Jolivet (2013) describe linguistic and cultural differences in attribution practice between French L1 and English L1 writers. In contrast, Fløttum et al. (2006) argue that language seems to have a smaller effect than discipline on academic writing practice. Nevertheless, they observe significant differences between English and Norwegian linguistics articles concerning the frequency of attribution (ibid.: 221). Such a frequency difference is also observed between BAWE and VESPA. Yet, we should not conclude that L1 influence is responsible for the patterns of attribution in VESPA. As the VESPA writers are students of English, they have not necessarily read enough academic articles in Norwegian to copy their attribution style. But the influence may be less direct: Norwegian patterns of argumentation (at least within the humanities) seem to involve more dialogism, as indicated by the greater visibility of reader and writer (Fløttum et al. 2006, Paquot et al. 2013) and less intertextuality (as indicated by the present study). The Norwegian-produced academic texts in English may thus reflect a transfer of cultural practice. However, since Borg (2000) saw similar features in texts by learners from different cultural and linguistic backgrounds, further research, involving comparison with other learner groups, is needed to elucidate this point.

6. Closing remarks

The aim of this investigation was to survey the extent and the nature of academic attribution in English by EL1 and EL2 novice writers. There were both similarities and differences between the two writer groups. A major difference was quantitative; however, though expected, this difference was hard to explain due to the relatively sparse data (if all the known variables are considered) and due to the fact that a number of variables remain unknown in spite of the relatively rich metadata accompanying both corpora, such as assignment instructions and amount of previous training in academic writing (see also Swales 2014: 136). Thus, the emphasis was on more qualitative aspects of attribution, where the overall picture is that the two novice groups use attribution in similar ways, but that they have some different practices in the choice of subtypes of integral and non-integral attribution, some of which may be ascribed to L1 influence and potentially to cultural differences in attribution practice.

The comparison between the novices and expert writers in the same discipline in both English and Norwegian could shed light on the extent to which the novices had acquired the attribution practices of their discipline and indicate whether differences between the EL1 and EL2 novices might be due to L1 influence. This brought the study close to an integrated contrastive model of analysis (Gilquin 2000/2001), in which learner corpus analysis is juxtaposed with a cross-linguistic analysis. However, the analysis of the expert texts was less

systematic than that of the novice corpora (though the KIAP corpus was analysed by Fløttum et al. 2006), and it is uncertain whether the academic writing in English by Norwegian learners are influenced by expert texts in Norwegian. Clearly, the contrastive interlanguage analysis would benefit from a comparable student corpus of academic L1 Norwegian to shed more light on the way in which Norwegian academic novices use attribution and how they construct their academic texts and their academic persona more generally.

This study has not aimed to identify errors in the ways that novice EL1 and EL2 writers use academic attribution. Furthermore, given the fact that the EL1 writers are novices, the differences uncovered between the corpora should not be taken to mean that the learners should change their practices (Granger 2009: 22). Still, where the differences between VESPA and BAWE seem to reflect differences between English and Norwegian academic style, as described here and in Fløttum et al. (2006), there is a potential for applications in EAP teaching. As the Norwegian learners already master the basic principles of academic attribution in their discipline, further instruction concerns adjustments of their practice. This might pertain to a raised awareness of the intertextual requirements of the academic genre so that they engage more – or more explicitly – with the literature to "build a credible writer ethos using citation to construct factual reliability" (Hyland 2002: 115). At the micro-level it may involve attention to the form of attribution, especially the use of naming attributions with subject function as an alternative to prepositional phrases.

The variation in attribution practice has been well established in a number of previous studies. The present focus on a single academic discipline is thus a clear limitation on the generalizability of its findings. It would be instructive to compare the results obtained here to other disciplines, for example to investigate the hypothesis of cultural differences further, in the vein of Fløttum et al. (2006), but this requires a substantial expansion of the VESPA corpus. Finally, the findings of more writer visibility and less attribution in Norwegian-produced EL2 texts would be interesting to study more systematically across languages and disciplines, to gain more insight into the relationship between averral and attribution in academic texts.

Primary data

BAWE – British Academic Written English Corpus:
    http://www.coventry.ac.uk/research/research-directories/current-projects/2015/british-
    academic-written-english-corpus-bawe/

KIAP – Cultural Identity in Academic Prose: http://www.uib.no/fremmedsprak/23107/kiap-korpuset

VESPA – Varieties of English for Specific Purposes dAtabase, Norwegian component: http://www.hf.uio.no/ilos/english/services/vespa/

References

Ädel, Annelie. 2008. *What uh the folks who did this survey found*: Expert attribution in spoken academic lectures. *Nordic Journal of English Studies* 7 (3): 83-102.

