# Independent learning through the use of data driven learning

***Amin Dehghan***
***Pornapit Darasawang***
*King Mongkut's University of Technology Thonburi*

**Abstract**

Technology-based approaches to language learning are assumed to provide learners with opportunities to self-direct their own learning (Benson, 2001). Data driven learning typically involves exposing learners to large amounts of authentic data; the electronic corpus, therefore, can play an active role for students' exploration of the language and detecting patterns in it. It poses basic concepts of learner-centeredness, discovery learning, authentic language, autonomisation and strikes many as revolutionary (Boulton, 2009). In this approach, students are at the focus of the process, taking responsibility for their own learning rather than being taught rules in a more passive mode. This study attempts to explore how using available web-based corpora and specifically, self-compiled corpora of academic papers collected by graduate students can be used as a tool to promote independent learning in academic writing. This study highlights the ways in which independent learning is manifested, facilitated, and encouraged through data driven learning. Moreover, the students' self-awareness in field specific genres and independent problem-solving of the students in their academic writing will be discussed.

## 1. Background

The concept of independence and autonomy has come to play an increasingly important role in language learning. There are several principles that highlight independent learning: optimizing or extending learner choice, concentrating on the needs of individual learners, transmission of decision-making to learners, and the accounting for learners' responsibility (Hurd & Lewis, 2008). Independent language learning represents a move towards more learner-centered approaches which consider learners as individuals who can develop and exercise responsibility for their learning based on some specific needs and rights. In other words, learners have both a capacity for active learning and an attitude towards learning which lead them to take responsibility for their learning. (Little, 1996). Independent language learning could encourage the raising of learners' awareness and knowledge of themselves, recognition of their needs and preferences, and the understanding of their beliefs and motivation and the strategies to develop second language learning. The notion of independent language learning has had a ubiquitous influence on language learning for more than three decades (Rubin, 1975; Holec, 1981; Nunan, 1988; Tudor, 1996). Moreover, studies such as self-access learning (Sheerin, 1997), distance learning (Hurd, 2005; Murphy, 2005; White, 2006), resource-based learning (Guillou, 1996), and self-directed learning (Carver, 1984) support the importance and the belief that language learners can be independent.

Studies have attempted to conceptualize independent language learning as either a particular context for learning, a philosophy or approach to learning, and learner attribute and skills (Hurd & Lewis, 2008). From the context or setting's perspective, the emphasis is on independence from the mediating presence of a teacher during the course of learning. Moreover, learners possess a certain degree of freedom to make choices (Anderson & Garrison, 1998), to select learning opportunities, and to use resources pertinent to their needs. Next, considering

independent language learning as a philosophy or approach to learning aims to develop and foster independence in learners. This dimension of independent learning, according to Paul (1990), claims that the ultimate challenge would be developing the individual's capacity to oversee his or her own learning needs. Finally, independent learning in terms of learner attribute accounts for the students' understanding of their own needs and interests. The understanding of the students' own needs is promoted by the creation of relevant opportunities and experiences which encourage student choice, over what and how they will learn, and self-reliance. Figure 1. illustrates how these dimensions of independent language learning are interrelated.

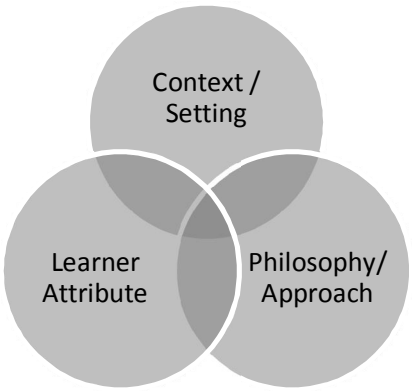


Figure 1. Interrelated dimensions of independent language learning (Hurd & Lewis, 2008)

### *1.1 Technology and independent language learning*

Technology-based approaches to language learning are assumed to be supportive of independent learning. This claim relies on the fact that they provide learners with opportunities to self-direct their own learning (Benson, 2001). Moreover, Motteram (1997) discusses that new technologies that are for learning purposes and independent learning have been related to each other for a long time such as found in computer-enhanced interactive video (Gardner & Blasco-Garcia, 1996), electronic writing environments (Milton, 1997), and concordancing (Aston, 1997). In these kinds of task it is either the technology and the interaction with it, or the potential assistance provided by technology that is seen to be supportive for independent language learning.

