

Google-Informed Patter-Hunting and Pattern-Defining: Implication for Language Pedagogy

Ebrahim Panah¹, Melor Md Yunus¹ & Mohamed Amin Embi¹

¹ Faculty of Education, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia

Correspondence: Ebrahim Panah, Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia. Tel: 60-17-678-6301. E-mail: panah_ebrahim@yahoo.com

Received: December 18, 2012 Accepted: January 11, 2013 Online Published: February 28, 2013

doi:10.5539/ass.v9n3p229

URL: <http://dx.doi.org/10.5539/ass.v9n3p229>

Abstract

The use of the Web as a corpus and Google as a concordancer, has been regarded as one of the promising areas that has a potential for revolutionizing language pedagogy in general, and second language (L2) writing, in particular. More specifically, it is believed that the functions of Google-Informed Pattern-Hunting (GIPH) and Google-Informed Pattern-Defining (GIPD) can promote natural L2 writing through Discovery Learning (DL) and Data Driven Learning (DDL), however, these advantages have mostly been given lip services than tested with first hand empirical studies, and only more recently some studies have been undertaken in this vein. Focusing on L2, this article explored how and to what extent this great potential of GIPH and GIPD has been recognized by reviewing the related studies, thereby some factors and themes (such as Learning Style, Training, Naturalness, Tidiness, Speed, Number of Retrieval, and Proficiency) have been extracted and elaborated on. However, due to the novelty of the area, the themes are mostly the outcome of researchers' descriptions and interpretations than empirical studies. The inclusion criteria for the present review were studies that focus on the application of the Web as a corpus and Google as a concordance for language learning and L2 writing based on researchers' and learners' evaluation of it. Seven studies included in the present review show that learners' use of GIPH and GIPD champions the promotion of their language learning and L2 writing, providing that proper training and scaffolding are provided. Future studies are also recommended based on the gaps and deficiencies identified in the reviewed researches.

Keywords: GIPH, GIPD, GALL, concordancer, web corpus, DL, DDL, natural

1. Introduction

1.1 The Web Corpus and Language Patterns

It is widely accepted that the Web corpus has merely been employed as linguistic datato corroborate linguists' intuitions about the frequency of occurrences of individual words, phrasal verbs, collocations, and idioms (Wierzbicka, 2009). However, more recently, as some researchers (Afendi Hamat & Mohamed Amin Embi, 2009; Chambers et al., 2011; Comelles et al., 2012; Conroy, 2010; Geluso, 2011; Park & Kinginger, 2010; Sha, 2010; Shei, 2008; Wu, Franken & Witten, 2009 ; Wu, Witten & Franken, 2010; Yoon, 2011) have discerned, the Web is also a particularly valuable source of authentic, natural and contextualized language patterns (concordances, collocations, colligations, chunks, phrases, idioms, Formulaic Sequences, etc)(definitions of pattern put forward by , Kennedy & Miceli, 2010; Wood, 1981) and on-line dictionaries invaluable for pedagogical applications including L2 writing. From the point of view of cognitive linguistics, what is perceived as authentic and natural language is in connection with phraseology, based on frequency of occurrences, rooted in usage-based theory of language acquisition (Bybee, 2006; Geluso, 2011). The Web, simply owing to its absolute size, is a well-suited recourse to provide insights into the frequency of occurrence of patterns, in turn natural language. Needless to say that, linguistic patterns are abundant in both written and spoken discourse (Bybee, 2006; Erman & Warren, 2000), hence for the ESL/EFL learner seeking for frequent linguistic patterns in L2 writing, the Web corpus can be a valuable tool (Chambers, Farr & Riordan, 2011 ; Comelles et al., 2012; Conroy, 2010; Geluso, 2011; Sha, 2010; Stapleton & Radia, 2009). To access the Web corpora, although there are many search engines (such as Google, Alta Vista, Yahoo, MSN, etc), Google has appeared to have become a popular search engine for this function, mainly because it is capable of doing much more than just ordinary search. As Chinnery (2008) expounds, Google "has the capacity to do much more than simply facilitate basic Boolean searches" (p. 3). He

then introduced the term Google-Assisted Language Learning (GALL), which later Shei (2008) adapted to the application of the Web corpus. Thus, the Web as corpus and Google as concordancer can be a great source of natural language patterns for writing improvement (Conroy, 2010; Geluso, 2011; Sha, 2010; Shei, 2008). Therefore, Google as a giant search engine can be used as a learning pedagogy application in the process of PH (finding appropriate pattern as a result of Google consultation) and PD (defining one's pattern through Google consultation, based on frequency) (PH & PD were put forth by Kennedy & Miceli, 2010). Since Google is the most popular search engine, representing the Web perfectly, in this review study Google-driven corpus and the Web corpus will be used interchangeably.

