

## A corpus linguistics sandwich Learners chewing over reporting verbs in academic writing

By Silvia Bernardini and Andy Cresswell (Università di Bologna, Italy)

### Abstract & Keywords

#### English:

Our lessons form part of a module on corpus linguistics from the Master's in Specialized Translation at UNIBO. The lessons are a "sandwich" because the central content, citations expressed through projecting clauses, is approached from the twin perspectives of the two module tracks, language and linguistics. In the language lesson, learners shadow the teacher's exemplification of citation functions, then "classify" corpus data, matching citation functions and examples, and generate functional descriptions. In the linguistics lesson, they apply the knowledge of citation acquired in the language lesson to pursue a corpus-based comparison of two language varieties (native and lingua franca English). Overall, students found that the sandwich was challenging and required critical, autonomous thinking; but arguably this is precisely what is required of future professional translators.

**Keywords:** Corpus analysis, translator education, academic writing, English as a Lingua Franca, indicatori di citazione, inglese lingua franca, scrittura accademica, ormazione dei traduttori, analisi di corpora, citation verbs

---

### 0. Foretaste

All those who teach are learners as well, and those who teach (with) corpus linguistics all the more so. We believe that our role is to guide by walking the road with our learners for a while, before progressively letting them lead the way. We try to teach by example, by reflecting on our own learning experience and distilling it for ourselves and for them. We try to share our enthusiasm and our failures, more than our knowledge. We are not always successful, but we keep trying. All this, we learnt from Guy.

### 1. A balanced and healthy diet? The context for our corpus linguistics "sandwich"

Our two lessons, which together form a corpus linguistics sandwich, are from the Corpus Linguistics module of the International Master's in Specialized Translation run by the Department of Interpreting and Translation of the University of Bologna at Forlì. The module takes place over 10 weeks in the first year, first semester of the Master's. It has two tracks: "linguistics" (two 2-hour sessions weekly), and "language" (one 2-hour session weekly). There are two groups for the linguistics track (of about 30 students) and three for the language track (of about 20); all classes are held in computer labs where each student has access to a computer.

The "linguistics" track focuses on corpus linguistics principles and corpus analysis skills. It introduces basic concepts and a variety of corpus resources relevant to translation students, and creates opportunities for learners to apply them to language (learning) and translation (studies) problems, first in guided activities, and then in increasingly self-directed and open-ended ones. The issues and approaches range from concordance analysis carried out on paper and aiming to describe a single unit of meaning in general English (along the lines of Sinclair's (2003) *Reading Concordances*), to the setting up of comparable (sub)corpora and the observation of typical features of the language used by different speakers (e.g., Kennedy vs. Nixon in their 1960's "Great Debates") or different groups of speakers (e.g., native English vs. English as a Lingua Franca (ELF) speakers).

The "language" track aims to develop learners' academic writing abilities, using corpus materials and methods, in the context of a single genre, the research article in applied corpus linguistics. The students are told that this single focus has two aims: on the one hand, it serves as a case study on genre awareness for future translators, who will have to tailor their lexicogrammatical choices to a working notion of the appropriate target language genre; on the other, it provides relevant experience for their end-of-course assessment, which consists of an academic essay on a corpus linguistics topic, illustrated with original corpus evidence.

As can be inferred from this sketchy outline, the module has a rather composite and ambitious set of purposes. First, we would like our learners to "learn about" and to "learn to exploit" corpora (Leech, 1997). Second, our aim is to develop awareness and skills for conducting linguistic analysis and for reporting on it in academic English. In recent years we have come across ever fewer students with a research orientation. The strong focus on the transmission of market-ready skills in translation departments means that limited attention is nowadays devoted to the enhancement of research skills. We believe this is wrong for at least two professionally-oriented reasons: first, because research education enhances critical thinking, which in turn "will prepare students to make well-founded decisions and choices in their [...] careers" (Mitchell-Schuitevoerder, 2014: 241); and second, because it contributes to establishing the status of the translation profession as one "whose members are competent and recognized academically" (Vandepitte, 2013: 144-145).

### 2. Sandwich ingredients and preparation

The lessons we describe in this contribution focus on in-text citations, aiming to develop learners' knowledge of how language choices reflect pragmatic ones in academic writing. The initial motivation therefore comes from the language track, but the topic is also tackled, with a more explicit research orientation, within the linguistics track. The topic is restricted to one specific structure (projecting clauses), and one specific pragmatic aspect (the

pragmatic implications of different citation verbs). We chose the projecting clause because it is a central structure of integral citation and also, more prosaically, because the structure is particularly easy to look for in a corpus. The focus on the evaluative role played by verbs used for citation is motivated by the existence of functional descriptions that could be used as a starting point for corpus explorations (Hyland, 2002; Thompson & Ye, 1991).

