

Article:

How EFL students can use Google to correct their "untreatable" written errors

Luc Geiller
ATILF/CNRS, Nancy University, France

[lgeiller @ gmail.com](mailto:lgeiller@gmail.com)

Abstract

This paper presents the findings of an experiment in which a group of 17 French post-secondary EFL learners used Google to self-correct several "untreatable" written errors. Whether or not error correction leads to improved writing has been much debated, some researchers dismissing it as useless and others arguing that error feedback leads to more grammatical accuracy. In her response to Truscott (1996), Ferris (1999) explains that it would be unreasonable to abolish correction given the present state of knowledge, and that further research needed to focus on which types of errors were more amenable to which types of error correction. In her attempt to respond more effectively to her students' errors, she made the distinction between "treatable" and "untreatable" ones: the former occur in "a patterned, rule-governed way" and include problems with verb tense or form, subject-verb agreement, run-ons, noun endings, articles, pronouns, while the latter include a variety of lexical errors, problems with word order and sentence structure, including missing and unnecessary words.

Substantial research on the use of search engines as a tool for L2 learners has been carried out suggesting that the web plays an important role in fostering language awareness and learner autonomy (e.g. Shei 2008a, 2008b; Conroy 2010). According to Bathia and Richie (2009: 547), "the application of Google for language learning has just begun to be tapped." Within the framework of this study it was assumed that the students, conversant with digital technologies and using Google and the web on a regular basis, could use various search options and the search results to self-correct their errors instead of relying on their teacher to provide direct feedback.

After receiving some in-class training on how to formulate Google queries, the students were asked to use a customized Google search engine limiting searches to 28 information websites to correct up to ten "untreatable" errors occurring in two essays completed in class. The findings indicate that a majority of students successfully use material from the various snippets of texts appearing on the Google results pages to improve their writing.

Keywords: Data-driven learning, Google-driven language learning, learner autonomy, error treatment, self-correction, language awareness.

1. Introduction

1.1. Data-driven learning (DDL)

"Data-driven learning" (DDL) was first used by Johns (1990) to refer to learners directly exploring authentic language by means of corpora, acting as researchers discovering language patterns, formulating and testing hypotheses. A number of recent studies have highlighted the usefulness of corpora and concordancers as tools to facilitate second language learning, particularly its impact on vocabulary acquisition and improved writing skills (Chambers, Conacher & Littlemore 2004; Chen 2004; Chen & Baker 2010; Jarvis 2004; Johansson 2009; Kennedy & Miceli 2010; Yoon 2008; Yoon & Hirvela 2004). As explained by Boulton (2009a: 83), DDL "can sensitise learners to issues of frequency and typicality, register and text type, discourse and style, as well as the fuzzy nature of language itself."

Reporting on their attempts to make concordance information accessible to lower-intermediate L2 writers as feedback to sentence-level written errors, Gaskell and Cobb (2004) explain that learners are willing to use concordances to work on grammar and that they are able to self-correct based on those concordances. They argue that online corpus exploration can reduce the burden on teachers, all the more so as the formal teaching of rules is not always effective in helping learners achieve more grammatical accuracy because "sentence-level writing errors seem immune to many of the feedback forms devised over the years" (p. 1). Similarly, Milton (2006) believes that encouraging learners to use online corpora for assistance "can help relieve teachers of the need to act as proofreading slaves" (p. 125). The rationale behind this is that maximizing learners' contact with English helps them detect recurring language patterns, thus increasing their language awareness in a data-driven learning process. The objective is for them "to acquire the means and confidence to self-edit in the future" (p. 131), which is in keeping with what Benson (2001) says about learner autonomy and language acquisition being dependent upon the capacity to initiate and manage one's own learning:

Many advocates of autonomy in language learning would [...] share Rousseau's view that the capacity for autonomy is innate but suppressed by institutional learning. Similarly, Rousseau's idea that learning proceeds better through direct contact with nature re-emerges in the emphasis on direct contact with authentic samples of the target language that is often found in the literature on autonomy in language learning. (p. 25)

But although the use of corpora in the classroom has imposed itself as an inescapable language learning tool, several barriers must be overcome before it goes mainstream. The activity is potentially time-consuming and tedious, and teachers and students can be reluctant to accept the changes to their traditional roles in the learning process. It may even be that they do not have a sufficient level of competence in ICT. More concretely, Widdowson (2000), argues that analyzing decontextualized and truncated concordance lines is an inauthentic activity and Johansson (2009) deplores the lack of empirical evidence supporting the theoretical benefits of DDL. Yoon (2008), for his part, suggests that learning style preferences can account for the slow acceptance of corpus use as an educational tool. As he puts it, "many corpus studies have regarded learners as a monolithic group rather than as idiosyncratic individuals" (p. 32). In other words, while some learners obviously benefit greatly from the approach, others do not. The challenge, then, is for teachers to adapt corpus exploration techniques to different learners so as to better cater to their individual needs.

1.2. Google-driven language learning

According to Rundell (2000: n. pag.), the web “is not a corpus at all according to any standard definitions: what it is is a huge rag-bag of digital text, whose context and balance are largely unknown.” Berg (2005: 2), for his part, argues that “the Web turns out to be a somewhat intractable collection of textual material, [...] a rather haphazard accumulation of digital text.” The acronym GALL (Google-assisted language learning) was first coined by Chinnery (2005) who described Google as an informative, productive, collaborative, communicative, and aggregative tool with lots of pedagogical uses. Substantial research on Google as a tool for second-language learners has since then been carried out (e.g. Guo & Zhang 2007; Milton 2006; Shei 2008a, 2008b; Wu, Franken, & Witten 2009) suggesting that it plays an important role in fostering language awareness and learner autonomy. According to Bathia and Richie (2009: 547), “the application of Google for language learning has just begun to be tapped.” A number of studies, however, point to problems associated with the use of Google and the web for language learning, namely the abundance of potentially unreliable data and the daunting task of scouring huge amounts of language (Berg 2005; Kilgarrieff 2001; Renouf 2003; Fletcher 2004; Robb 2003a, 2003b; Rundell 2000). Robb (2003a) calls it “a quick ‘n dirty corpus tool,” he warns about its use in class (2003b), explaining that queries are limited to specific words only, that there is no way of assessing the reliability of the language featured in the search results, and that these are not presented in a user-friendly format.