Using such technology can empower students when they improve confidence and self-discipline through taking more responsibility for their own learning (Warschauer, Turbee, & Roberts, 1994). Furthermore, Benson and Voller (1997) suggested that "computer software for language learning is an example of a technology which claims to promote independent learning simply by offering the possibility of self-study". Benson and Voller (1997) also discuss that educational technologies comprising the computer and the textbook can be essentially supportive of independent learning.

### *1.2 Data driven learning*

Data driven learning typically involves exposing learners to large amounts of authentic data. The electronic corpus, therefore, could be a possible tool for language learners to explore features and patterns in the target language. The language learners themselves are at the focus of the learning process, taking an increased responsibility for their own learning; instead of being

taught rules in a more passive mode (Boulton, 2009). Studies (e.g. Lee & Swales 2006, Gaskell & Cobb 2004, Charles, 2012) show that data driven learning can be most useful for extending or deepening knowledge of existing language items, differentiating close synonyms, noticing patterns of usage, collocation, colligation, morphology, to name a few. It can also raise student's awareness of issues such as frequency and typicality, register and text type, discourse and style, as well as the ambiguous nature of language itself.

Aside from aiding language learners in learning about the target language, there are other positive outcomes affecting the learning process as well. Seidlhofer (2000) states that motivation can be increased by allowing learners greater participation in creating the corpus, deciding what goes into it, or using their own productions. Closely linked to motivation is student agency. It has been suggested that data driven learning approaches, such as learning from a self-made corpora, could promote a sense of ownership over the learning progress. First, it makes students decide and choose what to put into the corpus. In other words, they have control over the content they use for learning. They can manipulate the data they have at any time to match their needs and preferences. Second, it makes them more independent because they don't have to always rely on outside agencies. Finally, its availability is an advantage. Self-made corpus is a resource that the students can use if freely almost anywhere (Charles, 2012). Furthermore, studies have shown that learners who have had experienced working with corpora would like to continue doing so in the future (e.g. Allan 2006; Lee & Swales 2006; Gaskell & Cobb 2004; Yoon & Hirvela 2004), which shows its potential benefits for learner autonomy. Charles (2012) echoes this sentiment through a study which found that about 90% of the subjects agreed that the corpus they made helped them develop their writing skills and the learners emphasized their intention to use it in the future. Nonetheless, an issue worth considering are the suitability for types of students. This is seen in Flowerdew's (2001) study where it was reported that science and engineering students used this approach easily compared to business students from the same institution.

### *1.3 DIY corpus-building*

Do It Yourself (DIY) corpus-building, which is a type of data driven learning, is an approach in EAP/ESP for the teaching of writing. This approach sees students creating and examining their own discipline-specific and individual corpora. Building individualized corpora is also multi-disciplinary, in that students from any discipline can take advantage of corpora that corresponds to the requirements and conventions of the students' discipline. This means that students are able to personalize materials to fit their academic needs. Another advantage of this approach is that it provides resources for disciplines where there may be few or no disciplinary materials or assistance available (Charles, 2012).

According to Charles (2012), DIY corpus-building can encourage independent learning in several ways. First, learners create their own resource, which is relevant to their needs. Second, consultation of students' own discipline-specific corpus may help the students to reduce dependence on teachers and other editing services. Third, it aids the students to take more responsibility for their own writing, being able to expand the resources, and having their own choices. Last but not the least, it provides opportunities for outside of class use and practice.

The present paper will attempt to shed further light on the ways Do-It-Yourself (DIY) corpus building corpora of academic papers can be used as a tool to encourage independent learning in graduate students. This study examines the stages the subjects went through the use of data driven approach and how they understand their own needs while creating opportunities

and experiences which encourage them to promote independence and self-reliance in academic writing.