In the language track, students work in pairs using AntConc (Anthony, 2014) with CRANE (Cresswell, 2013), a 65,000 word corpus of non-empirical research articles in the Social Sciences. The articles focus exclusively on previous research and/or theories, thus maximising the space given to citation. When CRANE does not provide enough evidence, it is supplemented by the downloadable, untagged version of the BAWE corpus. BAWE (Nesi, 2011) is a collection of assignments by students at British universities, containing 6.5 million words from a variety of disciplines and genres.

In the linguistics track, the annotated version of BAWE is accessed via the open Sketch Engine platform (<https://the.sketchengine.co.uk/open/>). For this lesson, information about students' first languages available in BAWE is exploited to define subcorpora of assignments by native English speakers and ELF speakers. These are then compared to identify quantitative differences in the use of citation verbs across the two groups. We consider students in BAWE as ELF writers, rather than language learners, since in the majority of cases they are not language students, and since these assignments are not evaluated based on the English proficiency of their authors.

The two lessons in the sandwich are informed by data-driven learning (DDL, Johns, 1991). In DDL, learners examine concordances and use multiple examples of authentic language to make generalisations, an approach that requires and fosters learner autonomy and creativity, but that can be demanding and make learners feel overwhelmed by data that is too abundant (Hafner & Candlin 2007:315), or by “the complexity and fuzziness of authentic data” (Boulton, 2009:41). Collaboration by learners working in pairs or small groups and teacher guidance (Yunus, 2017:143) can counteract these problems. Such a pedagogical approach reflects Vygotsky's social-constructivist theory of learning (Vygotsky 1978): learning occurs at a “zone of proximal development”, which is beyond the capacity of a learner working alone, but reachable with the aid of peers or competent adults.

The learner can be brought to this point through “scaffolding”, or guided support. Boulton (2010: 18) describes his version of scaffolding as follows: “rather than imposing hands-on DDL on the assumption that “teacher knows best”, a gentle lead-in would seem desirable [...], from pre-set exercises to more open-ended exploration”. The principle of the gentle lead-in is followed at two different, interconnected levels: first, because in the language lesson controlled exercises precede open exploration, and second, because the language lesson prepares the learners for their research-oriented exploration in the linguistics lesson.

In the language lesson before the one described here, learners are instructed on the relationship between citation structures, functions and interpersonal aspects of academic writing, with the teacher demonstrating how concordances can afford examples, and the students following on their own computers. Working in pairs, students then do semi-controlled exercises in which they rewrite the syntax of citation sentences. Focusing on reporting verbs which seem to be overused by Italian-speaking students (*affirm, analyse/analyze, prove, say, sustain, underline*), the teacher subsequently leads students in a search for these verbs in CRANE, using AntConc in order to demonstrate their absence or rarity in expert academic written discourse. Learners are then shown a list of 104 citation verbs that can be used as alternatives to overused verbs. These activities are all deductive, in the sense that information, principles and methods are provided by the teacher and the students learn by applying them.

### 3. The language slice: investigating citation functions in projecting clauses

#### 3.1. Introduction

The lesson follows the classic sequence modelled by Tim Johns (1991) of observe – classify – generalise. In this sequence, the “observe” and “classify” stages constitute further scaffolding, which makes the inductive data-driven learning in the “generalise” stage more manageable for the learners.

#### 3.2. Phase 1: observe

The first phase of the lesson, corresponding to Johns' “observe” stage, is a continuation of pre-DDL scaffolding, and is essentially deductive. In terms of language, the lesson begins with the teacher revising the projecting clause structure, deconstructing it as: subject (=the cited author) + publication date (with page number optional) + citation verb + *that* + author's cited views or information, for example “Hyltenstam and Abrahamsson note correctly that current versions of the CP do not make claims about speed of acquisition” (Marinova-Todd *et al.*, 2001). In terms of developing learners' query building skills, the structural deconstruction shows learners how understanding the relationship between lexis and structure is necessary when corpora are not annotated for the latter.

The “observe” stage scaffolding instruction extends the learners' experience of functions of AntConc (the learners have already made concordances and used the “Advanced” functions of “context words” and “horizons”, which respectively highlight instances of co-occurring words, and set an upper limit in word numbers on the distance between these words). Here, the principal skill extension is learning how to upload files containing lists of search terms. Learners download from the module Moodle page a list of reporting verbs to search for. Next, they open a document (also on the module page) containing precise instructions for uploading the research terms file to AntConc and searching (fig.1). Precise written instructions are necessary in concordance-focused lessons, to avoid the risk of a situation in which instructive dialogue oriented towards language learning is constantly submerged by requests for one to one demonstrations of how to get the software to work. Eventually, instructions can be made less detailed when the number of technologically challenged learners becomes low enough for peer support to be sufficient.

STUDENT A, STUDENT B, IN PAIRS.

1. Download the file called REPVERBS\_updated from Moodle. Save it on the Desktop.
2. Open AntConc.
3. Drag down the FILE menu, click on OPEN DIR.
4. Scroll through the directories till you find TEXTS/T.