Google, as a dynamic corpus, accesses a bulk of ever-growing and ever-changing Internet resources. For example, the pattern *perfect balance* occurred in the British National Corpus 37 times, in AltaVista search engine 8,020,000, and Google search resulted 11, 1000,000 hits. Accordingly, Shei (2008, p. 70) argues, "Google can offer solutions to many of the research questions in phraseology which even a billion-word corpus can hardly handle". Thus, Google can keep abreast of a wide range of genuine language patterns invaluable to ESL writing course. In addition, its user-friendly interface and high-speed search are appreciated by numerous users (Conroy, 2010; Sha, 2010). Google also has the power of offering correct spelling, a characteristics which is of paramount importance, but most of other concordancers lack (Islam & Inkan, 2009; Sha, 2010). Additionally, Google can play the role of a huge dictionary encompassing approximately all English vocabulary items, providing their definitions, synonyms and antonyms as well as contexts of usage. To illustrate, the learner can simply type the word 'define' succeeded by his/her word in question.

Finally, another benefit of Google search is that the TEFL/TESL (Teaching English as a Foreign/Second Language) community as well as the research community can take advantage of such colossal and ever growing amounts of the Web corpus data via Google. For example, Google can play the role of concordancer (Geluso, 2011; Sha, 2010; Smith, 2011). Hence, an ESL/EFL learner, dithering about the accuracy and naturalness of two resembling patterns such as *discuss about the issue* vs. *discuss the issue* or *the study concentrates on the issue of* vs. *the study focuses on the issue of*, can easily check the frequency of any given pattern on the Web by performing a simple search of the pattern in double quotation marks (Geluso, 2011; Stapleton & Radia, 2009). To illustrate with examples, consider the following two patterns: (1) *discuss about the issue* (2) *discuss the issue*. In November 2012, a search of both patterns, enclosed in double quotation marks on Google Scholar, retrieved approximately 79 hits for the first pattern and almost 41,800 retrievals for the second. Also, two other patterns such as: (1) *the study concentrates on* and (2) *the study focuses on* using Google Scholar, regarding the former the Google scholar result was 4,430, while that of the latter was 32,300. A native speaker of English is more likely to perceive the second patterns as more natural due to its frequency of occurrence and exposure, corroborating Google search results (Geluso, 2011). However, although direct uses of the Web as corpus and Google as concordancer have their naysayers, arguing that the Web corpus is not neat and tidy (Fletcher, in press; Wu et al., 2009 & 2010), their advantages outweigh their disadvantages, as they provide ever-growing natural linguistic data. Furthermore, Google Scholar as offers natural patterns produced by native speakers of English or expert writers (Sha, 2010), can compensate for the Web corpus deficiency in terms of tidiness.

Nevertheless, although studies on learner-generated corpora have demonstrated L2 learners' misuse, overuse, and underuse, of linguistic patterns (Hong, Abdul Rahim, Kim Hua & Salehuddin, 2011; Karimah Yunus & Swad Adwab, 2011; Yoon, 2011), or use of unnatural patterns (Geluso, 2011; Gilmore, 2009) in L2 writing, empirical studies on GIPH and GIPD, albeit with mixed results, have rarely been reported in the literature.

2. Literature Review

2.1 Importance of GIPH and GIPD in Language Learning and ESL Writing

In terms of language learning and L2 writing, it is stated that the Web armed with Google, based on Shei's (2008) ground-breaking work, allows ESL/EFL learners to study particular words and patterns to check whether the text they have produced represents standard and natural language usage (GIPD) (Geluso, 2011; Stapleton & Radia, 2009). It also enables learners to explore appropriate patterns and word combinations (GIPH) (Sha, 2010; Yoon, 2011). Learners can examine the words that most frequently, precede, follow or occur between specific words or patterns, based on frequency of occurrences (Wu, 2010; Geluso, 2011). In this section, some studies which have been conducted in this vein (Conroy, 2010; Geluso, 2011; Sha, 2010; Shei, 2008; Stapleton & Radia, 2009; Wu 2010; Wu et al., 2010), are reviewed and discussed.

2.2 Past Studies on GIPH

Based on the premise that Google allows ESL learners keep abreast of wide range of naturally occurring patterns, based on frequency occurrences, and thereby discover and hunt appropriate patterns in the process of language

learning and L2 writing, very few empirical studies, if any, have been reported in this vein in literature.

In a related study, Wu (2010) studied the students' use of corpora and the Web-based corpus, through an off-line collocation learning system, constructed based on Google, to examine how they use the function of pattern-hunting to expand their text. The findings of his observation and questionnaires show that three out of 12 students did some kind of pattern-hunting, and the result was promising. He reported that syntactic errors, erroneous sentence structures, and imperfect sentences were prevalent throughout their work. He argues "Because of the constraints of the topic—themselves and their family— and their limited language ability, their writing exhibited a narrow range of vocabulary and few idiomatic expressions. For example, the four most common words used were like, come, want and live. Sentence structure was simple and basic" (Wu, 2010, p.119). He concludes that **proficiency** is an issue which is a barricade making less proficient learners sluggish in relation to PH and text generation and confining their activities to text revision and correction.

However, this study was not on direct use of Google for pattern-hunting, but it paves the way for the further study. In addition, the number of participants conducting the function of GIPH was 3. Furthermore, although the off-line system that he constructed is neat and tidy in comparison with the language data offered by search engines such as Google, the system is very limited in terms of accessibility, because Google allows and enables ESL/EFL learners keep abreast of wide range of ever-growing, authentic, and natural language patterns.