Several attempts at harnessing and systematizing web output have been made though. Since 1998, the University of Central England in Birmingham has been developing WebCorp1, a system for extracting linguistic data from the web, presenting examples of word usage from the Web in a form suitable for linguistic analysis. Similarly, KWICFinder2 and WebAsCorpus.org3, launched in 2007 by William Fletcher, can produce concordances from webpages. Guo and Zhang (2007) have built a customized collocations collector that can be used by language users, and Wu et al. (2009), acknowledging the heterogeneous, uncontrolled, and messy nature of web data, have explored the use of web searches as a language learning tool and used the Greenstone digital library software4 to organize raw online data that can be sifted through by language learners. But if Google enthusiasts insist on using raw online data, one way of dealing with the messiness and potential unreliability of the search results can be to use Google Custom Search5, a service launched by Google in 2006 which allows creators to select what websites will be used to search for information, thus eliminating any unwanted websites. For language learning purposes, it is thus possible to create a search engine that will only search specific news websites, for example.

1.3. Google use and its impact on language development

Several studies have documented the impact of the web and search engines on language development and writing improvement (Acar, Geluso, & Shiki 2011; Clerehan, Kett, & Gedge 2003; Conroy 2010; Johnson 2004; Kennedy & Miceli 2010; Kenworthy 2004; Krajka 2000, Mansor 2007). Shei (2008a, 2008b) has shown that Google searches make it possible to compare the frequency of extended collocations (combinations of up to four words) and find the most commonly used and hence more formulaic ones. This suggests that Google output, however messy it is, *can* be used by second-language learners to explore native-speaker discourse and increase their language awareness.

Various studies have shown that some learners are keen users of information-related web services (e.g. Schroeder et al. 2010; Palfrey & Gasser 2008). Conroy (2010) reports that his students enthusiastically used Google and traditional concordancers for

language learning and error correction but that training was a key factor in getting them to use the approaches successfully. Although Google is a useful writing support tool, deciding which errors are amenable to correction needs further exploring. He also explains that students, being regular Google users, are more likely to favour the search engine than traditional corpora for which new interfaces have to be learnt, something learners sometimes find off-putting. Sun (2003) and Hafner and Candlin (2007) also found that learners preferred using Google to concordancers to learn about idiomaticity. As Shei (2008b) puts it, Google “remains a constant companion to the learner in the absence of the tutor. All the [teacher] has to do is to show the learner how to use this versatile tool” (p. 23). As explained by Boulton (2012): The objections [...] to using the web as ‘corpus’ and search engine as ‘concordancer’ have been shown to be largely theoretical, and based on criteria which are of little relevance in language teaching. The main conclusion is pragmatic and practical rather than dogmatic or ideological: if an approach or technique is of benefit to the learners and teachers concerned, it should not be ruled out automatically (Hafner & Candlin, 2007). As so often, there is likely to be a payoff between how much the teachers / learners are prepared to put in (ideally as little as possible) and how much they want to get out (ideally as much as possible). (n. pag.)

Kennedy and Miceli (2010) describe their use of the Contemporary Written Italian Corpus (CWIC) created at Griffith University to teach Italian to beginners, and especially to use corpus information to self-correct. Referring to Johns (1988), they sought to help their students develop observation strategies to extract information from concordances, developing what they call an “observe and borrow” mentality first, before progressing to an ‘observe and derive rules’ approach” (p. 1). They then explain that their aim was to “facilitate as much as possible their noticing the gap between their interlanguage and native speakers’ production,” encouraging them to explore the corpus “in search of words, expressions and even sentences that can be ‘plundered’ for use in their own compositions”—a “treasure-hunting” activity as they call it (p. 5).

1.4. Error treatment in second language writing

Whether or not error correction leads to improved writing has been much debated, some researchers dismissing it as useless (e.g. Hendrickson 1978; Kepner 1991; Sempke 1984; Truscott 1996; Zamel 1985) and others arguing that error feedback leads to more grammatical accuracy in students’ writing (e.g. Bates, Lane & Lange 1993; Bitchener et al. 2005; Bitchener 2008; Ellis 1998; Ferris & Roberts 2001; Ferris 2004; Hyland 2003; Chandler 2003). In her response to Truscott (1996), Ferris (1999) explains that it would be unreasonable to abolish correction given the present state of knowledge, and that further research needed to focus on which types of errors were more amenable to which types of error correction. In her attempt to respond more thoughtfully and effectively to her students’ errors, she made the distinction between “treatable” and “untreatable” ones: the former occur in “a patterned, rule-governed way” and include problems with verb tense or form, subject-verb agreement, run-ons, noun endings, articles, pronouns, while the latter include a variety of lexical errors, problems with word order and sentence structure, including missing and unnecessary words. Explaining that there is no handbook or set of rules to consult in order to avoid or fix those types of errors, she opted, in part, for direct correction hoping it “would, if nothing else, provide input for acquisition of these idiomatic forms” (p. 6). Noting that 50% of all errors she identified in her students’ compositions were “untreatable,” she argued that “ESL writing teachers would do well to give much more thought to how they provide error feedback regarding these different types of language forms and structures” (p. 6).

This study attempts to build on existing research into error treatment and especially the role Google can play in stimulating language awareness and enhancing self-editing

skills. “Untreatable” errors arguably occur when students are trying to emulate native speakers, working with their interlanguage, building on it using their acquired knowledge of rules and repository of words and expressions to formulate increasingly complex occurrences. The issue at stake is thus to find out if, during a self-correcting process, EFL learners can search the web and use raw online data, breaking down snippets of texts featured in Google search results, identifying and using various expressions and inherent language patterns to bring changes to their own non-native-like formulations.

2. Method

2.1. Participants

The *classes préparatoires aux grandes écoles section EC*, commonly called *prépa EC*, consist of two selective years preparing post-secondary students for competitive entry exams to France’s business schools. The program includes three hours of English teaching per week and consists in writing argumentative essays, answering reading comprehension questions, and translating newspaper articles and short excerpts from contemporary novels. The participants were 17 second-year French *prépa EC* students from a French *lycée*: 12 male and 5 female with an average age of 19 years. They all had French L1, had received at least six years of English instruction, and their levels varied from upper-intermediate to advanced (B2-C1). Since the beginning of their first year, they had been encouraged to read the press in their own time in order to complement the work done in class and gain a sense of self-direction, a key to learning languages and to learning how to learn languages (Holec 1980, 1981). It is generally agreed that autonomy cannot be taught and learned but only fostered and developed (Benson 2003:290) and the students were thus trained to scan newspaper articles in search of noteworthy linguistic material and also encouraged to compile their own lists of words and expressions spotted during in- and out-of-class “treasure-hunting activities” (Kennedy & Miceli 2010: 6).