5. Student A. Open TEXTS/T and select CRANE\_untagged. Click OK. Student B. Open TEXTS/T and select BAWE\_TXT. Click OK.
6. Click on ADVANCED.
7. Click on LOAD FILE.
8. Navigate to the Desktop and double click on REPVERBS\_updated.
9. Tick USE SEARCH TERMS FROM LIST BELOW.
10. Click APPLY.
11. A ONLY. Tick USE CONTEXT WORDS AND HORIZONS.
12. A ONLY. In CONTEXT WORDS, type *that*.
13. A ONLY. Click ADD.
14. A ONLY. In the CONTEXT HORIZON field, select 0/2R. Click APPLY.
15. A and B. Click START to get a concordance of selected reporting verbs.
16. A and B. Tick KWIC SORT and set as follows. Level 1 0 Level 2 1R Level 3 2R

**Fig. 1: Student search instructions (citation verbs)**

By the time they have followed all the instructions in fig. 1, each pair of students is able to view in AntConc two alphabetical lists of citation verb forms in context, retrieved from CRANE and BAWE respectively. The teacher then guides the students through a four-step procedure demonstrating the pragmatic functions of citation verbs using examples from CRANE, or, if there were no examples of a particular function in CRANE, from BAWE. The citation functions (listed in the first column of tab. 1) vary along a cline of “factivity” (Hyland, 2002), from factive (1) to counter-factive (10).

The observation of citation functions consists of four steps. First, the learners download a document showing ten citation verb functions (tab. 1, column 1). Second, for each function, the teacher directs the learners to examples by giving them a citation verb to use as a search term. Once the learners have concordances of this verb, they are given a key phrase (tab.1, column 2), which helps them find the line pre-selected by the teacher as an example. Third, each example is viewed with more context, as the learners double-click on the highlighted search term (or “keyword”), to show the full text. Two or three sentences of this wider context (pre-selected by the teacher) are read aloud by a student. Fourth, the teacher links the *pragmatic function* of the example (that is, the writer’s attitude to the cited view or information), to the potential *generic effect*, in terms of the orientation which the writer projects towards the academic discourse community through her/his choice of a particular verb.

CITATION FUNCTION	KEY PHRASE (keyword in italics)
1. Writer expresses his/her own assurance about the cited finding/information/view (because it supports his/her own argument).	The authors <i>argue</i>
2. Author's positive attitude: writer reports author as positive towards information/opinions author reports (writer may or may not view these opinions positively)	Model <i>pointing out</i>
3. Writer neutrally informs readers of the author's stated views/information.	<i>states</i> that, 'I have earlier
4. Research findings – non-factive verbs: writer adopts no clear attitude to the reported findings, or is neutral.	<i>found</i> that the learners
5. Writer neutrally informs readers of how reported information/opinions fit into the cited text (Thompson & Ye, 1991: 372).	<i>adds</i> that when people
6. Research procedures – verbs that are always neutral in attitude (Hyland, 2002: 119). These verbs are not generally used in the projecting clause structure with <i>that</i> .	Belz <i>examined</i> a 100,000-word (BAWE)
7. Author's tentative attitude – Tentative Cognition Verbs: writer is neutral, and reports the cited author as feeling a degree of caution about the views/information the author is reporting.	he <i>assumes</i> that 'all
8. Writer expresses own doubt about the cited information/views (tentative doubt) – the writer by no means rejects the view/information cited, but the tentativeness leaves space to later imply that there is room for improvement or development.	<i>claim</i> that these near-native
9. Writer implies his/her disapproval of the cited information/views indirectly, by presenting a negative view of the way the author presents the view/information	responded by <i>contending</i> that
10. Writer's direct criticism	(i) <i>does not mention</i> this (ii) <i>fails to account</i>

**Tab. 1: Main functions of reporting verbs in citation (with key phrases)**

To give an example, let us take function 8, “Writers expressing their own doubt about the cited information/views”, which Hyland (2002: 121) summarises as “tentative doubt”. The full co-text from CRANE is as shown in (1) below, with the keyword underlined.

(1) Further, Hyltenstam and Abrahamsson claim that these near-native speakers should be differentiated from the native speakers because ‘their L2 speaker background can be identified only when their L2 performance is scrutinized in detailed linguistic analyses.’ (Marinova-Todd *et al.*, 2001).

The generic effect in this case can be accounted for as follows. (i) To demonstrate disciplinary discourse community solidarity, the writer by no means rejects the view/information cited. (ii) On the other hand, the tentativeness implied by the choice of the verb *claim*, a non-factive verb, leaves the writer space to later imply that there is room for improvement or development, in order to justify her/his own research.

Finally, questions are taken, and the procedure is repeated for the other nine functions.

At the end of the “observe” phase of the lesson, the learners have observed ten citation functions, made ten concordances of citation verbs, related each of these ten verbs to their appropriate citation function, looked up the fuller context in each source text, and made sense of the fuller context by reading it aloud or by listening to another student read it. Through these procedures they have consciously experienced three ways in which concordances can be read – first, *paradigmatically*, by scanning down the concordance to look for a particular example; second, *syntagmatically*, by reading along the individual concordance line when given the key phrase; and third, *textually*, by looking at the search word set within the fuller context of several sentences in its original text.