Ädel, Annelie & Ute Römer. 2012. Research on advanced student writing across disciplines and levels. Introducing the Michigan Corpus of Upper-level Student Papers. *International Journal of Corpus Linguistics* 17 (1): 3-34.

Aijmer, Karin. 2002. Modality in advanced learners' written interlanguage. In *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching,* Sylviane Granger, Joseph Hung and Stephanie Petch-Tyson (eds). Amsterdam: Benjamins, 55-76.

Borg, Erik. 2000. Citation practices in academic writing. In *Patterns and Perspectives: Insights into EAP writing practice*, Paul Thompson (ed.). CALS, The University of Reading, 27-45.

Brezina, Vaclav. 2018. *Statistics in Corpus Linguistics. A Practical Guide*. Cambridge: CUP.

Carter-Thomas, Shirley & Elizabeth Rowley-Jolivet. 2013. Citation from a cross-linguistic perspective: the case of French researchers publishing in English. *ESP Across Cultures* 10: 111-126.

Charles, Maggie. 2006a. The construction of stance in reporting clauses: A cross-disciplinary study of theses. *Applied Linguistics* 27 (3): 492-518.

Charles, Maggie. 2006b. Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes* 25, 310-331.

Doccherty, Peter. 2018. An exploratory study in the use of direct quotation by L2 entry level Bachelor students. *Journal of English for Academic Purposes* 40: 36-40.

Ebeling, Signe O. & Alois Heuboeck. 2007. Encoding document information in a corpus of student writing: the British Academic Written English corpus. *Corpora* 2 (2): 241-256.

Fløttum, Kjersti 2003. Bibliographical references and polyphony in research articles. In *Academic discourse. Multidisciplinary approaches,* Kjersti Fløttum & François Rastier (eds). Oslo: Novus, 97-119.

Fløttum, Kjersti, Trine Dahl & Torodd Kinn. 2006. *Academic Voices*. Amsterdam: John Benjamins.

Gilquin, Gaëtanelle. 2000/2001. The integrated contrastive model. Spicing up your data. *Languages in Contrast* 3 (1): 95-123.

Granger, Sylviane. 2009. The contribution of learner corpora to second language acquisition and foreign language teaching: A critical evaluation. In *Corpora and Language Teaching,* Karin Aijmer (ed.). Amsterdam / Philadelphia: Benjamins, 13-32.

Granger, Sylviane. 2015. Contrastive Interlanguage Analysis. A Reappraisal. *International Journal of Learner Corpus Research* 1 (1): 7-24.

Groom, Nicholas. 2000. Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. In *Patterns and Perspectives: Insights into EAP Writing Practice,* Paul Thompson (ed.). Reading: University of Reading, 14-25.

Halliday, M.A.K. 1994. *An Introduction to Functional Grammar*. London: Arnold.

Hasselgård, Hilde. 2009. Thematic choice and expressions of stance in English argumentative texts by Norwegian learners. In *Corpora and Language Teaching*, Karin Aijmer (ed.). Amsterdam: Benjamins, 121-139.

Hasselgård, Hilde. 2017. Stating the obvious: signals of shared knowledge in Norwegian-produced academic English. In *Language, Learners and Levels: Progression and Variation*, Pieter de Haan, Rina De Vries & Sanne Van Vuuren (eds). Presses universitaires de Louvain, 23-44

Heuboeck, Alois, Jasper Holmes & Hilary Nesi. 2010. The BAWE Corpus Manual, version 3. http://www.coventry.ac.uk/Global/08%20New%20Research%20Section/Current%20Projects/BAWEmanual%20v3.pdf

Hunston, Susan. 1995. A corpus study of some English verbs of attribution. *Functions of Language* 2 (2): 133-158.

Hyland, Ken. 1999. Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics* 20 (3): 341-367.

Hyland, Ken. 2000. *Disciplinary Discourses. Social interactions in academic writing*. Ann Arbor: The University of Michigan Press.

Hyland, Ken. 2002. Activity and evaluation: Reporting practices in academic writing. In *Academic Discourse,* John Flowerdew (ed.). Harlow: Longman, 115-130.