2.3 Past Studies on GIPD

Some researchers have conducted studies on Google-driven corpus for language learning and ESL/EFL writing course specifically on GIPD (Conroy, 2010; Geluso, 2011; Sha, 2010; Shei, 2008; Stapleton & Radia, 2009).

The researches demonstrate that GIPD can assist ESL learners in editing the patterns in question through The Web corpus consultation, based on frequency of occurrences on the Web, leading to natural patterns (Geluso, 2011; Gilmore, 2009; Stapleton & Radia, 2009). Geluso (2011, p.1) argues that "Corpora and data-driven learning can offer useful insights into frequent patterns of naturally occurring language to second/foreign language learners who, unlike native speakers, are not privy to a lifetime of input and fine-tuning". He also adds that "native speakers perceive phrases that generated more results in Google searches to be more natural". Therefore, Google is a useful tool for keeping abreast of natural languages in the process of defining and correcting language patterns. Some studies, with mixed results, conducted in this vein are as follow:

Conroy (2010) studied the internet tools including Google for language learning and writing improvement. His motivation was Australian universities' challenges in terms of teaching EAL (English as an additional language) students. He selected four cohorts of students with different language proficiencies (165 students). Following training in online corpora and GALL tools and techniques, students used the function of PD. The results reveal that there is little evidence of universities' engagement with internet-based corpus tools and techniques for language promotion. Also, the findings of pre- and post training questionnaires and interviews show that students perceive concordancing and GALL consultation useful for language learning and writing improvement. In addition, he reported that "by inspecting frequently occurring word combinations the learner has improved his/her writing in a data-driven learning process" and "helped the student to acquire the means and confidence to self-edit in the future" (2010, p. 867), that is, they autonomously used concordances and Google for improving their writing assignment. As he points out, care should be taken that regarding **learning styles** students differ. He adds "The propensity for students to engage with GALL and concordancing might be a function of learning style" (p. 880). He concludes that unfamiliarity with inductive discovery learning by some students might be a factor in being reluctant and less impressive in using corpus. Nevertheless, yet this study has some limitations. For example it did not address the issue of the naturalness of GIPD and also the function of GIPH was not dealt with.

Correspondingly, Shei (2008) conducted a study on the use of Google to uncover the veil of phraseology. His motivation was the Achilles' heel of none Web-based corpora in justifying some patterns as boldfaced: "a baby boy **positively identified by DNA testing decades after he died** when the Titanic sank **has now been named as** someone else" (2008, p. 67). He found that Stubbs' (2002) Extended Lexical Unit and Hoey's (2005) Lexical Priming theory were corroborated in terms of the prevalence of language in pattern forms. He also discovered that Google search result can account for the patterns in question. He put forth a method, based on reducing the fragment on word at a time and keeping the record of the frequency of the remaining fraction, useful for both for DDL (data driven learning) and linguistic purposes.

Similarly, Sha (2010) researched the use of Google in comparison with British National Corpus (BNC) for L2 writing improvement. His study aimed at examining technical obstacles in using corpora exploring the reason why the DDL approach was unpopular in the process of language teaching and learning. In this qualitative study he observed and examined two types of learners' writing samples namely, TEFL writing classes and English

learning the Web sites. His samples patterns included bi-grams, tri-grams, and four-grams. He used Google advanced search with preferences and BNC. He found that Google retrieval is faster than BNC and yields far more results either in number or in comprehensiveness. He states that “the dynamic corpus or search-engine-based corpus is superior in usability, search speed, the number of solutions and above all, preference investigations” (2010, p. 377). He also shows that Google has spellchecker which BNC lacks. He concludes that there is “a strong evidence that static corpora are losing ground to the Web corpora” (2010, p. 390). Thus, he reminds us of the outstanding capability of Google as a concordancer which can take the place of conventional corpora in terms of GIPH and GIPD, there by promoting DDL and the Web corpus consultation.

Correspondingly, Geluso’s (2011) study focused on the naturalness of GIPD, based on the frequency of occurrences on the Web. He selected 25 Japanese EFL learners and set them to write essays about nine paragraphs. Following training, he got them to Google draft their essays and correct their erroneous patterns. Then he recruited 4 native speakers of English to blind rate learners’ Google-informed and non Google-informed patterns, in terms of their naturalness. The result of his study strongly suggests that by using the Web as a corpus and Google as a concordancer, students can improve the naturalness of their writing.

Accordingly, a descriptive study on the use of technology such as the Web corpus tools for writing by Stapleton & Radia (2009) strongly suggests that Google-driven corpora can be valuable source of information for ESL learners in the process of their ESL writing course, provided that they are trained in Google consultation. In relation to PD and correction, they state that “during the writing process, students check any doubtful phrases they have composed for their frequency counts in Google (advanced search with the phrase in quotations). A low-frequency count suggests that the composed phrase may be non-standard English”. They add “It appears that while corpus tools can bring clear advantages to the composing of some L2 students, both training and motivation to take the extra steps to use them are necessary” (p. 177). Hence, they highlight the role of training in relation to the Web corpus consultation.