### 3.3. Phase 2: Classify

The second, “classify” stage of the lesson is semi-autonomous. The functional descriptions and examples have been chosen by the teacher, and the outcome is predetermined, but the learners work independently in pairs without step-by-step direction.

Learners download the document containing the exercise. They follow instructions on loading CRANE into AntConc, and on making a concordance of projecting clauses with verbs in the third person singular of the present simple tense. So *\*s* is the search term, with context word *that*, horizons are set at 0/1R, and sort is set to Level 1 0/ Level 2 2L /Level 3 0. The instructions contain a warning that the concordance will contain plenty of lines that are not projecting clauses. This reminds students that using concordances implies critical reading.

Learners read a table which gives a list of verbs together with their citation functions. These functions correspond to the functions already outlined, but include additional detail that reflects the actual pragmatic contexts in texts that are written to communicate reasoned arguments. The learners' task is to read the concordance lines featuring that verb and identify which line exemplifies the detailed function. They then copy from the original text enough co-text to serve as a reference example for future study, and paste it into the table alongside the functional description, with instructor support as needed. Tab. 2 shows a functional description with a retrieved reference example pasted in by the learners.

SUBJECT	VERB	FUNCTION	RETRIEVED EXAMPLE
author	argues	The writer wants the reader to accept the author's view as reasoned because it supports the writer's own view	al-'Aqqad argues that Islam as traditionally understood is quite compatible with democracy as it is understood in the twentieth century world (Goddard 2002, in CRANE)

**Tab. 2: Citation verb functions: classification exercise**

By the time the learners have finished this “classify” activity, they are prepared for the final, “generate” stage of the citation functions lesson.

### 3.4. Phase 3: Generate

In this phase, which involves data-driven learning in the strictest sense, the learners create a concordance of a given citation verb, and arrive at their own functional descriptions. The task proposed is a largely autonomous exercise, though the learners work collaboratively in pairs, with the instructor ready to help.

Students are instructed to arrive first at a general functional account, and then to agree on an exact functional account. The general functional account is to be arrived at paradigmatically, by reading all the concordance lines for a given citation verb. The exact functional account must take into account the larger context of a single occurrence, viewed syntagmatically in the original text by double-clicking on the keyword. A downloadable template is provided, into which learners are asked to type their functional descriptions. They then upload them to a dialogic forum set up for the purpose on the Moodle learning platform. The descriptions remain available online in the forum for other groups investigating the same verb, for purposes of comparison and peer evaluation.

Here we present some of the learners' functional descriptions in order to illustrate the immediate outcome of the “generate” phase of the lesson. We have chosen to focus on functional descriptions that feature citation verbs whose functions had not already been encountered in the observation and classification phases. Hence we can reasonably attribute the learners' functional descriptions to autonomous data-driven learning, rather than to the information the teacher supplied about functions that was communicated in the demonstration and classify phases. According to Hyland (2002:119-121) and Thompson & Ye (1991), the verbs *note*, *report*, *show* and *warn* all fall into the general functional category that can be summarised as “assurance verbs”. This category of verbs can be used either factively or less so, “to pass on information without interpretation” (Hyland, 2002:121). These functions correspond to numbers 1 to 3 in tab. 1. Among the seven pairs or small groups of learners who studied concordances of these verbs, three produced functional descriptions comparable to the functional description of factive use (no. 1 in tab. 1). An example is shown in tab. 3.

Verb	Report
------	--------

General functional description	Neutral reporting: writer expressing their own assurance about the cited information because it supports their own argument - Factive verbs
Example occurrence	The first indication of language delay in an infant with Downs Syndrome is their delayed onset of canonical babbling. Oller (1986 as in Tager-Flusberg, 1999: 313) reported that infants who had been diagnosed with Downs Syndrome began babbling approximately two months after the control group of typically developing children. (BAWE)
Exact functional description	The writer neutrally reports a piece of information from a (sic) previous research supporting his own argument

**Tab. 3: Learner account of report**

A further four pairs/small groups arrived at descriptions comparable to functions 1 or 2 in tab. 1 (or both), sometimes after considering peer feedback. The exchange in tab. 4 shows how use of the online forum permitted extended peer dialogue to help bring about a more accurate outcome. Interestingly, the dialogue between the learners reflects the ambiguity in Hyland's account, which attributes functions of both positive and neutral evaluation to the same verbs.

