G. Philip (2007) ' *...and I dropped my jaw with fear* – The role of corpora in teaching phraseology.' Paper read at *TaLC –7th Teaching and Language Corpora Conference*. Paris, France 1-4 July 2006 Available: http://amsacta.cib.unibo.it

"…and I dropped my jaw with fear"
The role of corpora in teaching phraseology

1. Introduction

This research reports an ongoing study into phraseological production in advanced learner writing. While most studies of learner production, whether corpus-based or not, investigate problems and their likely causes, this study compares successful and less successful uses of language. The *Phrasebuilder* task, a semi-structured writing activity designed to elicit multiple versions of a single base text (see 2), makes this comparative approach possible, and is qualitatively different to the more commonly-used types of vocabulary elicitation, such as gap-fill, used in studies by Deignan et al. (1997) and Boers (2000) amongst others.

On-line corpora, paper and electronic dictionaries, internet search engines and encyclopaedias were all familiar sources of language reference for the learners whose work is examined in the following sections [1]. All the students were given basic training in corpus use during the course, including how to carry out some of the more frequently-used advanced search techniques such as those involving multiple nodes, POS tags and wildcards. Merely possessing these skills is no guarantee that corpora will be used efficiently, appropriately, or, indeed, at all. The task therefore was given a twofold purpose – to determine how proficient the students were at combining the keywords in running text, and to see which language resources they made use of in tackling the different phraseological challenges that the exercise presented. Some of the questions to be addressed in this contribution are: how can the effects of corpora on the language learning process be effectively assessed?; does corpus use aid students in a measurable way?; and is there a marked qualitative difference between the work of students who make use of corpora in their studies and those who do not? In other words, does corpus use necessarily lead to an improvement in students' language production, and if so, in what ways?

2. The *Phrasebuilder* exercise

The *Phrasebuilder* task is an exercise in guided creative writing. It consists of a fixed sequence of keywords and collocations, presented sequentially on flash cards or slides, forming the basic skeleton of a story which students must flesh out to produce their own short narratives. The keywords appearing on each consecutive slide collocate with each other, and they may also collocate with the keywords on the preceding or following slides. As with all language, however, they can also be used compositionally. The students' task is to write their own version of the story using all the words provided, sticking as closely as possible to the order in which the keywords are presented and avoiding changes to the word form provided (plurals should remain plural, tenses of verbs should not be modified). The keywords used in this study come from a version of the task devised by Goldsmith (2003): they are reproduced

1 The data presented in this study has been gathered from five groups of learners following advanced (C1) general English courses at the University of Bologna, Italy.

in their order of occurrence in the Appendix, alongside a model version of he story which incorporates them in the required sequence.

The *Phrasebuilder*'s guided approach to creative writing is similar to the one described by Martinez (in Lewis 1997). However, the specific exercise described in this study is more complex than Martinez's version, built up around a single delexical verbs and its collocates, despite the apparent constraints that fixed keywords in a fixed sequence might suggest. In the first place, students are provided with the raw materials for constructing a story which are general enough to allow them to use their imagination, yet structured in such a way as to provide a beginning, middle and end – to ensure that the story will be complete – and limited in scope so that the task can be completed in a single session (in this case, a 2 hour lesson). Secondly, the presence of a well-delimited set of keywords allows the teacher/researcher to identify areas of language where students are in need of additional instruction, whether grammatical or lexical.

Although not designed with corpus research in mind, the potential for keyword tasks in learner-corpus studies is immediately apparent. A small corpus can be compiled from the students' contributions, differing from other learner corpora in that the texts are all variations of a single, hypothetical original. This can be likened favourably to Bowker & Bennison's (2003) approach to translator training, in which the multiple texts are translations of a single source text. Such a corpus makes it possible to analyse general tendencies in language production over a large sample, in a much more detailed and systematic way than can be hoped for with manual correction alone. The guaranteed presence [2] of the keywords in all the texts means that the researcher can easily identify areas of language to analyse, safe in the knowledge that concordances will be retrievable in sufficient numbers to ensure that findings are relevant. Further to this, s/he is also in a position to observe how a restricted set of keywords are made to combine in text, whether as collocations proper or in a looser, more compositional way. It is this factor that will form the main focus of the present study.