2.2. Procedure

During the first step of the experiment, students were introduced in class to a customized search engine restricting searches to 28 information websites created using Google Custom Search (see Table 1), a service launched by Google in 2006 allowing creators to select what websites will be used to search for information, thus eliminating any unwanted websites and limiting the amount of potentially unreliable results. A set of explicit guidelines introduced students to working with Google by showing them how to perform simple and more advanced search options. It consisted of a description of the various search options, a series of search results screenshots, and sample corrections of untreatable errors performed with the help of the search results (details are provided in the next section). During the second step of the experiment, the students wrote two essays, I underlined a number of untreatable errors they contained, and the learners were then instructed to correct them at home using the customized search engine and send me their corrections via email. I then proceeded to analyze the types of searches they had performed, their use of the material featured in the search results and whether the correction was successful or not. At the end of the experiment, the students were given the opportunity to provide feedback on their use of Google Custom Search to self-correct their errors. They provided answers to a questionnaire featuring seven closed questions on a 5-point Likert scale and open questions for additional comments.

Home page	
http://www.google.fr/cse/home?cx=011764784480104570934:4qgipwv8a2q	
Indexed websites	
www.bostonglobe.com	www.uk.wsj.com
www.cbsnews.com	www.usatoday.com
www.chicagotribune.com	www.usnews.com
www.csmonitor.com	www.voanews.com
www.edition.cnn.com	www.washingtonpost.com
www.europe-wsj.com	www.bbc.co.uk
www.ft.com	www.economist.com
www.latimes.com	www.guardian.co.uk
www.newstatesman.com	www.independent.co.uk
www.nytimes.com	www.observer.guardian.co.uk
www.online.wsj.com	www.spectator.co.uk
www.reuters.com	www.telegraph.co.uk
www.thedailybeast.com	www.thesundaytimes.co.uk
www.time.com	www.thetimes.co.uk

Table 1. News websites indexed by the customized Google search engine.

2.2.1. First step: introducing learners to Google search

In the next two sections, simple and more advanced search options are presented respectively.

A) Searching for exact words and phrases using quotation marks and wild cards. Learners were first shown how to use the search engine to solve grammar problems and find collocations and idioms. By using the quotation marks around a search string, Google makes it possible to search for exact word combinations and whole phrases. It is possible, for instance, to compare prepositional constructions such as the number of hits for *"it depends on"* and *"it depends of"* (543,000,000 and 4,420,000 hits respectively) and find the most frequently used form (e.g. Shei 2008a). Another example: if learners are uncertain over the correct way of saying that a task or job requires no effort, they can enter *"it's as easy as"* in the search box and scour the results to find the answer (*it's as easy as pie*, *it's as easy as ABC*, and *it's as easy as falling off a log* being the recurring expressions). But learners can also use a wildcard (*) in the search string to leave open a slot for one or more words. Entering *"it's a * step forward"* in the search box enables them to retrieve a variety of adjectives used

with *step forward* in the snippets of text listed by Google. They can then select and compare the number of hits and choose the most frequently used ones (*it's a great step forward* occurs 4,170,000 times, *it's a big step forward* 676,000 times, *it's a major step forward* 496,000 times, and *it's a huge step forward* 319,000 times).

B) Searching for expressions using word combinations. In-class training then moved on to more advanced Google searches that rely on word combinations meant to generate snippets of texts that can be explored in search of words and expressions to plunder for use in personal sentences. The rationale behind this was that learners could scour the results and borrow the native-like linguistic material their interlanguage precluded them from formulating themselves, and then weave it into their own formulations. For example, if learners want to write about the need for politicians to implement an assault weapons ban, they were shown that by entering *ban* followed by *assault weapons* in the search box, Google generates a series of results which can then be observed and borrowed from (see Figure 1).

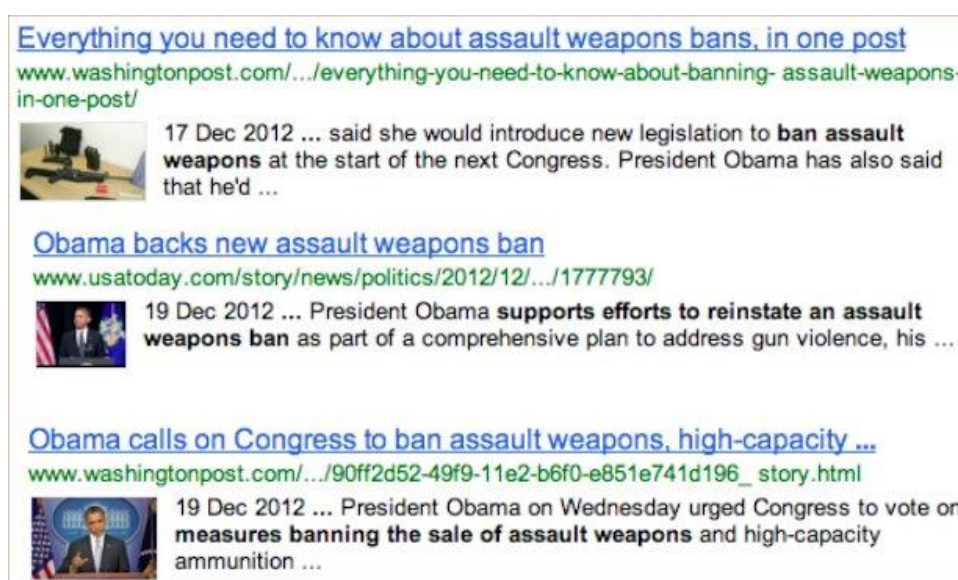


Figure 1. Selected search results for *ban assault weapons*.

Using these examples, it is possible to write a series of forceful arguments like "politicians need to *introduce new legislation to ban assault weapons*" (using the first snippet), "US politicians must make *efforts to reinstate an assault weapons ban as part of a comprehensive plan to address gun violence*" (using the second snippet), and "politicians must *vote on measures banning the sale of assault weapons and high-capacity ammunition*" (using the third snippet).

Another example: if learners are trying to express the idea that immigrants are sometimes discriminated against but don't know how to combine their words, they can enter "immigrants" followed by "scapegoats" (see Figure 2).



Figure 2. Sample search result for "*immigrants scapegoats*".

We see that "Immigrants are scapegoats for high unemployment rates" is one possibility. And using material from one snippet, the learners can then find other

noteworthy elements. Here they can enter the sentence builder "immigrants are scapegoats for" (not forgetting quotation marks) to find how else it is complemented in the press (see Figure 3).