Verb	Show
General functional description	The writer informs the readers about the author's stated opinion or findings, keeping a neutral or slightly positive attitude.
Example occurrence	1. Rosch showed clearly that humans do not regard all items within a category as equal, instead, they rank some as being better than others, in the sense of being more typical examples. (Aitchison, 1993, in CRANE) 2. Simons and Keil (1995), and Gelman and Wellman (1991), show that four and five year old children understand the differences that exist in how animate and inanimate objects are supposed to look on the inside as opposed to the outside. (BAWE)
Exact functional description	In the first example the writer presents the author's findings in a slightly positive way; while in the second example the writer neutrally reports the authors' stated view.
Reply by another pair	Unfortunately, we happen to disagree with the general function description of the verb <i>show</i> , as in our opinion the verb is connected with a positive evaluation of the statement which follows. In fact the writer is committing to the truth of it by presenting evidence of its reliability and opposing it with another one previously presented.

**Tab. 4: Learner account of show (with peer feedback)**

A further example of the role of interaction through the online forum is shown in the exchange about the verb *warn* (tab. 5).

Verb	Warn
General functional description	Research findings – non-factive verbs - writers adopt no clear attitude to the reported findings, or are neutral.
Example occurrence	The growing consumer ideology in health care is giving patients increasing choice, rights, and opportunities to be involved in decision making regarding care (Hinchliff <i>et al.</i> , 1998). Fulford <i>et al.</i> (1996 p151) warn that 'An unthinking acceptance of patients' rights is dangerous, because introducing the wrong sort of rights would be as damaging to patients as continuing to ignore their rights altogether.' They also state that, unfortunately, in health care 'patient rights are more rhetoric than reality' (Fulford <i>et al.</i> , 1996 p152). (BAWE)
Exact functional description	The verb <i>warn</i> has generally a negative semantic prosody, as it often co-occurs with negative words such as "unthinking", "dangerous", "wrong". In this example, the subject of the verb is specific (Fulford <i>et al.</i> ) and the writer has a neutral attitude to the authors' argument
Comment by another pair	"To warn" seems to be a neutral reporting verb. However, in our research we found that it is often used by the writer to emphasize the negative consequences of the issue the author is discussing. In other words, the writer agrees in considering the given issue as negative.
Learners'	We examined both examples and we concluded that you are right.

**Tab. 5: Learner account of *warn* (with peer feedback and reply)**

The learners' general account of the pragmatics, obtained paradigmatically, corresponds to function 3, but the close examination of the specific example, examined syntagmatically, shows a more original learning development, which illustrates the meaning of the "generate" stage of data-driven learning. To paraphrase the learners' account of the pragmatics, *warn* is used when the author's negative view is cited because it supports the writer's negative view of the same view or fact. This functional description corresponds plausibly to the cited example, and it arguably corresponds to function 1 in tab. 1. But the learners' description is pedagogically clearer. This is because Hyland does not mention the negative counterpart of function 1 – writers citing authors because the authors represent the writer's disapproval of a view or fact. The learners seem to be thinking critically and autonomously, to be adapting functional descriptions to the needs of the co-text, a development in a functional sense of Gavioli & Aston's (2001) "discourse authentication".

#### **4. The linguistics slice: comparing ELF and native use of citation verbs**

The lesson on citation verbs from the linguistics track of the module is designed to take place after the corresponding language lesson, and makes use of the substantial scaffolding it provides. Knowledge about citation verb functions and categorisations, and experience with corpus analysis in this specific area (i.e., query design for retrieving citations, inferring and generalizing from the retrieved concordances) are presented as particularly important, and often referred to. In this way it is hoped that the learners will be able to appreciate the close ties between the two tracks of the module, a feature not always apparent in linguistics-with-language modules in Italian universities.

A further advantage of this sequencing is that it provides a context for the learners to "play the researcher". Given the many requirements and the limited time available, a corpus linguistics module whose aims include introducing learners to empirical research methods is faced with the dilemma of either tackling several trivial language issues, or focusing on a single complex one in detail. Although in the course described we have chosen breadth over depth, the sandwich lesson on citation contributes to counterbalancing this tendency, offering a more plausible example of "proper" corpus analysis than is otherwise possible.

In previous lessons in the linguistics track, most of the corpus analyses carried out by the students focus on a specific word or set of words investigated in a single corpus, with attention dedicated, along the lines of Sinclair (1996), to looking at collocates, colligates, semantic preferences and semantic prosodies. In this lesson the focus shifts from language features to describing a language variety. The aim is to compare the variety with another one that acts as a baseline and that differs from it (ideally) only with respect to the variable under study. This is a more abstract problem than any the students have faced before, and one that is closer to the local concerns of a researcher, rather than to the local concerns of a translation or language student.

After an introduction to contrastive interlanguage analysis (Granger, 2015) and a presentation of some learner corpus resources, the learners read the first section of the classic Granger study (1998) on amplifier collocations in native and learner English. Granger finds that learners use fewer collocations than native speakers, and that when collocations are used, they are likely to be transferred from the learners' native language (French), or include general words, unrestricted in their collocational behaviour, such as "very". Granger (1998) thus provides a set of initial questions/hypotheses for this inquiry-based lesson, in which students explore whether similar differences can also be observed (a) in a corpus like BAWE – which includes ELF writing by students from several language backgrounds – and (b) in the use of citation verbs.