The decision to turn the paper exercise into a digital one was based on several factors, corpus construction actually being quite low in the order of priorities. Several disadvantages became immediately apparent when working on paper, first among which was the difficulty of reformulating the texts in the light of new keywords as the story progressed. Space on paper is limited, and students soon tired of having to score out text and re-write chunks. There were also, unsurprisingly, repercussions on the legibility of the finished compositions. Timing also proved problematic. Students were presented the keywords on slides, and while some students were quick to incorporate the new elements into their existing texts, others inevitably fell behind and missed some parts of the sequence. Finally, it also became evident that students working at their desks with pencil and paper did not think of switching on their computers to make use of the various reference sources available to them via the internet connection. As a result, the students completing this task the first time it was run rarely used reference tools at all, which was hardly ideal as the task was intended as a language learning exercise, not an assessment.

Switching the exercise from paper to computer solved all three of these problems, with the added advantages of obtaining finished texts that were legible enough to correct and easily

[2] Obviously, misspellings and typographical errors can have an affect on the location of keywords in the corpus.

convertible into corpus material. Students worked one to a computer, with several windows open on the screen in front of them – one with the slides, so they would work at their own pace, one with the word-processor, and at least one other with monolingual or bilingual dictionaries, a search engine (usually Google), or a concordancer (typically WordBanks Online). Students also had paper copies of the main advanced learner dictionaries [3] available in the classroom.

3. Outline of the study

The findings discussed in this study are the result of the analysis of five groups' work, with data gathered over four academic semesters. A total of 41 students completed the exercise, and for the purposes of analysis they have been divided into four groups: *corpus users*, *Googlers*, *mixed-resource users*, and *dictionary users*. These groupings were made on a post-hoc basis, after observing which resources each individual tended to make most use of. As a means of evaluating the usefulness of corpora and the internet in particular, the presence of a control group was indispensable. This group of ten students completed the task without the aid of any on-line resources, and account for ten of the *dictionary users*. Of the remaining 31 students, the *corpus users* numbered six, the *Googlers* three, *dictionary users* two, with the vast majority of students falling into the *mixed-resource users* group.

The *corpus users* were those who not only used the corpus, but, most importantly, were also happy to experiment with advanced searches, for example in order to find a suitable adverb to complete a verb phrase. Some even used the corpus as a source of ready-made examples which they could copy and paste into their text. The *Googlers* eschewed traditional forms of language reference, preferring to use fairly elementary single-word searches on the internet as a means of locating possible expressions, by counting the number of hits, and uses by looking at the short extract accompanying the retrieved web pages and occasionally going into the full text. The *dictionary users* not belonging to the control group appeared uncomfortable with corpora and the internet, preferring to use only dictionaries, usually in paper or hand-held electronic form, which habitually brought to class with them, while those in the control group may have belonged elsewhere had they been given the option The *mixed-resource users* used all resources to a greater or lesser extent, although it should be pointed out that within this group, there is a range running from the students who selected particular sources to address different kinds of linguistic problems and those who displayed no apparent sense of purpose in their choices, and/or made minimal use of reference tools. Were this experiment to be repeated, it might be useful to adopt a query record sheet such as that used by Frankenberg-Garcia (2005) to log the students' referencing behaviour.

A few unanticipated problems arose within the task itself. On some occasions, students changed the order of occurrence of the keywords, or indeed omitted one or more of them, despite being able to self-pace. As far as the individual compositions are concerned, this matters very little, but if viewed in the broader light of investigating students' ability to recognise potential collocations, omissions and the like caused some disruption to the data set as a whole. In a similar vein, despite instructions to avoid alterations to the word forms, many students were tempted to use alternative inflected forms of the keywords provided, changing

[3] MED, LDOCE, COBUILD, OALD.

their meanings in the process. Different inflected forms collocate in distinct ways, and the apparently innocuous change from e.g. singular to plural noun can play havoc with phraseological meaning: *I lost my nerves* does not express the same meaning as *I lost my nerve*. A third problem, from the point of view of corpus analysis, is that keywords were at times misspelled, whether as a result of typographical mistakes at the keyboard (e.g. *sleeep* for *sleep*), persistent spelling errors (e.g. *toughts* for *thoughts*), or imprecise copying of some less-familiar words from the slides to the composition (e.g. *jew* for *jaw*). Inaccuracies such as these affected the retrieval of the keywords with corpus query software.