In a somehow related study, Wu et al. (2010) reported that IELTS learners made use of concordances, offered by a collocation learning system empowered by Google, with high success rate of error correction, they corrected 73 out of 108 without assistance, and consequently, their attempts resulted in natural and native like collocations. However, although they argued that due to the messy status of the Web corpus, direct use of Google as concordance would be challenging for ESL/EFL learners, as most researchers have strongly suggested, training in GIPH and GIPD can alleviate this problem.

All in all, the findings of the studies on GIPH and GIPD are promising and pave the ground for further research. They all emphasized the need for training prior to Google consultation. However, there are some gaps and deficiencies which are needed to be filled and addressed. For example, in some studies, students were ignorant of natural patterns or the issue was not effectively brought to their attentions (Conroy, 2010; Sha, 2010). Shei (2008) conducted only a descriptive study, without getting into learners’ perception of GIPH and GIPD. Geluso’s (2011) study was small scale one (in terms of the numbers of raters, and patterns). Geluso did not either get into learners’ perceptions of GIPD. While overall the studies reported positive outcome of the use of corpora, in what ways they are conducive to constructivist and inductive learning and the factors, including challenging ones, affecting the use of GIPH and GIPD were not explored.

3. Discussion

In this section, the seven reviewed studies are summarized in terms of focus of investigation, main findings; functions (GIPH & GIPD) used and extracted themes (see Table 1).

Table 1. Summary of seven studies

Study	Focus of Investigation	Main Findings	Functions & Themes
Conroy, 2010	The use of internet-based tools in Australian universities to promote language learning. Students' perceptions & use concordancers/Google for language learning and writing improvement. Students' use concordances / Google output to successfully correct their errors. Students' independent use of Google and concordances following training.	There is little evidence of engagement with internet-based corpus tools and techniques for language promotion by universities.	GIPD
		Students found concordancing and Google consultation useful for language learning and writing improvement.	Training
		They independently used concordances and Google for improving their writing assignment.	Learning Style
		The survey findings show a considerable increase in the use of concordancers following training, as well as intention to maintain this motivation for the future use.	
Geluso, 2011	On the naturalness of GIPD, based on the frequency of occurrences on the Web	The result of his study strongly suggests that by using the Web as a corpus and Google as a concordancer, students can improve the naturalness of their writing.	GIPD Naturalness Training
		Google retrieval is faster than BNC and yields far more results either in number or in comprehensiveness.	GIPD & GIPH
Sha, 2010	To examine technical obstacles in using corpora and b) to see why the DDL approach has not made major inroads to mainstream language teaching.	Google is more dominant in usability, search speed, the number of solutions and above all, preference investigations.	Speed
		Google has spellchecker which BNC lacks.	Number of Retrieval
		Google is like online library with update language resources.	
		There is strong evidence that static corpora such as BNC are losing ground to the Web corpora.	Training
Shei, 2008	To review some key concepts in the search of patterns and formulaic sequences and identify the specific type of phraseology. To explain the Google-based method for verifying patterns and formulaic sequences. To explain how his proposed methodology can contribute to computer assisted language learning	Google accounts for patterns which the conventional corpora found them as their Achill's heel.	GIPD & GIPH
		Stubbs' (2002) Extended Lexical Unit and Hoey's (2005) Lexical Priming theory were corroborated in terms of the prevalence of language in pattern forms.	Number of Retrieval
		He put forth a method, based on reducing the fragment on word at a time and keeping the record of the frequency of the remaining fraction, useful for both for DDL (data driven learning) and linguistic purposes.	Training
Stapleton & Radia, 2009	Describing the application of the Web as medium for PD & error correction	During the writing process, students can check any doubtful phrases they have composed for their frequency counts in	GIPD

		Google (advanced search with the phrase in quotations).	Naturalness
		A low-frequency count suggests that the composed phrase may be non-standard English	Training
		Observation and questionnaires show:	
		Students using the system adopted one of two strategies. Most finished their writing first and then used it to check text they were uncertain of (PD).	PH & PD
Wu, 2010	Students' use of corpora and the Web-based corpus, collocation learning system, to examine how they use PD &PH to generate and expand their text	Some (three) used the system to help generate text by finding the correct usage of a word and suggesting suitable sentence structures (PH).	Proficiency
Wu et al., 2010	Use of collocation learning system	Students corrected 73 out of 108 errors without assistance, and consequently, their attempts resulted in natural and native like collocations.	PD Tidiness

Studies show that Google (Google Scholar) has a great potential in assisting ESL/EFL learners in using the functions of GIPH and GIPD, thereby appreciating and keeping abreast of wide range of naturally occurring patterns on the Web (Conroy, 2010; Geluso, 2011; Yoon, 2011). Seven studies conducted in this area were reviewed and some themes such as Learning Style, Training, Naturalness, Tidiness, Speed, Number of Retrieval, and Proficiency were extracted, which are discussed at length.