Figure 3. Selected search results for "Immigrants are scapegoats for".

Finally, learners can use Google to check the idiomaticity of their formulations and find alternatives in case they are not native-like. To that end, they can combine the quotation mark search with the keyword search. For example, is it native-like to write "privacy issues involving Google and Facebook"? Entering the expression in the search box with the quotation marks generates no result at all. But it is not the case when the same expression is entered without the quotation marks as Google now lists a series of articles combining the words in one way or another (and not in the exact order we want them to occur as is the case when using the quotation marks). The material featured in the snippets (see Figure 4) can now be used to write alternatives like "Google and Facebook are *involved in an online privacy row*" (using the third snippet, "the latest *privacy rows involving Facebook and Google*") or "Facebook and Google have raised *privacy concerns*" (using the last snippet, "the *privacy concerns raised by Facebook and Google*").

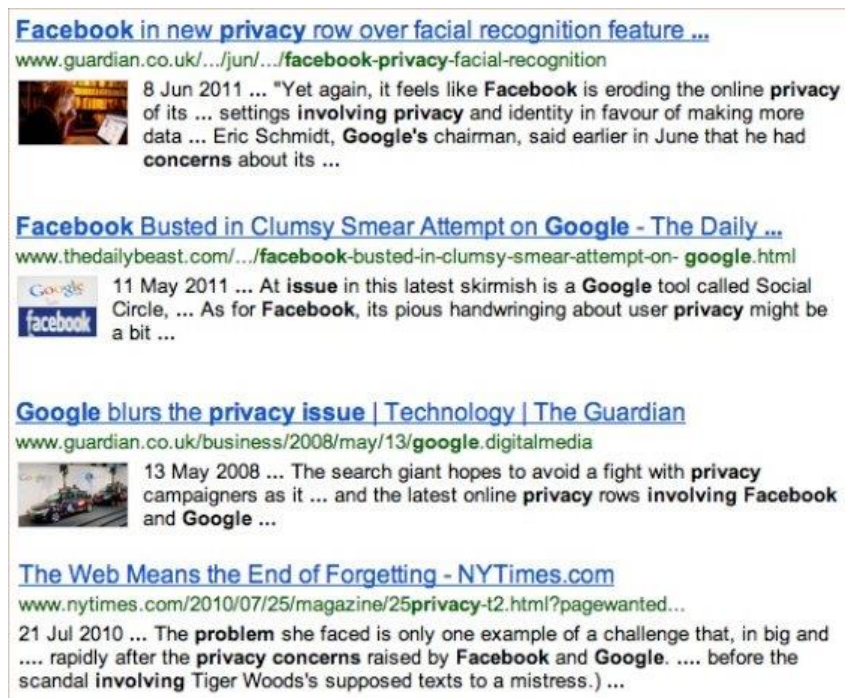


Figure 4. Selected search results for privacy issues involving Google and Facebook.

Following that initial search, the keywords spotted in the original snippets can then be used for a subsequent search. Learners will then be directed to other relevant examples. Entering "online privacy row involving Facebook and Google" (without quotation marks) generates a list of results, among which one formulation clearly stands out (see Figure 5).



Figure 5. Sample search result for *online privacy row involving Facebook and Google*.

2.2.2. Second step: data collection by the instructor, self-correction by the learners

In week one, the students wrote their first in-class essay ("Should society restrict some forms of expression in order to protect its members from violence or hatred?"). The essays were then collected and one to five "untreatable" errors were identified in each of them. All students were then emailed personal charts containing the untreatable errors to be revised and were given one week to correct them on their own using the customized Google search engine. In order to exert some control over their search activities, they were instructed to submit revised passages explaining in detail how they had used Google results to improve their original passages. In week 5, the students wrote a second essay in class ("What do you think about the European Union recently winning the Nobel Peace Prize?"), received their personal charts containing up to five errors and were given one week to submit revised passages explaining the corrections.

3. Findings

3.1. Error analysis

A total of 129 untreatable errors were identified in all 34 essays. The total number of segments improved is 67, equivalent to a success rate of 52%. The number of segments for which the correction was not successful is 36 (28%) and the number of segments for which the correction was partly successful is 16 (12.4%). Six errors (4.6%) were left uncorrected or partly so, and in four cases (3%) the students did not specify whether they had used Google in the correction process. The students' personal charts detailing the corrections made with Google Custom Search reveal six types of searches performed by the students (see Table 2 for details). One way for students to correct their errors is to perform searches on fragments of a non-native-like segment containing an untreatable error. They either initiate a direct correction that they check on Google, or use various approaches (wild card search, word combinations, etc.), and they then use elements featured in the snippets to make the necessary corrections (search type #1, used 70 times). Two other strategies consist in formulating queries after consulting a dictionary (search type #2, used 6 times) or using Google's auto-correct (alternate spelling or wording) to revise a segment (search type #3, used 3 times). In other cases the students decide to perform searches on a whole segment (or syntactically whole fragments of it). In the result snippets, they identify elements of the segment they have to correct which they use to make the necessary changes (search type #4, used 19 times). Yet another strategy consists in entering the whole segment (or syntactically whole fragments of it) in the search box. In the result snippets, although the students do *not* see elements of the segment they have to correct, Google lists articles dealing with their topic. In the snippets of text they then identify what they need to correct themselves (search type #5, used 12 times). Finally, the students sometimes perform keyword searches to which Google responds by listing articles

dealing with their topic. The students then use elements featured in the snippets to correct their segments (search type #6, used 10 times).

Search type #1		
Original segment	Revised segment	Comments
Even if war is no more a reality in Europe, there is no denying that <i>the economical war has remplaced it</i> .	Even if war is no more a reality in Europe, there is no denying that <i>Europe is in an economic war now</i> .	<p>1. I first entered <i>economical war</i> in the search box and Google's auto-correct offered <i>economic war</i> as an alternative.</p> <p>2. I then entered <i>economic war</i> and saw that David Cameron once said <i>Britain is in an economic war</i>. So I used the whole expression instead of my original segment.</p>

Search type #2		
Original segment	Revised segment	Comments
<i>The liberty of expression is necessary in democratic countries but we must warn to violence.</i>	We must take steps to prevent such violence / We must pay attention to violence	I used an online dictionary to check how to say <i>faire attention à</i> in English. I then used GCS to check my correction.

Search type #3		
Original segment	Revised segment	Comments
EU is one of <i>the hugest weapons solder of the world</i> .	EU is one of <i>the biggest weapons soldier of the world</i> .	I entered the segment and Google's auto-correct offered an alternative, <i>EU is one of the biggest weapons soldier of the world</i> .