The problems outlined above were only a minor cause of trouble, as they were infrequent. On the whole, students had no difficulty in carrying out the task and every one of them came up with an individual interpretation of the underlying story, in spite of the apparent constraints imposed by the keywords. The next section discusses in detail the ways in which students dealt with four different types of phraseological meaning that the keywords were intended to elicit, demonstrating some of the differences in quality of the writing produced by the *corpus users*, the *Googlers*, the *mixed-resource users*, and *dictionary users*.

4. Phraseological production in the *Phrasebuilder* task

The *Phrasebuilder* task is designed to elicit phraseological language through the presentation of particular word forms in a predetermined sequence. Essentially, four types of phraseological language are involved, and each of these is discussed in the subsections that follow. The most simple is the formation of collocations, such as the decision to combine keyword nouns to make a compound or the addition of known collocates in order to incorporate the keywords into running text. This can be contrasted with non-formation of collocations, whereby the keywords are treated compositionally, occurring at some distance from one another in the text. The second type discussed is the generation of phrases from a single keyword – typically one which has relatively low frequency and/or is presented in a precise inflected form which is expected to trigger the recognition (or location in reference works) of a limited range of expressions. The third type of phraseological language considered is the use of 'filler' phraseology, the lexical and grammatical choices made by students in order to link keywords which collocate strongly in the real world but rather more weakly in text. Finally, the problems posed by figurative and idiomatic meaning are examined by observing how well students are able to recognise when a term is not to be used in its literal sense, to appreciate its intended meaning and tackle its integration into their text.

4.1 Collocation versus free combination: wine + party

The keywords used to investigate students' awareness and use of collocation, as opposed to a compositional treatment of language, are *wine* and *party* (see appendix for their place in the overall sequence). These words can form a compound or be used in a looser collocation which still conforms to the standard definition of collocation as being words which co-occur within five words of the node (Sinclair et al. 1970/2004: 5).

It might be surprising to learn that only one student produced the compound *wine party*, with all others settling on the looser type of collocation relating to the real-world occurrence (especially true for these Italian students) of *wine* being drunk at *parties*. This is an example

of the influence of 'psychological collocation' (Partington 1998, 16-17) intruding on textual collocation, and this tendency is nicely illustrated in the data present for these keywords. Apart from the lone occurrence of *wine party*, one other compound was present (*wine cellar*), and five occurrences of other collocations: *bottle of wine* (2) and *glass of wine* (1), as well as *red wine* (2). By far the most frequent collocate for *wine*, however, was *drink* (14), usually accompanied by the colligational feature of quantifiers (of which *too much* occurred in 9 of the 14) completing the picture of people drinking wine to excess at parties, as illustrated in example 1.

1. It was a strange **party**, the alchool was flowing like a river, I had drunk too much **wine** and I was very drunk.

Given that these keywords are both high-frequency and easily contextualisable, there were no great differences in approach to their inclusion in students' texts, though *dictionary users* produced remarkably consistent results, shown in Figure 1. The student who used *wine party* was an assiduous *dictionary user*, not from the control group, while all but one of the control group students used a variant of the pattern *drink too much wine* +, with *party* appearing in the cotext nearby.

```
 -----Yesterday, we drank a lot of wine at the party. When I went to bed
to a party, where I drank too much wine. Not only the headache and
 to sleep because I drank too much wine at the party, and so I had a
  -----Last night I drank too much wine to a party and when I went to
confusing...maybe I drank too much wine. Now I feel a terrible headache
    in my house and drank too much wine. How could I lay down and sleep
  to sleep. Maybe I drunk too much wine at the party. A terrible headache
 went to a party. I drunk too much wine. Now, I feel a terrible headache
like a river, I had drunk too much wine and I was very drunk. That
and I think to have drunk too much wine, in fact when I came back home
 I had really exaggerated with red wine! Considered that I mustn't have
my standard. Because I went to the wine party with my colleagues. It
```

Figure 1: KWIC concordance for *wine* – dictionary users only

There were few apparent instances of free combination with this pair of keywords, though in a number of instances they were separated too far to qualify statistically as collocates. These were produced by all groups of students except the *dictionary users*, though the worst culprits were the *Googlers*. Examples 2 and 3 show typical examples of free combination produced by the students, while example 4 illustrates a case where the keywords were moved around or indeed omitted (*party* should occur between *sleep* and *wine*, with *headache* at the end of this sequence).