The basic method can be summarised through the following description of how the lesson was actually taught in 2017/18. First, after a discussion of ways of reducing evidence to manageable quantities, the teacher and learners decided to limit the analysis to the Arts and Humanities (AH) assignments of BAWE, and to active structures only. Learners were allowed to experiment with queries until they were satisfied with the concordances retrieved. They were informed that if faced with a choice between precision and recall, precision was to be favoured, given the quantity of evidence, the need to otherwise perform tedious manual cleaning, and the general principle that, when comparing two language varieties, completeness is less important than unbiased-ness. In other words, they did not necessarily need to see the whole picture, provided they were confident that the differences observed were reliable. In the end, the learners settled on a very simple query – a proper noun followed by a date, followed by a verb; this search allowed for no internal variation but returned a sufficiently precise and manageable amount of evidence. Working in pairs, they performed parallel queries on native-written AH assignments and ELF-written AH assignments, saved the concordances to a text file, removed false positives and identified the complete citation verbs. Several decisions had to be made; for instance, in a case like "Leitner (1992) agrees with this method writing", learners had to decide whether the citation verb was "agree", "write", or both.

Once the cleaning was done, learners were instructed to produce two lists, one of verb tokens, and one of verb types. To obtain the latter, the token list was sorted, and repetitions were removed, partly using utilities in text editors, partly through manual removal of inflected forms and other variants (a sort of manual lemmatisation).

Before proceeding with the counting and normalisation, some potential sources of bias were identified. For instance, the second most frequent citation verb in the native subcorpus is "write", a verb that is totally absent from the ELF results. Further scrutiny showed that 8 out of 14 occurrences of "write" came from a single text; this fact brought home to the students the need to carefully analyse data to limit the impact of single files and single authors.

After recording the raw number of types and tokens, the learners' attention was drawn to the difficulty in comparing these numbers, given the different sizes of the two subcorpora. After normalisation, it became clear that almost twice as many citation verb tokens had been retrieved from the ELF subcorpus as from the native subcorpus (21.1 vs. 12.7 per 100,000 words, tab. 6). While this result is difficult to interpret without further analysis (native speakers may be citing less, or may be using other structures, e.g. passive ones), it does serve as a basis for evaluating the (much smaller) difference in terms of types (16.5 vs. 14.7 per 10,000 words, tab. 6).

ELF students in BAWE seem to cite more using the pattern we searched for in the corpus, but use only a slightly higher number of different verbs for this purpose. It could be hypothesised that ELF writers favour general words, as was the case in Granger’s (1998) data. Partial support for this hypothesis comes from the presence, in the ELF token list, of general verbs such as “give”. This verb occurs four times in the ELF type list (“gives a better solution”, “gives an example”, “gives an interesting argument”, “gives more detailed explanation”), but never in the native one. Browsing the latter, one finds instead several cases in which two citation verbs are used together, allowing the writer to express her views about the cited work in a more precise manner (“went even further, arguing”, “agrees with this method writing”, “retaliates by criticising”, “expand on this by suggesting”). This strategy, as students were able to observe, is virtually absent from the ELF subcorpus.

	ELF sub-corpus		Native sub-corpus			
	Raw	Normalised	N. words	Raw	Normalised	N. words
Tokens	59	21.1 (per 100K words)	278,606	200	12.7 (per 100K words)	1,571, 762
Types	34	16.5 (per 10K words)	20,493	80	14.7 (per 10K words)	54,333

**Tab. 6: Raw and normalised frequencies of citation verb types and tokens in the native and ELF subcorpora of BAWE**

In concluding the lesson, it was pointed out that the activities proposed were only a starting point, and that further work could be conducted to investigate, for example, the typical co-textual patterns of the most frequently used verb types, the preference for a given category of verbs (e.g., *real world*, *cognition* or *discourse*, following Thompson & Ye’s (1991) taxonomy), or the presence of other citation patterns (such as those targeted by Nesi, 2013). Finally, the topic of the following lesson was introduced, namely the ways in which quantitative data such as those gathered in this lesson on citation verbs could be represented graphically (e.g. through bar plots) and tested for significance using the  $\chi^2$  statistic in MS Excel.

### 5. Food for thought

In this contribution we have described how a single object of linguistic analysis, i.e. citation verbs in academic English, can be tackled from both a data-driven language learning perspective and a more research-oriented perspective, within a single module on corpus linguistics run on parallel tracks.

In terms of student evaluation, at the time of writing we have limited feedback, mainly in the form of interaction in class and in the forum. The general impression we have is that the corpus linguistics part of the sandwich was rather challenging for these learners, who had no previous experience of empirical linguistic research. They seemed to react to the language learning lesson (more) positively, first through intensive concentration on reading the concordances and individual examples, then through committed discussions of the functional descriptions they had generated, discussions which sometimes became animated when learners evaluated the descriptions produced by other pairs/small groups. Overall, the impression was that the lesson scored high in terms of motivation.