Search type #4		
Original segment	Revised segment	Comments
Freedom is the backbone of the driving force behind a "good society."	Freedom is the backbone of AND the driving force behind a "good society."	I entered the sentence and found a snippet making me realize that "the backbone of" and "the driving force behind" were two different expressions.

Search type #5		
Original segment	Revised segment	Comments
The newspaper Charlie Hebdo published <i>some comics which critic Islam</i> .	The newspaper Charlie Hebdo published <i>some cartoons that mocked Islam</i> .	I entered the whole passage and saw that <i>cartoons</i> was more appropriate than <i>comics</i> . I saw a better sentence than mine in the first snippet and so I used it.

Search type #6		
Original segment	Revised segment	Comments
The recent scandals in Iraq about prisoners detention.	The Iraq prison abuse scandal.	I entered <i>Iraq scandals detention</i> and found what I needed.

Table 2. Sample search types and comments.

The general coding of errors (see Table 3) reveals that the students are very creative, sometimes combining various search methods (e.g. student #13, error #8), or have an obvious predilection for one type of error correction (e.g. student #5 mainly using search type #1).

Student #	Error #									
	1	2	3	4	5	6	7	8	9	10
1	4 +	X PB3	4 +	X PB2	- PB1	4/5 ±	3 -	- PB1		
2	5 -	1 +	1 +	5 +	- PB1	? ±	1 +	1 +	1 +	
3	1 -	4 +	1 ±	2 -	2 +	4 -	1 -	1 -		
4	1 +	1 -	4 +	1 -	4 +	1 -	1 -			
5	1 +	4 -	??	1 +	1 +	1 -	1 +	1 +	1 +	? -
6	1 -	3 -	1/5 +	1 +	5/1 +					
7	1 +	1 +	1 +	4 -	1 +	1 +	2 -			
8	1 +	1 +	X	4 +	1 -	2 +				
9	4/1 ±	5 +	1 +	? -	1 +	X PB2				
10	6 +	6 +	1 +	6 +	1/6 ±	6 +	1 +			
11	1 +	??	1 -	1 -	1 ±	1 -	1 -	6 +	1 +	
12	4 +	3 +	X	? -	6 ±	5 ±	1 -			
13	? -	1 ±	1 -	1 -	1 +	1 -	1 ±	1/4/5/2 +	2 +	

14	4 +	1 ±	5 +	5 +	4/1 ±	6 ±	6 ±	1 +	4/1 +	6 +
15	1 +	X	1 +	1 +	1 ±					
16	? +	1 +	1 +	? -	1 +	1 +	1 +	1 +		
17	4/5 +	5/1 -	4 +	4 +	??	? -	??	4/1 ±		

Table 3. General error coding.

Note: The errors were identified in essays 1 and 2. To correct each error, the students performed various search types. Each search type number (1 to 6) is followed by a positive (+), a negative (-), or a plus-minus (±) sign depending on whether the correction was successful, not successful, or partly successful. The students sometimes combine various search methods, hence the succession of numbers in some cases (cf. student #13, error #8). A question mark (?) is used when the correction is not explained although a Google search was performed. Two question marks (??) are used when the correction is not explained and there is no indication that a Google search was performed, and a cross (X) is used when the segment is left uncorrected. PB1 is used when students initiate a correction after entering the whole segment in the search box and say they do not know how to use the results. PB2 is used when students say they do not know what query to formulate, and PB3 when they see elements in the search results but do not know how to use them.

3.2. Feedback on Google-driven language learning

Sixteen completed questionnaires were returned via email (the responses to the seven 5-point Likert-scale questions are given in Table 4). Questions 1 to 4 show that a majority of students felt comfortable with the use of basic Google search options. Question 5 indicates that the students view Google use as a good way to correct their errors and improve their English, and question 6 indicates that a majority view it as a good way to find native-like formulations in the search results. However, only nine students said that they intended to use it in the future for linguistic purposes. In the answers they provided to the open-ended questions the students explained in more detail what they liked about Google search but also raised a number of issues.

Eight students explained that the main difficulty for them was to find appropriate ways to formulate their queries. They sometimes found it difficult to identify alternatives to their non-native-like formulations because they couldn't think of any other word or expression to enter in the search box. Three of them argued that in order to use Google effectively, it is necessary for them to know what they are looking for, which implies knowing what is wrong in a segment underlined by the teacher. Other students explained that they liked how Google Custom Search could be used to discover word combinations and noteworthy formulations. One for example said she enjoyed using Google to check the idiomaticity of formulations by using quotation marks around search strings. Another student liked the idea of restricting searches to specific websites, while another one enjoyed making serendipitous discoveries when scouring the snippets of text. Two of them, however, said that they found it more effective to read newspaper articles to find noteworthy formulations. Three others said they sometimes found it tedious to have to use a search engine to correct their errors while they had other, more effective tools at their disposal (grammar handbooks, dictionaries, etc.). Two of them in fact said that they used Google Custom Search in conjunction with online dictionaries. Two others confessed they found it difficult to adapt the search results to have them fit into their original sentences. They also said it was a little frustrating to find ideas that did not exactly express the ideas they had in mind

although they constituted obvious alternatives to their original non-native-like formulations. Three students said that they sometimes felt overwhelmed with the results and simply did not know what to make of them.

Closed questions (5-point Likert scale)	1 strongly disagree	2 disagree	3 neither agree nor disagree	4 agree	5 strongly agree
1. I find it easy to use Google search options.	0	12,5 %	6,25 %	31,25 %	50 %
2. I can differentiate between searches using quotation marks and searches not using quotation marks.	0	6,25 %	6,25 %	18,75 %	68,75 %
3. I know how to use wild cards in my queries.	0	6,25 %	25 %	31,25 %	37,5 %
4. I know how to use keywords in my queries.	0	6,25 %	0	43,75 %	50 %
5. I think that using Google Custom Search is a good way to correct my errors and improve my English.	0	6,25 %	12,5 %	68,75 %	12,5 %
6. I think that using Google Custom Search is a good way to find native-like formulations used in the press.	0	6,25 %	12,5 %	37,5 %	43,75 %
7. I intend to use Google (Custom Search) in the future for linguistic purposes.	6,25 %	6,25 %	31,25 %	50 %	6,25 %

Table 4. Responses to the 5-point Likert scale questions.