2. This morning I went to a **wine** cellar, to buy 50 liters of Sangiovese for the **party** I and my girlfriend would give in the evening.

3. I went to my kitchen desperately looking for a glass of **wine**, maybe there was one bottle left after the party!

4. They say: "If you can't sleep, if you've got an headache...just drink some red **wine**, and you'll be better in a while!".

That only one student used the compound *wine party* may serve as proof that students on the whole did not feel the need to look *wine* up, being quite convinced that they already knew its meaning(s). Had any of them checked in the corpus, they would not only have found confirmation that *wine* and *party* do collocate closely, but they would also have discovered that the pair actually tend to form a trio with *cheese*. It is easy to forget, both as teachers and as learners, that looking up common words can be a worthwhile exercise in vocabulary building and, as in this instance, in the unearthing of cultural information.

4.2 Building collocations: **nerves**

Nerves – note the plural form – is the keyword that was chosen in order to investigate how students tackled the generation of phrases from a single keyword. *Nerve* is not a very frequent word, but it is polysemous, with some of its meanings being restricted to the plural inflection alone. This detail proved crucially important, as did the mismatch between *nerve*(*s*) and its Italian equivalent, *nervi*, whose semantic range and phraseological patterns only partially overlap. But apart from the expected learner errors caused by calquing, some of the unexpected collocations that occurred cannot reasonably be ascribed to L1 interference (see ex. 5 and 6). There appears to have been real confusion regarding the meaning of this keyword, yet the data indicates that few students actually turned to a dictionary to clarify the meaning or to look up possible phrases to use.

5. I shooked my **nerves**, trying to forget it was less than ten minutes I was awake.
6. Noise and silly conversations were the cause of Jack's headache; his **nerves** were too tired to bear meaningless smiles and honey words.
7. In addition to this, it started rainning. In fact, it really got me **nerve**.
8. In my personal point of view rain is the third kind of weather that stand on my **nerves**, snow is the second and windy is the first.

Examples 7 and 8 also appear not to have been produced with the help of a dictionary, but merely used in a half-remembered, or back-translated form. In these examples, the underlying meaning can be identified, but the word forms used are not the ones that typically communicate the intended meaning (*got on my nerves* in both instances). As Hanks points out, "linguistic behaviour among users of a language is highly stereotypical, even in matters of fine detail" (2004: 246). Such fine detail includes limitations on which inflected forms can express particular meanings, as the examples above illustrate. One pedagogical question that inevitably arises is how and when such detail should be taught, given that language teaching tends not to devote as space much to fine details of lexis as it does to those of grammar.

Of the four groups of students, none produced flawless texts, but the use of appropriate collocations with nerves was quite obviously the preserve of those who had looked the word

up in one or other of the language reference tools available[4]. The *Googlers* produced expressions which verge on the incoherent, featuring the collocates *calm* and *break down* (see examples 9 and 10). Even allowing for the fact that *calm + nerves* is acceptable, its use in the text points to a lack of comprehension, and a consequent inability to contextualise the phrase appropriately. This student may well have found the phrase on the internet, but failed to look further than the search results page to check its meaning in cotext.

9. … so the party finished and we had to calm ours **nerves** backing home a little bit drunk
10. I had spent all the morning working hard - and the rain didn't make things easier -, my **nerves** broke down and I went into a heavy depression.

The *dictionary users* were also guilty of making up collocations including *irritate one's nerves* (ex. 11), and they also fell into the trap of thinking that the plural form could substitute the singular in collocations such as *lose one's nerve* (ex. 12). Overall, however, when the students in this group used acceptable collocations such as *get on one's nerves*, and *nerves + on edge* (ex. 13 and 14), they did so correctly, presumably due to the presence of examples in the dictionary entry.

11. More, a strong wind and an incessant rain irritated my **nerves**.
12. …I lost my **nerves** and I had so much fear that my jaw was moving itself without my will.
13. …a terrible storm was shaking my house, and the rain on the windows was getting on my **nerves**.
14. I began to walk towards my home when it began to rain. My **nerves** were on edge.
15. Outside the window an heavy rain was falling. It hurted my **nerves**.
16. I shooked my **nerves**, trying to forget it was less than ten minutes I was awake.
17. That night were getting on my **nerves**!
18. …I could even perceive the light smell of the rain I began to be a bundle of **nerves**.
19. I listen a strong wind, which is starting to get on my **nerves**; it's pouring with the rain...
20. …I felt anxious, I made an effort to calm my **nerves**.