In relation to **Learning Style**, Conroy (2010) argues that the training should be offered for both inductive and deductive learning. Sha (2010) also highlights the issues of inductive and deductive learning. Geluso (2011) believes that inductive learning outperform deductive learning in relation to the Web corpus learning. On the same note, Creswell (2007) argues that sometimes the teacher is also needed to assist the student interpreting the search solutions, which is referred to as “deductive DDL” with teacher’s mediating and facilitating and paper-based materials, which is different from fully-fledged autonomous, self-directed hands-on “inductive DDL”. Thus, in the process of Web corpus consultation, a teacher as a facilitator and research assistant (O Sullivan, 2007) ought to address the learning issues of both inductive and deductive learners.

Another extracted theme is Training. In terms of **Training**, all researchers have emphasized the critical role of training, prior to the Web corpus consultation. Chang (2010, p. 160) argues that “At certain point, learners need teachers who provide feedback, clarify questions and concepts, and support learners’ progress”. He further adds without training and feedback they become unconfident about the knowledge they have constructed and the progress they have made.

Regarding the **Naturalness** of the Web corpus consultation, some researchers state that frequency of occurrences on the Web accounts for the naturalness of language patterns (Conroy, 2010; Geluso, 2011; Sha, 2010; Shei, 2008; Stapleton & Radia, 2009; Wu et al., 2010). Geluso (2011, p.1) states that “Corpora and data-driven learning can offer useful insights into frequent patterns of naturally occurring language to second/foreign language learners who, unlike native speakers, are not privy to a lifetime of input and fine-tuning”. Thus, ESL/EFL learners can consult Google Scholar and keep abreast of wide range of natural patterns, based on frequency of occurrences on the Web (rooted in usage based theory), and employ them in their language learning and academic writing through the functions of GIPH and GIPD.

An issue which is challenging is lack of Tidiness. In relation to the ‘**Tidiness**’ of the Web corpus, Wu et al. (2009 & 2010) argue that the pattern, *cause actual result*, which occurs much less frequently in other corpora, is the top hit in the Web corpus. They conclude that this reflects the sometimes anomalous or ‘dirty’ nature of the Web collection. However, Google Scholar, a corpus produced by native speakers of English or expert writers does not have such deficiencies (Sha, 2010). Moreover, as almost all researchers have stressed the critical role of training, it is believed that training in GIPH and GIPD can alleviate this problem (Geluso, 2011; Stapleton & Radia, 2009).

Another issue is speed. In terms of **Speed**, Google surpasses the conventional concordancers such as British National Corpus (BNC), Collins Birmingham University International Language Database (COBUILD), Corpus of Contemporary American English (COCA), etc (Geluso, 2011; Sha, 2010). Hence, an ESL/ EFL learner can access and discover his/ her appropriate natural patterns within seconds, noticeably faster than conventional corpora.

Regarding the number of **Retrieval**, no concordance or search engine can match Google (Geluso, 2011; Sha, 2010; Shei, 2008). Shei (2008, p.70) argues, "Google can offer solutions to many of the research questions in phraseology which even a billion-word corpus can hardly handle". Thus, Google can keep abreast of a wide range of genuine language patterns and phrases invaluable to teaching ESL writing course. Therefore, an ESL/EFL learner can appreciate an immense amount of natural patterns applicable to GIPH and GIPD.

Another issue highlighted by Wu (2010) is learners' **proficiency**. He points out "Because of the constraints of the topic—themselves and their family—and their limited language ability, their writing exhibited a narrow range of vocabulary and few idiomatic expressions. For example, the four most common words used were like, come, want and live. Sentence structure was simple and basic" (Wu, 2010, p.119). However, his study was not conducted on direct use of the Web corpus. Since Google provides the translation of learner's language to and from English as well as provides the definition (with synonym, antonym and usage) of almost every word in English, it can be a great boon in helping the learners even those with low proficiency levels.

On the whole, however, although some themes related to GIPH and GIPD have been located in the literature, an important point needs to be made that they are mostly the outcome of descriptive studies and based on researchers' inferences and interpretations than longitudinal first-hand empirical studies on Google, specially Google Scholar, getting into learners' perceptions through their real use of GIPH and GIPD. Therefore, a longitudinal exploratory study on the use of GIPH and GIPD is crucial.

4. Implications and Contributions

Future studies on GIPH and GIPD can contribute to the body of literature in this vein. In addition, regarding its contribution to ESL/EFL language learning context, the following groups of people can appreciate the benefits of GIPH and GIPD: teachers/educators, researchers, curriculum planners, text book writers and ESL/EFL learners.

Since the findings and the results from native-speakers' perceptions about ESL/EFL learners' GIPD in teaching ESL/EFL writing course are very relevant to the work of teachers and lecturers in English-medium universities, who have students from non-English-speaking backgrounds, and who have to teach and assess these students, this would be great evidence to support the use of GIPH and GIPD at English-medium universities and ESL/EFL teaching contexts. More specifically, educators/teachers may take advantage of the functions of GIPH and GIPD for various purposes. First, they will be introduced to ever growing authentic materials useful for their classroom syllabi or extracurricular activities through employing GIPH and GIPD (Leech, 1997; Romer, 2011), and target the specific language pattern and relate it to the context under study (Wu, 2010). Second, based on needs analysis, teachers will also be able to design a well-suited literacy curriculum by consulting the Web corpus (Chambers et al., 2011; Pérez-Paredes, Sánchez-Tornel, Alcaraz Calero, & Jiménez, 2011), compatible with ESL learners' needs in real communication events. Third, it can have implications for the teaching of patterns and multiword units via GIPH and GIPD.