4. Discussion

The purpose of this study was to document the way in which internet searches can act as “a tool helping second language writers make decisions about their writing” (Acar et al. 2010: 6). It can now be argued that using Google Custom Search and restricting searches to information websites is a way to increase the reliability of raw online data in so far as it maximizes the students’ chances to be exposed to grammatically accurate English. For teachers who generally choose to reformulate “untreatable” passages in their students’ papers, this can surely “help relieve [them] of the need to act as proofreading slaves” (Milton 2006: 125). One student for example said he found that Google was a good way to go about correcting his errors when the teacher was not around. So it seems that Google acts as a gateway to a repository of formulations that they can choose by themselves instead of relying on their teacher to provide alternatives. However, some students confessed they sometimes felt overwhelmed with the results or did not know how to formulate their queries. Several studies bearing on corpus use have reported that students feel frustrated (Lavid, 2007) or overwhelmed by considerable amounts of data (Ädel, 2010; Johns et al., 2008; Liu & Jiang, 2009; Kennedy & Miceli, 2010). Others said they found it difficult to formulate corpus queries and various studies also report on the same problem (Ma, 1994; Kennedy & Miceli, 2001; Miceli & Kennedy, 2002; Sun, 2003; Cheng et al., 2003; O’Sullivan & Chambers, 2006; Hafner & Candlin, 2007). Others still explained that analyzing Google output was no easy task, another recurring problem in studies documenting learner analysis of concordancer output (Ma, 1994; Bowker, 1998; Kennedy & Miceli, 2001; Miceli &

Kennedy, 2002; Cheng et al., 2003; Sun, 2003; Yoon & Hirvela, 2004; Lavid, 2007; Johns et al., 2008; Boulton, 2009b; Liu & Jiang, 2009;). The challenge for teachers is thus to provide learners with appropriate training and make sure they are “adequately equipped” (Kennedy & Miceli, 2001: 81) before exploring corpora on their own.

When working on Google output, teachers are also faced with the difficult task of encouraging learners to assimilate the formulations they identify because they will inevitably risk being stigmatized for working too closely with their sources and accused of plagiarism. Donahue (2008) points to this major problem that language teachers are grappling with and makes the case that copying should nonetheless not be castigated as plagiarism:

How do we determine at what point something is “owned”? [...] Students come to learn and we want them to appropriate knowledge and be comfortable in the discourse of the field; at what point does something —class discussion, a professor’s discourse— no longer get cited? (p.102)

We can indeed wonder what students are supposed to make of what they read in their own time. Where to draw the line between what ought to be copied and what ought not to be? If we take a sentence like *Human cloning may be the thin end of the wedge*, it is difficult to decide whether or not, if a student reads it in a news article and subsequently uses it in an essay, the accusation of micro-plagiarism is justified. Research on the subject (e.g. Grossberg 2008; Murray 2008; Emerson 2008; Senders 2008; Bloom 2008; Bloch 2008; Adler-Kassner et al. 2008) explains that accusations of plagiarism are most often sweeping generalizations of otherwise skillful use of appropriated material. It may not be really fair to accuse students who borrow and use without referencing of intellectual theft as, when copying, they are learning to situate their discourse in relation to others’. Within the framework of this experiment, it has been shown that selective reading of Google results is a way for EFL students to write better English by skillfully copying and integrating prefabricated ideas and language into their own essays. The students never transfer extensive verbatim passages to their essays but select relevant multi-word fragments and the result is language hybridity (i.e. a combination of material identified in Google snippets and personal utterances). And while it is difficult to decide whether or not Google search is a tool helping EFL learners gain in grammatical accuracy, it is a way for them to find alternatives to their non-native-like formulations. The keyword search, used by many students, is particularly effective to that end.

For example, seeking to improve *a cartoonist who draws Mahomet*, student #10, who is writing about a scandal which recently flared up in France, enters *who draws Mahomet* and realizes that the result snippets feature the word *cartoon*. He then performs a search with a series of three keywords, *charlie hebdo cartoon* (*Charlie Hebdo* being the name of the newsweekly which originally published the controversial cartoons), and finds a *satirical weekly publishes cartoons of the Prophet Mohammed*, which he decides to use to rephrase his original idea. The same student, trying to improve *The contestation wave in Middle East against a disgusting film*, explains that he knew that *contestation wave* was incorrect yet could not come up with anything better when writing his in-class essay. So he explains that entering *protesters middle east* in the search box resulted in Google producing a link to a *New York Times* article whose title (“Protests spread in the Middle East”) he used to correct his sentence.

A successful keyword search is thus arguably the first step on the road to writing clarity. Yet it is obvious that it does not solve other problems that the students also have to tend to. When the same student uses *publishes* (instead of *published*) to refer to a scandal which erupted a few months ago, it is difficult to decide whether or not he is

aware that *spread*, which is transferred to the original essay, is used in the present tense and not the simple past in the title. In a word, while it is obvious that the students generally do recognize what they need when they see it in Google results, they are not always successful at accommodating the syntax of the segments they seek to weave into or substitute for their original written productions.

Student #1, for instance, writing about free speech and asked to improve *If the society do not established a red border, it can be a vicious circle*, explains that he doesn't know how to use Google to improve the sentence. He performs a search with the entire sentence and doesn't break it down to explore meaningful elements (e.g. *society establish a red border*) to find out if they are combined in a particular way or if Google lists articles dealing with the topic, featuring expressions that can be borrowed. In most cases, this shows that the students must already have a repository of alternatives they can use to perform their searches. These alternatives don't need to be whole syntactical segments but can be collocations or single lexical items that the student is not sure how to articulate in a complete sentence. For instance, if students realize that *establish a red border* is incorrect but know the expression *draw the line*, they can perform a search meant to find out how it is contextualized in the press. Furthermore, in order to maximize their chances of finding what they need, the students must also be able to self-correct a number of treatable errors first (i.e. write *if society does not establish* and not *if the society do not established* in the example). Indeed, Google is more likely to produce relevant examples when searches are performed with grammatically accurate, albeit awkwardly formulated, segments. In other cases, it was found that the students did make changes but on some elements only. In other words, they did not see what was wrong in their sentences. For example, student #5, asked to improve *freedom of expression is being turned into ideological injures* only corrects *injures*, opting for *injuries*, unaware that *ideological injuries* is an unlikely collocation and that it is in fact the whole idea that needs to be reformulated.

5. Conclusion

The web should not be dismissed as an unreliable source of data. Although it is arguably not a corpus, EFL learners can nonetheless profitably use Google for quick and easy access to authentic language in the form of selected passages from a great number of articles. In that sense, Google output is very much adapted to students who need to keep up with world events and whose ultimate goal is to emulate the language of the press. Depending on their competence, it is a vast repository of formulations that they can identify and borrow for further use in their own writing. Students can be given a significant linguistic boost if encouraged to plunder formulations featured in Google results. Such an approach implies for the students to go through an initial stage of teacher-controlled imitation (or micro-plagiarism) because initially copying native speakers will, arguably, make it possible to emulate them.