The collocations produced by the *mixed-resource users* were much like the *dictionary users*' ones: a few invented collocations such as *hurt one's nerves* (ex. 15) and *shake one's nerves* (ex. 16) some standard collocations (typically *get on one's nerves*), and a couple of instances of more idiomatic phrases such as *bundle of nerves*. The general level of textual accuracy was lower than that of the *dictionary users* group: examples 17 and 18 illustrate how the phrase itself can be faultless but yet not run on seamlessly from the preceding text. This error was not present in the *corpus users*' texts (ex. 19 and 20), whose use of collocation was phraseologically accurate, with the chunk being incorporated into the rest of the text with no apparent breaks. It may be the case that when corpora are used, the phrasal boundaries are less clearly cut than they appear in a dictionary entry, so a student using corpus data for reference will have a better model of how to mould the cotext around the chunk to make it fit properly. Clearly, this is the reason why dictionaries include example phrases, but a KWIC concordance does make repeated patterns more immediately recognisable, and so may have more influence on a student's choice of words.

[4] It is worth remembering that students are unlikely to have looked up every single keyword in the task: by comparing the tendencies of each group, however, a general picture emerges, and from this it is possible to determine which choices were influenced by reference tools and which were not.

4.3 Filler phraseology: **heart + chest**

While identifying correct or suspect collocates for a keyword is a fairly straightforward procedure, the analysis of filler phraseology is often relegated to second place in collocation studies, especially those dealing with learner data. A lot of filler phraseology is grammatical, and perhaps this is the reason why it tends to be ignored in favour of the more colourful lexical collocates. However filler phraseology, both grammatical and lexical, is an important part of language production. It is the glue with which lexical chunks are stuck together. This subsection investigates the lexical and grammatical choices made by students linking the keywords *heart* and *chest*.

In the real world, *hearts* are found in *chests*; or, to be more specific, the heart muscle is located within the chest cavity. It is not normally deemed necessary to make this sort of specification, but both of these keywords are polysemous, and it is really only by virtue of their proximity that we are inclined to interpret them both as parts of the body. One student (ex. 25) did exploit the polysemy: for the others there was no question that the meaning was anatomical, especially since the language was intended describe the emotions felt by the narrator in the light of the events up to this point in the text. Fear, trepidation and anticipation are all manifested by physical responses, typically including an increased heart rate. But how did the students express this familiar concept in a foreign language?

The four groups of students followed distinct patterns of usage in their filler phraseology, though the same general structures were adopted in order to express the idea of the heart beating fast within the chest. One interesting recurring feature in the students' filler phraseology was the presence of verbs in the progressive aspect (generally underused by Italian learners of English). This often followed *start* or *begin*, emphasising the immediacy of the events. A further interesting aspect was the presence of adverbs, which were not used a great deal elsewhere in the texts (see Figure 2).

his **heart** beating in his chest:
my **heart** is beating fast in my chest.
my **heart** was beating more fast
My **heart** started beating madly.
My **heart** started beating incredibly fast
My **heart** began to beat so rapidly that...
my **heart** started to beat very quickly
the **heart** in my chest and his beating was almost deafening
The **heart** begins bumping in my chest
my **heart** began to thump wildly in my chest.
My **heart** jumped out of my chest!
Figure 2: Verbs collocating with *heart*

The *Googlers* used simple, high-frequency verbs and did not use any adverbs and used compositional language throughout (see ex. 21 and 22).. While the cartoon-style image of a heart literally jumping out of somebody's chest is clearly intended in both these examples, they both lack the conventional wording which would make this meaning immediate. Similar

meanings are expressed more successfully in examples 23 and 24 (produced by a *dictionary user* and a *mixed-resources user* respectively).