## 2. Methodology

### *2.1 Participants*

The participants of this study were a group of 5 Thai graduate students (1 female and 4 males) with an average age of 29. A PhD student in Computer and Electronic engineering field of study and 4 master students from the Institute of Field Robotics (FIBO) at King Mongkut's University of Technology Thonburi, Thailand. They voluntarily took part in this study which lasted for 5 weeks. The course was designed to teach academic writing through data driven learning. The participants were perceived to be motivated to improve their academic writing skills since they are expected to write their dissertation in the English language and one of them was working on writing a journal paper in the English language. It's worth noting that the students had no prior experience in the use of data driven learning approach.

### *2.2 Procedure*

This paper is a part of a pilot study for an ongoing research. In order to see how data driven learning would enhance learners' independence particularly in academic writing , the researcher attempted to train the students to use data driven learning as a useful resource as a means to address some of their problems in academic writing. It is hoped that consequently, learners will develop a sense of independence to deal with future problems in academic writing. The training progressed in the following steps. First of all, the materials and course syllabus were designed. It is worth noting that in designing the materials, the needs of the engineering students were taken into consideration. Moreover, the syllabus designed for this study was regarded as a dynamic syllabus, which provided the use of creative and research tasks, and options to adapt it to the needs, abilities, and interests of the participants used in this study. Dynamic syllabus refers to an open and flexible syllabus that evolves as the sessions move along (Lee & Swales, 2006). As a result, the materials were modified and improved after each session of the course, according to needs and the feedback from the participants. The role of the teacher in this syllabus was considered as a facilitator that played an important role in fostering independent learning. Second, the materials were selected, and adapted to fit the students' needs in academic writing. Besides, a resource booklet was designed for the students who need to practice and read more about the software used in the sessions.

#### *2.2.1 Training*

In the first session of the class the students were introduced to data driven learning approach and were asked to collect some accredited journal articles in their own fields. The researcher asked the students to use journals they are familiar with or papers they had already read. They were also asked to add as many papers as they can to their corpus. The students were trained to convert PDF files to text since making their own corpus would mean that they needed .txt files to be compatible with AntConc software. Furthermore, the students were also introduced to general corpora such as Corpus of Contemporary American English (COCA) and British National Corpus (BNC) and also the AntConc software to analyze their own texts. The sessions were about 2 hours long each, except for the first session which lasted for 3 hours when the introduction to data driven learning was given. During each session, related handouts were given to the participants to explore both general corpora and their own corpus. The participants were given numerous tasks to deal with and they had the choice of looking for the possible

answers in both the general and their self-made corpora. Table 1. shows an overview of the course.

Table 1 Overview of the course

| Week | Session title |
|---|---|
| 1 | An Introduction to Data Driven Learning |
| 2 | Prepositions in Academic Writing |
| 3 | Transitions in Academic Writing |
| 4 | Collocations in Academic Writing |
| 5 | Dealing with Verbs in Academic Writing (reporting verbs, reported speech) |

In each session of the course the students were given the choice of working either individually or in pairs. However, the students preferred to work in pairs rather than individually. The class environment was welcoming and participants were encouraged to ask questions. In addition, the teacher constantly monitored all the participants during each session and tried to support them whenever needed.

## 3. Results and discussions

The results based on the observations notes and the interviews with three participants in the study show how the students after the given training and support through the use of data driven approach became more independent to use a new learning resource and improve their academic writing based on their own needs. The support given and the independent behaviors of students are discussed as follows.

### *3.1 Support*

The data from the class observation notes and interviews showed how students have reached a certain level of confidence. The students also demonstrated independence from the presence of the teacher support in using data driven learning approach at the end of the course. Some early research notes such as "***the students seem puzzled with the lots of results in COCA and asked a lot questions on different features of it***" show "***Students needed more support to convert text files***" (1st & 2nd week respectively) are instances of students' need for support in the early stages of their exposure to data driven learning. The students tend to ask questions regarding different features of the corpus such as display modes (List, Chart, KWIC), search string, corpus sections and different genres of the corpus (Spoken, fiction, academic, etc.). The researcher tried to provide explanations for each feature and had the students understand by using some tasks to see the results for each feature. Moreover, the researcher interacted with each student individually to monitor and respond to their questions. In some cases, such as providing more support to help them with converting the files, the researcher introduced them to some online tools such as www.convertfiles.com to make it easier for them. On top of that, the researcher demonstrated converting a file both manually and online using a data projector and then asked the students to do the same procedures by themselves. Similar procedures were taken for supporting the students with downloading the AntConc software and helping the students with different features and windows of the software.