We do not yet know, however, whether learners’ citation strategies have improved as a result of the substantial work done on this topic in the language lesson. Longitudinal data on learners’ use of logical connectors in Cresswell (2007: 282) shows that information about language is retained better when it has been acquired through the detailed and focused investigation of multiple concordance examples. Of course, acquisition in a productive sense of the meanings and functions of the citation verbs investigated is not guaranteed, but findings suggest that the investigative activity may increase the possibility of acquisition (Cresswell, 2007). We intend to focus specifically on citation practices when correcting end-of-course assignments.

Concerning the linguistics lesson, we hope that the research activity on citation verbs in ELF vs. native English academic writing has resulted in improved corpus comparison skills. To evaluate this, we will compare the original research activity reported in the end-of-course assignments with those from the previous student cohort. These assignments constitute our own corpus of learner academic writings (CLAWS), which, with the learners’ permission, we intend to expand yearly.

In terms of self-evaluation, in hindsight and time permitting, we would have included, in either the language or the linguistics lesson, activities involving direct learner access to the CLAWS corpus. As well as providing us with a resource for evaluating our own teaching, we believe that this corpus may be a useful “local” addition to the set of corpora available to our students, and one that they may find it easier to relate to, following Seidlhofer’s (2002:220) suggestion that “[foreign language] pedagogy, and presumably any pedagogy, has to be local, designed for specific learners and settings”.

### 6. Post-prandial musings: why we did what we did

The two lessons we have described in this contribution exemplify our attempt at offering Master’s students of translation practical experience of corpus work, not only for language learning and translation practice, but also for research applications. We are fully aware that few if any of our students will go on to become full-fledged corpus linguists, but we are convinced that research skills will make them not only better, but also more satisfied professional translators. Quoting from Kiraly (2000: 182):

[t]hrough our very teaching methods, we language teachers demonstrate to our students our own understanding of how language works. If we teach language as a set of artefacts, and translation skills as objectifiable, transmittable strategies, we can expect our students to develop a translator’s self-concept that sees their role as that of insignificant bilingual scribes, mechanically transcoding from one language into another.

We believe that the ambitious aims we set ourselves in this module can only be achieved if the linguistics and language tracks, and ideally other modules as well, reinforce each other. The citation unit we describe here was language-oriented, but offered the necessary context for a linguistics-oriented activity. At the same time, reflection on units of meaning and practice identifying their constituents, as experienced by the learners earlier in the linguistics track, provided essential background knowledge that they applied to the functional descriptions shown in tab. 3 to 5. Similarly, knowledge of regular expressions and familiarity with the basic functions of concordancers could be assumed (though the assumption did not always prove correct), since they are covered in a concurrent module on information mining and terminology for translators.

We see the “sandwich” presented here as an attempt at working toward “aligned” or coordinated learning, which Kelly (2005:78) suggests is needed for the development of those less content- and more process-oriented competences “that would never constitute individual modules on a [translator] training programme, so generic or cross-curricular are they in nature”. But the approach has the further, more local advantage of allowing us to downplay technical aspects and focus instead on analytical challenges that require and foster critical thought. This remains, we would suggest, the critical challenge in teaching (language with) corpus linguistics, or any other approach to empirical language study. Most of the difficulties we experienced occurred because the activities we designed for our students required “a level of analytical skill and attention to detail which [some of them] had simply not yet acquired” (Braun 2007: 323).

The currently prevailing professional orientation of quality translation Master’s (see e.g. the strong focus on translation provision competences within the EMT competence framework(s), Toudic & Krause 2017) raises the question of whether our teaching methods, course contents and learning objectives are fully appropriate for achieving our ultimate goal, that of providing the best possible education for language professionals. Since “corpus use is anti-economic in the short term, and [therefore] has not yet become widely established among professional translators” (Aston 2009: ix-x), it may be difficult to convince learners of its potential if our teaching approaches and course objectives have an exclusively product-oriented, instrumentally-focused orientation, attempting to simulate working conditions as closely as possible, as is so fashionable these days.

Further research is needed to confirm that data-driven learning is more effective than other language learning approaches, and to explore the constraints and conditions for its use, in terms of learning settings, competence levels, linguistic features etc. (Boulton 2009:51). Yet we would suggest that it is equally important for researchers to address a more intangible and arguably more challenging question, namely, in the words of Mitchell-Schuitevoerder (2014: 30), “whether the use of corpora [...] helps students develop a critical mind and whether it enhances their actual translation skills”.