Fourth, according to Liu and Jiang (2009, p. 70) "Another positive effect of corpus searches for the instructors is that by working with students in their searches and reading their search assignments, they had the opportunity to observe and understand students' discovery learning and thinking processes", hence GIPH and GIPD can help teachers assess students hands-on activities and experience of Google consultations in the process of DDL and DL. Fifth, for busy teachers, GIPH and GIPD can lessen their workloads (Stapleton & Radia, 2009) by supporting ESL/EFL learners in making corrections autonomously and independently, without the need for protracted explanations in the margins. As Gilmore (2009, p. 369) expounds on this issue "Underlining problem areas in students' work is quick to do and frees up time to concentrate on more global issues of cohesion and coherence, which the corpora cannot easily highlight", accordingly, the burden of educators reduces to indirect feedback in ESL writing course. Also, teacher can spend more time with learners who are in need of extra help, while giving advanced learners more freedom to study on their own—which is particularly useful in classrooms with large student numbers (Wu, 2010). Finally they may, also, be able to enhance their own knowledge of patterns (Liu & Jiang, 2009) through GIPH and GIPD.

Researchers may also benefit from the GIPH and GIPD study results in different ways. They may be introduced to new areas of language pedagogy and pedagogical applications (Geluso, 2011; Sha, 2010) with great potential for exploratory and empirical research studies.

The third groups who can benefit from GIPH and GIPD are curriculum planners. They may appreciate authentic materials offered by GALL through GIPH and GIPD for curriculum designing and material writing (Romer, 2011). Thus, they are able to determine the contents of the English language syllabus for schools/universities, consulting Google-driven corpus (GIPH & GIPD).

The fourth groups who may benefit from the outcomes of the studies on GIPH and GIPD can be material and textbook writers. The results can also have implications for and impacts on the content of language teaching and learning. GIPH and GIPD may provide information for textbook writers on decisions of the types of materials and activities (Norwati Roslim & Mukudan, 2011; Romer, 2011) that they can look into while preparing for the next cycle of English language textbooks.

The last but not the least, benefiting from GIPH and GIPD are ESL/EFL learners, they can benefit from GALL tools and techniques through GIPH and GIPD in different ways in order to discover appropriate patterns and improve the naturalness of their ESL/EFL writing (Conroy, 2010; Geluso, 2011; Sha, 2010; Shei, 2008; Stapleton & Radia, 2010; Yoon, 2011). First, ESL students can appreciate user friendly, free, ever-growing language learning tools and techniques such as GALL facilitating GIPH and GIPD. Second, they can do inductive and discovery learning and learning by doing (Bernardini, 2000; O'Suvillan, 2007) by means of GALL (Conroy, 2010; Sha, 2010), accordingly, they can perform the functions of PH and PD consulting Google driven corpus and thereby improve their ESL writing, by deepening their understanding of GIPH and GIPD. Finally, they can get familiar with the process of constructing their own concordancers employing Google search engine as a result self-study is encouraged (Smith, 2011).

Moreover, the factors, challenges and problems facing ESL/EFL learners in using GIPH and GIPD in language learning in general, and ESL writing in particular, will be explored to be addressed. Additionally, the results also might be applicable to the other field of studies and disciplines, as ESL/EFL writing problems are prevalent in all disciplines (Nor Aslah Adzmi, Samsiah Bidin & Syazliyati Ibrahim, 2009; Karimah Yunus & Suad Adwab, 2011). Further, in the same venue, it may accelerate the integration of GIPH and GIPD into ESL/EFL learning curricula and activities.

In sum, in addition to the contribution to the body of knowledge, studies on GIPH and GIPD can have advantages for the following groups of people: Educators/teachers, researchers, curriculum planners, material writers, and ESL/EFL learners.

5. Recommendation for Future Research

All in all, as literature demonstrates there is a huge gap in relation to researches on corpora-informed PH and PD, in general, and GIPH and GIPD, in particular. For example, very few empirical studies on corpora-informed PH and PD have been conducted (Yoon, 2011). Furthermore, very few studies on GIPD (such as Conroy, 2010; Geluso, 2011; Sha, 2010) have been reported. Moreover, very few studies, if any, have been conducted on GIPH. Only two small scale studies on naturalness of corpora-informed PD, of which one (see Geluso, 2011) was on GIPD, were conducted. In addition, Geluso's (2011) study was merely product-oriented. On the whole, while there is a rising trend in pedagogically-oriented study in the Web corpus consultation where the ESL/EFL learners' voice has gradually come to be heard, up to now there is yet little evidence about how students approach a corpus autonomously (Chang, 2010; Liu & Jiang, 2009).