Hyland, Ken & Feng Jiang. 2019. Points of reference. Changing patterns of academic citation. *Applied Linguistics* 40 (1): 64-85.

Hyland, Theresa Ann. 2009. Drawing a line in the sand: Identifying the borderzone between self and other in EL1 and EL2 citation practices. *Assessing Writing* 14: 62-74.

Larsson, Tove. 2017. *The importance of*, *it is important that* or *importantly*? The use of morphologically related stance markers in learner and expert writing. *International Journal of Corpus Linguistics* 22 (1): 57-84.

Li, Shuguang & Klaus-Uwe. Panther. 2014. 'Author (date)' constructions in academic discourse. *English Text Construction* 7 (2): 215-248.

Nesi, Hilary & Sheena Gardner. 2012. *Genres across the Disciplines. Student Writing in Higher Education*. Cambridge: Cambridge University Press.

Mansourizadeh, Kobra & Ummul K. Ahmad. 2011. Citation practices among non-native expert and novice scientific writers. *Journal of English for Academic Purposes* 10 (3): 152-161.

Nesi, Hilary. 2021. Sources for courses: Metadiscourse and the role of citation in student writing. *Lingua* 253, 103040. https://doi.org/10.1016/j.lingua.2021103040

Paquot, Magali, Hilde Hasselgård & Signe O. Ebeling. 2013. Writer/reader visibility in learner writing across genres: A comparison of the French and Norwegian components of the ICLE and VESPA learner corpora. In *Twenty Years of Learner Corpus Research: Looking back, Moving ahead,* Sylviane Granger, Gaëtanelle Gilquin & Fanny Meunier (eds). Louvain: Presses Universitaires de Louvain, 377-387.

Parkinson, Jean & Jill Musgrave. 2014. Development of noun phrase complexity in the writing of English for Academic Purposes students. *Journal of English for Academic Purposes* 14: 48-59.

Pecorari, Diane 2006. Visible and occluded citation features in postgraduate second-language writing. *English for Specific Purposes* 25: 4-29.

Pecorari, Diane & Philip Shaw. 2012. Types of student intertextuality and faculty attitudes. *Journal of Second Language Writing* 21: 149-164.

Petrić, Bojana 2007. Rhetorical functions of citations in high- and low-rated master's theses. *Journal of English for Academic Purposes* 6 (3), 238-253.

Petrić, Bojana & Nigel Harwood. 2013. Task requirements, task representation, and self-reported citation functions: An exploratory study of a successful L2 student's writing. *Journal of English for Academic Purposes* 12: 110-124.

Scott, Mike. 2012. WordSmith Tools version 6, Stroud: Lexical Analysis Software.

Sinclair, John M. 1986. Fictional worlds. In *Talking about Text* (Discourse Analysis Monographs No. 13)*,* M. Coulthard (ed.). Birmingham: English Language Research, University of Birmingham, 43-60.

Swales, John M. 1990. *Genre Analysis. English in Academic and Research Settings*. Cambridge: Cambridge University Press.

Swales, John M. 2014.Variation in citational practice in a corpus of student biology papers: From parenthetical plonking to intertextual storytelling. *Written Communication* 3 (1): 118-141.

Tadros, Angela. 1993. The pragmatics of text averral and attribution in academic texts. In *Data, Description, Discourse*, Michael Hoey (ed.). London: Harper Collins. 98-114.

Thompson, Geoff & Yiyun Ye. 1991. Evaluation in the reporting verbs used in academic papers. *Applied Linguistics* 12 (4): 365-382.

Thompson, Paul & Chris Tribble. 2001. Looking at citations: using corpora in English for academic purposes. *Language Learning & Technology* 5 (3): 91-105.

Thompson, Paul. 2005. Aspects of identification and position in intertextual reference in PhD theses. In *Strategies in Academic Discourse*, Elena Tognini-Bonelli & Gabriella Del Lungo Camiciotti (eds). Amsterdam: Benjamins, 31-50.

Verheijen, Lieke. 2015. The language of quoting in academic writing. *Dutch Journal of Applied Linguistics* 4 (1): 101-121.

Wiemeyer, Leonie. 2019. Direct quotation in second language writing. A corpus-based study of intertextuality in academic learner English. In *Learner Corpora and Language Teaching*, Sandra Götz & Joybrato Mukherjee (eds). Amsterdam/Philadelphia: Benjamins, 129-156.