21. That's the reason why, I suppose, when I suddenly heard someone knocking at the door, my **heart** tried to escape from my **chest**.
22. I couldn't open the door and my **heart** was got off my **chest**
23. My **heart** started beating incredibly fast I thought it would have popped out of my **chest**.
24. My **heart** began to beat so rapidly that I feared it was going to jump out my **chest**.
25. My **heart** started beating madly. How could I save myself? The **chest**! I jumped inside.
26. The **heart** begins bumping in my **chest** with the decise sound similar to footsteps on a stairs.
27. I suddenly heard something strange and my **heart** began to thump wildly in my **chest**.
28. I suddenly heard a door slammed just over me. Even more violently slammed the **heart** in my **chest** and his beating was almost deafening.

Of the remaining three groups, patterns in verb and adverb usage are discernable. The *dictionary users* opted almost invariably for the verb *beat* accompanied by a range of adverbs relating to speed, including *quickly* and *fast*, as well as adverbs of manner such as *madly* (see ex. 23 and 25). *Mixed-resource users* drew upon a wider range of verbs, including *bump*, *run* and *pound*. While *beat* was invariably accompanied by adverbs (see ex. 24), the more specific verbs tended not to be, though there are some interesting run-ons, such as that in example 26. Finally, *corpus users* used a wide range of verbs, usually in combination with specific adverbs, as illustrated in examples 27 and 28.

4.4 Figurative language: **my jaw...**

Jaw is not a word in everyday usage: *jaws* do not arouse a great deal of interest and are not brought up in conversation very often. It was therefore unsurprising that many students felt perplexed by the presence of this keyword near the end of their task. While *heart* and *chest* were easy to fit into the story, *jaw* appeared not to be. Why should this part of the head be important?

A proficient user of English should recognize that *my jaw*... (and not *jaws* plural) leads into a restricted set of possible verb collocates, namely *drop* or *clench* and their synonyms. Even an advanced student may not have come across either of these expressions, through lack of exposure to the language, and it should not be surprising that even at this level students do not realise that the keyword is intended to elicit one of these metaphors (probably the former, given the context). Yet problems are not wholly negative. The confusion that *my jaw* caused had the fortunate consequence of prompting students to look up possible uses, and as a result only eight students chose to interpret the word literally (see ex. 29).

29. Too tired and angry to switch the light on I jumped in the bed hitting **my jaw** on the floor: it was the wrong direction.

The *Googlers*, in line with their previous tendencies, either gave a literal interpretation or miscued on the necessary wording (ex. 30 and 31). Most *dictionary users* completed the phrase with *dropped*, although *clenched* also appeared once, alongside four literal, compositional interpretations (see ex. 32 – 34).

30. I was surprised too when opened **my jaw** and I realized that it was the wrong house.
31. **My jew** fell down to the floor.
32. **My jaw** dropped, my heart started to beat very quickly and my chest was almost hurting!
33. My mind was completely flooded by frightful thoughts, **my jaw** was clenched.
34. Ok, now is the moment - I'm at the end of the bed! I feel a soft ache in **my jaw**. I can call my neigbour, he's doctor.

The tendencies displayed by the *mixed-resource users* (ex. 35-36) were similar to the *dictionary users*' ones, using *dropped*, but were several instances of the less common active form of the verb (ex. 36). This also occurred amongst the *corpus users*, who also produced some unexpectedly inventive interpretations of the keyword (see ex. 37). This burst of inventiveness is strikingly uncharacteristic in the light of the corpus users' performance elsewhere in the task, as is the literal interpretation shown in example 38. Why does this apparent lapse occur at this stage?

35. It took some seconds, but when I realized what had just happened **my jaw** dropped.
36. My teeth were chattered and I dropped **my jaw** with fear.
37. **My jaw** was clicking like a typewriter for the fear.
38. **My jaw** hardened to an unsustainable degree, all my muscles were stiff and, because of the fright, I could hardly breath.

At this point, it is enlightening to take a closer look at the corpus data. There were only 14 occurrences of *my+jaw* in the WordBanks Online data when these students completed their task, all of which are reproduced in Figure 3. Although the conventional collocate *dropped* is present three times – twice to the right of the node (lines 3-4), and once to the left (line 12) – these frequencies are inadequate for any tendencies to be identified. This may well account for the presence of active forms of the verbs amongst the *mixed-resource group*.