Moreover, the teaching and learning potential of corpora-informed PH and corpora-informed PD, in general, and GIPH and GIPD, in particular, remains unrealized, because almost all teachers and students are unfamiliar with GIPH and GIPD and therefore do not use it in their language teaching and learning (Breyer, 2009; Chambers et al., 2011; Conroy, 2010; Yoon, 2011). In terms of Google consultation, researchers recommend exploratory researches. For example, Yoon (2011) states that "longitudinal case studies may explore how, to what extent, and for what purposes learners use Internet search engines as a reference/research tool when they compose" (p. 138). On the same note, Conroy (2010, p. 879) recommends that "The majority of studies so far have only looked at concordancing very few have looked at GALL techniques it would be very useful to investigate how and why they use GALL".

Thus, it would be useful to conduct longitudinal studies on GIPH and GIPD to explore how and why ESL/EFL learners use and take up these functions in the process of language learning. Also, studies on ESL/EFL learners' perceptions of and attitudes to GIPH and GIPD can accelerate the use of these functions and marry the gaps between the past study findings and the classroom practice (Romer, 2011). In addition, a study with mixed approach (by looking into the breadth and depth of the issue) may be valuable to test the located themes. However, due to the novelty of this research area, an exploratory study is recommended. Furthermore, Breyer (2009, p. 155) points out "If they [teachers] find their learning experience with corpora to be beneficial, then this

is likely to influence positively their decision to use corpora for their own teaching later on". Therefore, this sort of detailed exploring and understanding of the use of GIPH and GIPD in language learning and ESL writing course is needed to encourage the use of these constructs in the process of language learning and teaching ESL writing course.

6. Conclusion

As it was discussed, the use of the Web as a corpus and Google as a concordance, by providing the learners with huge amount of authentic natural language patterns, has been regarded as one of the promising areas with great potential for revolutionizing language pedagogy and L2 writing. Particularly, the functions of GIPH and GIPD can promote natural L2 writing through DL and DDL. For example a learner who is vacillating between two patterns in terms of the accuracy and naturalness, s/he can check the patterns in hand based on frequency of occurrences on the Web via Google Scholar or even search for and discover appropriate patterns. However, as the literature demonstrates, these advantages have mostly been talked about than tested with first hand empirical studies, and only more recently some researchers have initiated conducting studies in this area (Conroy, 2010; Yoon, 2011). Focusing on L2 writing, this review study explored how and to what extent this potential of GIPH and GIPD has been recognized by reviewing the related studies. Also, some themes related to the functions of GIPD and GIPH were extracted and elaborated on. Seven studies included in the present review show that learners' use of GIPH and GIPD champions the promotion of their language learning and L2 writing, providing that proper training and scaffolding are provided. However, to fill and bridge the gaps future studies are crucial. Although some studies have been reported in the literature, due to the novelty of this area more studies are recommended to explore the 'how' and 'why' in relation to the use of GIPH and GIPD, whereby discovering and addressing the challenges facing ESL/EFL learners in this vein.

References

- Adzmi, N. A., Bidin, S., & Ibrahim, S. (2009). *English Language Teaching*, 2(4), 171-178.
- Breyer, Y. (2009). Learning and teaching with corpora: reflections by student teachers. *Computer Assisted Language Learning*, 22(2), 153-172. <http://dx.doi.org/10.1080/09588220902778328>
- Bybee, J. (2006). From usage to grammar: The mind's response to repetition. *Language*, 82, 711-733. <http://dx.doi.org/10.1353/lan.2006.0186>
- Chambers, A., Farr, F., & O'Riordan, S. (2011). Language teachers with corpora in mind: from starting steps to walking tall. *The Language Learning Journal*, 39(1), 85-104. <http://dx.doi.org/10.1080/09571736.2010.520728>
- Chang, P. (2010). Taking an effective authorial stance in teaching ESL writing course: Inductive learning for second language writers using a stance corpus. PhD dissertation. University of Michigan.
- Chinnery, G. M. (2008). You've got some GALL: Google-assisted language learning. *Language Learning and Technology*, 12(1), 3-11.
- Comelles, E., Laso, J. N., Forcadell, M., Castaño, E., Feijóo, S., & Verdager, I. (2012). Using online databases in the linguistics classroom: dealing with clause patterns. *Computer Assisted Language Learning*, 1-13. <http://dx.doi.org/10.1080/09588221.2012.658407>
- Conroy, M. A. (2010). Internet tools for language learning: University students taking control of their writing. *Australasian Journal of Educational Technology*, 26(6), 861-882.
- Cresswell, A. (2007). Getting to 'know' connectors? Evaluating data-driven learning in a Writing skills course. In E. Hidalgo, L. Quereda, & J. Santana (Eds.), *Corpora in the foreign language classroom* (pp. 267-287). Amsterdam: Rodopi.
- Erman, B., & Warren, B. (2000). The idiom principle and the open-choice principle. *Text*, 20, 29-62. <http://dx.doi.org/10.1515/text.1.2000.20.1.29>
- Fletcher, W. H. (in press). Corpus analysis of the World Wide the Web. In A. Carol (Ed.), *Encyclopedia of Applied Linguistics*. Wiley- Blackwell. Chappelle.
- Geluso, J. (2011). Phraseology and frequency of occurrence on the Web: native speakers' perceptions of Google-informed second language writing. *Computer Assisted Language Learning. I First article*, 1-14.
- Gilmore, A. (2009). Using online corpora to develop students' writing skills. *ELT Journal*, 63(4), 363-372. <http://dx.doi.org/10.1093/elt/ccn056>
- Hamat, A. B., & Embi, M. A. (2009). The Web Technologies for language learning. Enhancing the course management system. *Advanced Technologies*, 357-368.