The rationale behind customizing a search engine to explore linguistic material from a selection of online newspapers is in keeping with Tribble's recommendation that the most useful corpus for EFL learners is "the one which offers a collection of expert performances in genres which have relevance to the needs and interests of the learners. Collections of relevant expert performances will exemplify the results of the desired forms of language behavior that learners are trying to achieve" (1997: n. pag.). The main objection raised by a certain number of students who took part in this study was that they sometimes felt overwhelmed with search results or could not think of ways to formulate their queries. Further research could thus profitably focus on how best to train EFL learners to use Google search results in order to self-edit.

Websites

1. <http://www.webcorp.org.uk/live>
2. <http://www.kwicfinder.com>
3. <http://webascopus.org>
4. <http://www.greenstone.org>
5. <http://www.google.com/cse>

References

- Acar, A., Geluso, J. & Shiki, T. (2011). How can search engines improve your writing *CALL-EJ*, 12 (1): 1-10.
- Ädel, A. (2010). Using corpora to teach academic writing: challenges for the direct approach. In: Campoy-Cubillo, M. C., Belles-Fortuño B. & Gea-Valor M. L. (eds). *Corpus-based Approaches to ELT*. London: Continuum, 39-55.
- Adler-Kassner, L., Anson, C.M. & Howard, R.M. (2008). Framing plagiarism. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 231-247.
- Bates, L., Lane, J., & Lange, E. (1993). *Writing clearly: responding to ESL compositions*. Boston: Heinle & Heinle.
- Benson, P. (2001). *Teaching and researching autonomy in language learning*. Harlow: Pearson Education.
- Bergh, G. (2005). Min(d)ing English language data on the web: what can Google tell us? *ICAME journal*, 29: 25-46.
- Bhatia, T. K. & Ritchie, W. C. (2009). Second language acquisition: research and application in the information age. In: Ritchie, W.C. and Bhatia, T.K. (eds.), *The new handbook of second language acquisition*. Bingley: Emerald, 545-565.
- Bitchener, J. (2008). Evidence in support of written corrective feedback. *Journal of second language writing*, 17 (2): 102-118.
- Bitchener, J., Young, S. & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of second language writing*, 14: 191-205.
- Bloch, J. (2008). Plagiarism across cultures: is there a difference? In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 219-231.
- Bloom, L. Z. (2008). Insider writing: plagiarism-proof assignments. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 208-219.
- Boulton, A. (2009a). Data-driven learning: reasonable fears and rational reassurance. *Indian journal of applied linguistics*, 35(1):81-106.
- Boulton, A. (2009b). Corpora for all? Learning styles and data-driven learning. In: M. Mahlberg, González-Díaz, V. & C. Smith, C. (eds.), *Proceedings of the 5th Corpus Linguistics Conference*. Liverpool: UCREL.
- Boulton, A. (2012). What data for data-driven learning? *EUROCALL 2012: Proceedings*. Nottingham: The University of Nottingham.

- Bowker, Y. (1998). Using specialized monolingual native-language corpora as a translation resource: a pilot study. *Meta*, 4: 631-651.
- Chambers, A., Conacher J. & Littlemore J. (eds.) (2004). *ICT and language learning: integrating pedagogy and practice*. Birmingham: University of Birmingham Press.
- Chandler, J. (2003). The efficacy of various kinds of error feedback for improvement in the accuracy and fluency of student writing. *Journal of second language writing*, 12(3): 267-296.
- Cheng, W., Warren, M., & Xun-feng, X. (2003). The language learner as language researcher: putting corpus linguistics on the timetable. *System*, 31: 173-186.
- Chen, Y. H. (2004). The use of corpora in the vocabulary classroom. *The internet TESL journal*, 10(9): n. pag.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language learning and technology*, 14(2): 30-49.
- Chinnery, G. M. (2008). You've got some GALL: Google-assisted language learning. *Language learning and technology* 12(1): 3-11.
- Clerehan, R., Kett, G. and Gedge, R. (2003). Web-based tools and instruction for developing it students' written communication skills. In: *Exploring Educational Technologies Conference Proceedings*. Monash University. Retrieved from http://www.monash.edu.au/groups/flt/eet/full_papers/clerehan.pdf. Last accessed 25/09/2014.
- Conroy, M. (2010). Internet tools for language learning: university students taking control of their writing. *Australasian Journal of educational technology*, 26(6): 861-882.
- Donahue, C. (2008). When copying is not copying: plagiarism and French composition scholarship. In: Eisner, C. and Vicinus, M. (eds), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 90-103.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Emerson, L. (2008). Plagiarism, a Turnitin trial, and an experience of cultural disorientation. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 183-195.
- Ferris, D. R. (2004). The "grammar correction" debate in L2 writing: where are, and where do we go from here? (and what do we do in the meantime...?). *Journal of second language writing*, 13 (1):49-62.
- Ferris, D. R. and Roberts, B. (2001). Error feedback in L2 writing classes: how explicit does it need to be? *Journal of second language writing*, 10(3): 161-184.
- Ferris, D. R. (1999). The case for grammar correction in L2 writing classes: a response to Truscott (1996). *Journal of second language writing*, 8(1): 1-11.
- Fletcher, W. H. (2004). Making the web more useful as a source for linguistic corpora. In: Connor, U. and Upton, T. (eds.), *Applied corpus linguistics: A multidimensional perspective*. Amsterdam: Rodopi, 191-205.
- Gaskell, D. & Cobb, T. (2004). Can learners use concordance feedback for writing errors? *System*, 32(3): 301-319.
- Grossberg, M. (2008). History and the disciplining of plagiarism. In: Eisner, C. and Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 159-173.