A second glance at line 8 in the concordance explains the mystery of ex. 37.

```
 1   and my camera, not to mention my jaw, free-fell to the floor. Japan's
 2  in a hotel without dislocating my jaw on my knees. I can sleep in tents
 3   you propose to get it out [p] My jaw dropped Oh God, I never thought
 4 in open water. I glanced up and my jaw dropped, for only thirty feet
 5  and I felt horribly miserable. My jaw ached from grinning all night
 6    out at about 2.45am. Whether my jaw muscles simply retired hurt, or
 7        me but about the pain in my jaw. Mind you, judging by the stares
 8 Worse still, my teeth ached and my jaw was clicking like a typewriter.
 9   me on the head Another said: 'my jaw fell to the floor and fractured
10     from my stomach clear up to my jaw. The peace symbols swirled
11  thigh, then landed squarely on my jaw. I kept both hands on the wheel.
12     I asked, trying not to drop my jaw. I admire Gail. She's one of the
13            [M02] No I dislocate my jaw all the tie. I [ZGY] smashed
14 ha ha ha. No. No. Please no more my jaw's aching all [tc text=pause]
```

Figure 3: *my+jaw* (WordBanks Online: all concordances)

Is it right to criticise students for cherry-picking their way through corpora rather than restricting their interest to frequent, recurrent forms? Corpora provide native-speaker models, real examples of language in use that are there for the taking, and this student should probably

be applauded for having embraced corpora so warmly. High-frequency patterns have to be learned in order to communicate effectively, but this does not necessarily mean that low-frequency forms have to be ignored completely. Should learners be denied the right to be creative and opportunistic for no better reason than the fact that they are non-native speakers? This previously hidden aspect of learner corpus use is incredibly interesting, and certainly worthy of more detailed investigation before any pedagogical conclusions can be drawn legitimately.

5. Corpora, users, and language learning

Relatively few students in this study chose to use the corpus as their main source of language reference. One immediate reason for this is that it is a newly-encountered tool: even those who had used corpora in previous courses had only ever encountered them in their university studies, while they have been using dictionaries since primary school. A further reason, unlikely as it may sound, is that there are still students reaching degree level who are barely computer literate and who avoid using the computer in the language class whenever there is an alternative available. In order to gain the maximum benefit from what corpora have to offer, students need to have training in how to formulate search queries so as to obtain the data that is most relevant to their needs. Such training takes precious time away from other aspects of language learning, and not all teachers are prepared to invest time in cultivating corpus skills, many advanced students fail to reap the full benefits offered by corpus data. Although the success of a query is dependent on the effectiveness of the search syntax used, learning the query language for any corpus is worthwhile, not least because of its knock-on effect of improving one's use of search-engine syntax too. In this way, an accomplished *corpus user* also becomes a more efficient *Googler*.

So corpora and corpus-based reference tools both aid students' vocabulary learning demonstrably, but this can only be true if they are indeed consulted. The data presented here, as well as other learner data discussed in Philip (2005; 2006) confirms the fears of many language teachers: advanced students limit their looking up to the bare minimum, usually only to check the meaning of new words. Some students see consultation as a sign of weakness, believing that only less-proficient students need check meanings (form appears to come even lower down the priority list). Thus, out of embarrassment or fear of shaming themselves, they relegate their dictionaries and corpora to tasks where they are explicitly required. Other students fall foul of the belief that once a word has been encountered, it has been 'learned', forgetting that such learned lexis is rarely revisited, hindering the learning of sub-senses, figurative extensions and phraseology. Students generally encounter words in their literal sense first, match them to a translation equivalent in their L1, and from then on, unless instructed otherwise, use the word in calqued forms of the L1 phraseology. The relative success of this strategy effectively masks the underlying problem, which is more serious than simply getting collocations wrong. Persistent calquing actually prevents students from acquiring a sense of the word's conceptual range in the L2, negatively affecting textual fluency and cohesiveness.

The data presented in 4 demonstrate that corpora do indeed have a positive effect on language production, but the advantages are not restricted to hands-on use. Corpus-based advanced learner dictionaries condense and select information from the corpus and also provide examples and usage notes. The differences identified between *corpus users* and *dictionary*

users are not easily quantified statistically, but this study has revealed some of the qualitative differences present in students' written production when aided by dictionaries or by corpora, or indeed by a combination of the two. At the same time it has found that the internet, while undeniably useful as a linguistic ready-reckoner, fosters neither accuracy nor variety in the acquisition and use of lexis and phraseology.