- Hoey, M. (2005). *Lexical priming: A New Theory of Words and Language*. London/New York: Routledge.
- Hong, A. L., Abdul Rahim, H., Kim Hua, T., & Salehuddin, K. (2011). Collocations in Malaysian English learners' writing: A corpus-based error analysis. *3L: The Southeast Asian Journal of English Language Studies*, 17(Special Issue), 31-44.
- Islam, A., & Inkpen, D. (2009). Real-word spelling correction using Google, the Web 1T 3-grams. Wood Stocks' 97 El Paso, Texas USA.
- KarimahYunus, B., & Adwab, S. (2011). Collocational competence among Malaysian undergraduate law students. *Malaysian Journal of ELT Research*, 7(1), 151-202.
- Kennedy, C., & Miceli, T. (2010). Corpus-assisted creative writing: Introducing intermediate Italian learners to a corpus as a reference resource. *Language Learning & Technology*, 14(1), 28-44.
- Leech, G. (1997). Teaching and language corpora: A convergence. In Wichmann, A., Fligelstone, S., McEnery, T., & Knowles, G. (Eds.), *Teaching and language corpora* (pp. 1-23). New York: Addison Wesley Longman.
- Liu, D., & Jiang, P. (2009). Using a corpus-based lexicogrammatical approach to grammar instruction in EFL and ESL contexts. *The Modern Language Journal*, 93(9), 61-78. <http://dx.doi.org/10.1111/j.1540-4781.2009.00828.x>
- O'Sullivan, Í. (2007). Enhancing a process-oriented approach to literacy and language learning: The role of corpus consultation literacy. *ReCALL*, 19(3), 269-286. <http://dx.doi.org/10.1017/S095834400700033X>
- Park, K., & Kinginger, C. (2010). Writing/thinking in real time: Digital video and corpus query analysis. *Language Learning & Technology*, 14(3), 31-50.
- Pérez-Paredes, P., Sánchez-Tornel, M., Alcaraz Calero, M. J., & Jiménez, A. P. (2011). Tracking learners' actual uses of corpora: guided vs. non-guided corpus consultation. *Computer Assisted Language Learning*, 21(1), 55-75.
- Romer, U. (2011). Corpus research applications in second language teaching. *Annual Review of Applied Linguistics*, 31, 205-225. <http://dx.doi.org/10.1017/S0267190511000055>
- Roslim, N., & Mukundan, J. (2011). An overview of corpus linguistics studies on prepositions. *English Language Teaching*, 4(2), 125-131. <http://dx.doi.org/10.5539/elt.v4n2p125>
- Sha, G. (2010). Using Google as a super corpus to drive written language learning: A comparison with the British National Corpus. *Computer Assisted Language Learning*, 23, 377-393. <http://dx.doi.org/10.1080/09588221.2010.514576>
- Shei, C. C. (2008). Discovering the hidden treasure on the internet: Using Google to uncover the veil of phraseology. *Computer Assisted Language Learning*, 21, 67-85. <http://dx.doi.org/10.1080/09588220701865516>
- Smith, S. (2011). Learner construction of corpora for general English in Taiwan. *Computer Assisted Language Learning*, 24(4), 291-316. <http://dx.doi.org/10.1080/09588221.2011.557024>
- Stapleton, P., & Radia, P. (2009). Tech-era L2 writing: towards a new kind of process. *ELT Journal*, 64(2), 175-183. <http://dx.doi.org/10.1093/elt/ccp038>
- Stubbs, M. (2002). *Words and phrases*. Oxford: Blackwells.
- Wierzbicka, A. (2009). Exploring English phraseology with two tools: NSM semantic methodology and Google. *Journal of English Linguistics*, 37(2), 101-129. <http://dx.doi.org/10.1177/0075424209334338>
- Wood, M. (1981). *A Definition of Idiom*. Manchester, England: Center for Computational Linguistics, University of Manchester.
- Wu, S. (2010). Supporting collocation learning. PhD Thesis .Department of Computer Sciencel, the University of Waikato. Hamilton, New Zealand.
- Wu, S., Franken, M., & Witten, I.H. (2009). Refining the use of the Web (and the Web search) as a language teaching and learning resource. *Computer Assisted Language Learning*, 22(3), 83-102. <http://dx.doi.org/10.1080/09588220902920250>
- Wu, S., Witten, I. A. N. H., & Franken, M. (2010). Utilizing lexical data from a The Web- derived corpus to expand productive collocation knowledge. *Computer Assisted Language Learning*, 22(1), 249-268.
- Yoon, C. (2011). Concordancing in L2 writing class: An overview of research and issues. *Journal of English for Academic Purposes*, 10(3), 130-139. <http://dx.doi.org/10.1016/j.jeap.2011.03.003>