- Guo, S. & Zhang, G. (2007). Building a customised Google-based collocation collector to enhance language learning. *British journal of educational technology*, 38(4): 747-750.
- Hafner, C. A. & Candlin, C. N. (2007). Corpus tools as an affordance to learning in professional legal education. *Journal of English for academic purposes*, 6(4): 303-318.
- Hendrickson, J. M. (1978). Error correction in foreign language teaching: recent theory, research, and practice. *The modern language journal*, 62(8): 387- 398.
- Holec, H. (ed.) (1988). *Autonomy and self-directed learning: present fields of application*. Strasbourg: Council of Europe.
- Holec, H. (1980). Learner training: meeting needs in self-directed learning. In: Altman, H. B. & James, C. V. (eds.). *Foreign language learning: meeting individual needs*. Oxford: Pergamon, 30-45.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31(2): 217-230.
- Jarvis, H. (2004). Investigating the classroom applications of computers on EFL courses at higher education institutions in the UK. *Journal of English for academic purposes*, 3(2): 111-137.
- Johansson, S. (2009). Some thoughts on corpora and second-language acquisition. In: Aijmer, K. (ed.). *Corpora and language teaching*. Amsterdam: John Benjamins, 33-44.
- Johns, T. (1988). Whence and whither classroom concordancing? In: Bongaerts, P., De Haan, P., Lobbe, S. & Wekker, H. (eds.), *Computer applications in language learning*. Dordrecht: Foris, 9-27.
- Johns, T. (1990). From printout to handout: grammar and vocabulary teaching in the context of data-driven learning. *CALL Austria*, 10: 14-34.
- Johns, T., Lee H. C. and Wang L. (2008). Integrating corpus-based CALL programs in teaching English through children's literature. *Computer Assisted Language Learning*, 2: 483 -506
- Johnson, A. (2004). Creating a writing course utilizing class and student blogs. *The internet TESL journal* 10(8).
- Kennedy, C. & Miceli, T. (2001). An evaluation of intermediate students' approaches to corpus investigation. *Language Learning and Technology*, 5: 77-90.
- Kennedy, C. & Miceli, T. (2010). Corpus-assisted creative writing: introducing intermediate Italian learners to a corpus as a reference resource. *Language learning and technology*, 14(1): 28-44.
- Kenworthy, R. C. (2004). Developing writing skills in a foreign language via the internet. *The internet TESL journal*, 10(10).
- Kepner, C. G. (1991). An experiment in the relationship of types of written feedback to the development of second language writing skills. *The modern language journal*, 75(3): 305-313.
- Kilgariff, A. (2001). Web as corpus. In: Rayson, A., Wilson, T., McEnery, A., Hardie & Khoja, S. (eds.), *Proceedings of the corpus linguistics 2001 conference*. Lancaster: UCREL, 342-344.
- Krajka, J. (2000). Using the internet in ESL writing instruction. *The Internet TESL Journal*, 6(11).
- Lavid, J. (2007). Contrastive patterns of mental transitivity in English and Spanish: a student-centred corpus-based study. In: Hidalgo, E. Quereda, L. & Santana J. (eds.). *Corpora in the foreign language classroom*. Amsterdam: Rodopi, 237-252.

- Liu, D. & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93: 61-78.
- Ma, B. K. C. (1994). Learning strategies in ESP classroom concordancing: an initial investigation into data-driven learning. In Flowerdew, J. & Tong, A. (eds.). *Entering Texts*. Hong Kong: Language Centre, The Hong Kong University of Science and Technology, 197-214.
- Mansor, N. (2007). Collaborative learning via email discussion: strategies for ESL writing classroom. *The Internet TESL Journal*, 13(3).
- McCarthy, M. (2008). Accessing and interpreting corpus information in the teacher education context. *Language Teaching*, 41(4): 563-574.
- Miceli, T. & Kennedy, C. (2002). An Apprenticeship with the CWIC Corpus: a tool for learner writers in Italian. In: Kennedy, C. (ed.) *Proceedings of Workshop Innovations in Italian Teaching*. Brisbane: Griffith University, 83-94.
- Milton, J. (2006). Resource-rich web-based feedback: helping learners become independent writers. In: Hyland, K. and Hyland, F. (eds.), *Feedback in second language writing*. New York: Cambridge University Press, 123-139.
- Murray, L. J. (2008). Plagiarism and copyright infringement: the cost of confusion. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 173-183.
- O'Keeffe, A., McCarthy, M. & Carter, R. (2007). *From corpus to classroom: language use and language teaching*. Cambridge: Cambridge University Press.
- O'Sullivan, Í. & Chambers, A. (2006). Learners' writing skills in French: corpus consultation and learner evaluation. *Journal of Second Language Writing*, 15: 49-68.
- Palfrey, J., & Gasser, U. (2008). *Born digital. Understanding the First Generation of Digital Natives*. New York: Basic Books.
- Renouf, A. (2003). WebCorp: providing a renewable data source for corpus linguists. *Language and computers*, 48: 39-58.
- Robb, T. (2003a). Google as a quick 'n dirty corpus tool. *TESL-EJ*, 7(2).
- Robb, T. (2003b). Google as a corpus tool? *ETJ Journal*, 4(1).
- Rundell, M. (2000). The biggest corpus of all. *Humanising language teaching*, 2(3).
- Schroeder, A., Minocha, S., & Schneider, C. (2010). The strengths, weaknesses, opportunities and threats of using social software in higher and further education teaching and learning. *Journal of Computer Assisted Learning*, 26: 159-174.
- Senders, S. (2008). Academic plagiarism and the limits of theft. In: Eisner, C. & Vicinus, M. (eds.), *Originality, imitation, and plagiarism: teaching writing in the digital age*. Michigan: The University of Michigan Press, 195-219.
- Shei, C. (2008a). Discovering the hidden treasure on the internet: using Google to uncover the veil of phraseology. *CALL*, 21(1): 67-85.
- Shei, C. (2008b). Web as corpus, Google, and TESOL: a new trilogy. *Taiwan Journal of TESOL*, 5(2): 1-28.
- Sun, Y. (2003). Learning process, strategies and web-based concordancers: a case study. *British journal of educational technology*, 34(5): 601-613.

Tribble, C. (1997). Improvising corpora for ELT: quick-and-dirty ways of developing corpora for language teaching. In: Melia, J. & Lewandowska-Tomaszczyk, B. (eds.) *PALC 97 Proceedings*, Lodz: Lodz University Press.

Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language learning*, 46(2): 327-369.

Widdowson, H. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21(1): 3-25.

Wu, S., Franken, M., & Witten, H. (2009). Refining the use of the web (and web search) as a language teaching and learning resource. *CALL*, 22(3): 249-268.

Yoon, H. (2008). More than a linguistic reference: the influence of corpus technology on L2 academic writing. *Language learning and technology*, 12(2): 31-48.

Yoon, H. & Hirvela, A. (2004). ESL student attitudes toward corpus use in L2 writing. *Journal of second language writing*, 13: 257-283.

Zamel, V. (1985). Responding to student writing. *TESOL Quarterly*, 19(1): 79-97.