One of the most interesting points of note to come out of this study was the observation that higher-quality texts are not produced by *corpus users* simply because they are using a corpus, but because the *corpus users* displayed several traits associated with the archetypal 'good language learner' (Oxford 1990). These traits are shared to a lesser extent by the more discerning amongst the *mixed-resource users*, as well as the control-group *dictionary users* (although these students were explicitly asked to consult their dictionaries as often as possible). Adventurous *corpus users* are students who are motivated enough to make the effort to learn to use the query syntax. They then take care to identify precisely what forms of language they wish to locate, and are prepared to try various permutations of their queries before they settle on the data that best satisfies their needs. They appear to actively avoid translating directly from their L1, especially when the language is figurative, preferring to copy the native-speaker models present in the data. In other words, they are students who have simply added corpus use to their repertoire of effective language learning strategies.

References

Boers F. (2000) Metaphor awareness and vocabulary retention. *Applied Linguistics* 21, 553-571.

Bowker, L. and P. Bennison 2003. Student translation archive: design, development and application. In Zanettin, F., S. Bernardini & D. Stewart (eds) *Corpora in translator education*. Manchester: St Jerome.

Deignan A., D. Gabrys & A. Solska (1997) Teaching English metaphors using cross-linguistic awareness-raising activities. *ELT Journal* 51, 352-360.

Frankenberg-Garcia, A. (2005) A peek into what today's language learners as researchers actually do. *International Journal of Lexicography* 18 (3), 335-355.

Goldsmith. P. (2003) Working wonders with words. Paper presented at 37th IATEFL conference and exhibition, Brighton 22-26 April 2003.

Hanks, P. (2004) The Syntagmatics of Metaphor and Idiom. *International Journal of Lexicography* 17 (3), 245-274.

Lewis, M. (1997) *Implementing the Lexical Approach: Putting Theory into Practice*. Hove: Thomson Heinle LTP.

Oxford, R. (1990). *Language Learning Strategies. What Every Teacher Should Know*. Boston: Heinle & Heinle.

Partington, A. (1998) *Patterns and Meanings. Using Corpora for English Language Research and Teaching*. Amsterdam and Philadelphia: John Benjamin.

Philip G. (2005) From Concept to Wording and Back Again: Features of learners' production of figurative language. In J. Barnden, S. Glasbey, M. Lee, A. Wallington & L. Zhang (eds) *Proceedings of the Third Interdisciplinary Workshop on Corpus-Based Approaches to Figurative Language*. Birmingham: University Press, Birmingham, 46-53.

Philip G. (2006) Metaphor, the dictionary, and the advanced learner. In E. Corino, C. Marello & C. Onesti (eds.) *EURALEX Proceedings of the XII International Lexicography Congress*. Alessandria: Edizioni dell'Orso, 895-905.

Sinclair, J., S. Jones & R. Daley (1970/ 2004) *English Collocation Studies: The OSTI Report*. 2nd Edition, R. Krishnamurthy (ed.): London & New York: Continuum.

Appendix

Phrasebuilder: ***keywords in sequential order***

I; bed; unable to sleep
went to
wine; party
headache; stomach; wind
rain; nerves
remember; film
turn; key; lock
suddenly heard; door
heart; chest
footsteps; stairs
thoughts; my mind
think; save myself
door; open
something; end of the bed
my jaw
neighbour; wrong house

Phrasebuilder: ***model text***

Last night **I went** to **bed** but I was **unable to sleep**. I had got home late after my friend's **party**, and I'd definitely had too much **wine**. I had a terrible **headache** coming on, and my **stomach** was churning. Outside the **wind** was howling, and the **rain** battering on my window was getting on my **nerves**. For some strange reason, it made me **remember** a **film** I'd seen.
The sound of a **key turn**ing in the **lock** downstairs brought me back to earth. I **suddenly heard** my own front **door** slam shut. My **heart** started pounding in my **chest** as I realised that I had an intruder in the house! I heard **footsteps** steadily making their way up the **stairs**. All manner of **thoughts** went through my **mind**, as I tried desperately to **think** how I could **save myself**. But there was no time. I lay in terror as the **door open**ed slowly, and in the darkness of my bedroom I could make out **something** near the **end of the bed**.
My jaw dropped when the light was abruptly switched on and I came face to face with my **neighbour**, James, who had been at the party with me. He was in a worse state than I was, and had come back to the **wrong house**!