## References

- Anthony L. (2014), AntConc (Version 3.4.3) [Computer Software], Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Aston G. (2009), Foreword, in Beeby, A., Rodríguez Inés P. & Sánchez-Gijón P. (eds), *Corpus use and translating: Corpus use for learning to translate and learning corpus use to translate*, Benjamins, Amsterdam: IX–X.
- Boulton A. (2009), Testing the limits of data-driven learning: language proficiency and training, in *ReCALL* 21(1): 37–54.
- Boulton A. (2010), Data-driven learning: taking the computer out of the equation, in *Language Learning* 60(3): 534–572.
- Braun S. (2007), Integrating corpus work into secondary education: from data-driven learning to needs-driven corpora, *ReCALL* 19(3): 307–328.
- Cresswell A. (2007), Getting to ‘know’ connectors? Evaluating data-driven learning in a writing skills course, in Hidalgo E., Quereda L. & Santana J. (eds), *Corpora in the foreign language classroom*, Rodopi, Amsterdam: 267–287.
- Cresswell A. (2013), Both on and under the surface of discourse: Tagged corpora for the functional description of conjunctive language, in *Procedia—Social and Behavioural Sciences* 95: 116–125.
- Gavioli L. & Aston G. (2001), Enriching reality: Language corpora in language pedagogy, in *ELT Journal*, 55(3): 238–246.
- Granger S. (1998), Prefabricated patterns in advanced EFL writing: Collocations and formulae, in Cowie A. P. (ed.), *Phraseology. Theory, analysis, and applications*, Clarendon Press, Oxford: 145–160.
- Granger S. (2015), Contrastive interlanguage analysis. A reappraisal, in *International Journal of Learner Corpus Research* 1(1): 7–24.
- Hafner C. & Candlin, C. (2007), Corpus tools as an affordance to learning in professional legal education, in *Journal of English for Academic Purposes*, 6: 303–318.
- Hyland K. (2002), Activity and evaluation: Reporting practices in academic writing, in Flowerdew J. (ed), *Academic discourse*, Longman, London: 115–130.
- Johns T. (1991), Should you be persuaded: Two samples of data-driven learning materials, in *English Language Research Journal (New Series)*, 4:1–16.
- Kelly D. (2005), *A handbook for translator trainers*, St Jerome, Manchester.
- Kiraly D. (2000), *A social constructivist approach to translator education*, St Jerome, Manchester.
- Leech G. (1997), Teaching and language corpora: a convergence, in Wichmann A., Fligelstone, S., McEnery A. & Knowles G. (eds), *Teaching and language corpora*, Longman, London: 1–23.
- Mitchell-Schuitevoerder R. (2014), *A project-based syllabus design - Innovative pedagogy in Translation Studies*, Durham theses, Durham University. Available at Durham E-Theses Online: <http://etheses.dur.ac.uk/10830/>
- Nesi H. (2011), BAWE: An introduction to a new resource, in Frankenberg-Garcia, A, Flowerdew, L. & Aston G. (eds), *New trends in corpora and language learning*, Continuum, London: 213–228.
- Nesi H. (2013), Citation in student assignments: A corpus-driven investigation, in Hardie A. and Love R. (eds.), *Proceedings of Corpus Linguistics 2013*, UCREL, Lancaster: 225–227.
- Seidlhofer B. (2002), Pedagogy and local learner corpora: working with learning driven data, in Granger S., Hung J. & Petch-Tyson S. (eds), *Computer learner corpora, Second Language Acquisition and foreign language teaching*, Benjamins, Amsterdam: 213–234.
- Sinclair J. (1996), The search for units of meaning, in *Textus* 9(1): 75–106.
- Sinclair J. (2003), *Reading concordances*, Longman, London.



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

Toudic D. & Krause A. (2017), *EMT Competence Framework 2017*, European Commission, Brussels. Online: [https://ec.europa.eu/info/sites/info/files/emt\\_competence\\_fw\\_2017\\_en\\_web.pdf](https://ec.europa.eu/info/sites/info/files/emt_competence_fw_2017_en_web.pdf)

Vandepitte S. (2013), Research competences in translation studies, in *Babel* 59(2): 125–148.

Vygotsky L. (1978), *Mind in Society. The development of higher psychological processes*, Ed. by Cole M., John-Steiner V., Scribner S. & Souberman E., Harvard University Press, Cambridge (MA).

Yunus K. (2017), Corpus Linguistics: Pedagogic application in the 21st century, in *International Journal of Academic Research in Progressive Education and Development* 6(3): 137–152.

#### *Cited articles from CRANE*

Aitchison J. (1993), Birds, bees, and switches: psycholinguistic issues 1967-2017, in *ELT Journal*, 47(2): 107–116.

Goddard H. (2002), Islam and Democracy, in *International Affairs*, 80(1): 92–94 .

Marinova-Todd S., Marshall D., & Snow C. (2001), Missing the Point: A Response to Hyldenstam and Abrahamsson, in *TESOL Quarterly*, 35(1): 171–1

#### **Notes**

[1] Throughout this paper, following the precedents set by Thompson and Ye (1991) and Hyland (2002), “writer” refers to the person who is doing the citing , and “author” refers to the person who wrote the text that is being cited.

---

©inTRAlinea & Silvia Bernardini and Andy Cresswell (2018).

"A corpus linguistics sandwich Learners chewing over reporting verbs in academic writing", *inTRAlinea* Special Issue: Translation And Interpreting for Language Learners (TAIL).

**Stable URL:** <http://www.intraline.org/archive/article/2298>