Later the students found familiarity with the approach and tried to be more confident in the use of the corpora. "***This session the students are very independent in using the corpus. They easily do searches by their own without asking for help***" were some observational remarks noted at the 4th week of the course, which shows how the students developed using the approach

independently. It was also observed that the students experimented with different corpora to see what corpora will give them the most appropriate answer. The students' use of the corpus for each task was an example developing this ability. In this regard, the students' were asked to write what corpus or corpora they have used for each task.

Moreover, the interviews at the end of the sessions showed that the students felt quite confident in using the corpora and particularly their own corpus. In response to the question whether they needed more training for this approach a student said "***There was enough sessions***" and one said "***I know how to use and formulate by now***" which shows they probably gained enough support to use this new tool and approach for them. On the other hand, in response to the question whether the resource booklet on how to use the AntConc software was used, none of them had to use it. Two of the students said,

S1:"***No…the sessions were enough for me***"

S2: "***No, but I may use it if I want to know more"***

Whether it is due to the clarity of the training sessions, or perhaps the students themselves who do not see the point to learn more about the software or just the simplicity of the software is not clear and this needs to be investigated more in details.

From the observational notes and the interviews, independence, within the context of this study, may refer to learners who do not need much teacher's intervention for language learning (Benson & Voller,1997; Wright, 2005 cited in Hurd & Lewis, 2008). Moreover, it appears that data driven learning can provide a fundamental context or setting for independent language learning, giving the students ample learning opportunities to use the resources according to their needs (Anderson & Garrison, 1998).

### *3.2 Independent Behavior*

*3.2.1 Making decisions*

The results based on the observations and the interview questions showed how students could evaluate the resources they were introduced to, and make decisions on their own to complete the tasks and use the corpora. There were early notes from the observations such as "***The students found out that they have to shift to COCA because their corpus is small***". This notation made in the second week of the training shows how students were getting familiar with how their corpus works can how it can work better if it is bigger in size. The excerpt from observation notes in the 5th week of the training when their corpora got bigger in size shows how students realize the importance and distinguish the differences to make decisions which corpora suit their needs to do the tasks. "***Students encountered some specialized words in their own corpus that they couldn't find in the COCA."***

In another instance, observation notes shows that students used to Google as a means to establish whether the keywords or phrases they were analyzing existed. Later, the researcher mentioned this issue in the interview and asked all the students if they still used Google to find out similar results. Below are experts from the students' response to this question.

Student 1: "***if I am writing a paper, I will use the corpus***"

Student 2: "***I still use Google but for writing emails, something like that…it is like a habit… I use the corpus for specific words."***

Student 3: "***I sometimes use it as a habit, but not for searching specialized words anymore***"

The results show that the students have had always needed an agent to solve some of their problems in writing and since the students were not familiar with using corpus they used to rely on "Google" as a resort to clarify their unfamiliarity. However, they could easily evaluate and see the differences in the resources as soon as getting to know the data driven learning approach

as an expert from the observation in the first week of training shows:
***"One of the students said he used to use Google for the same thing as a problem solving agent for her writing but she sees this way [corpus approach] much more effective."***

Additionally, in the interview, one of the students highlighted his use of dictionary and the corpus approach as follows:
***"I use the dictionary for meaning mainly, but not many examples ...the main purpose of the dictionary is just give the meaning but corpus approach teach me how to use that keyword"***

Furthermore, the excerpts from the students' reply to the question of "What corpora do you use the most to solve your problems in writing?" such as
*S2:****"It depends on the words, if it very specific I will use my corpus and if it is general I will use COCA corpus because I can see more examples"***
*S3:****"COCA first and for specific word, my corpus"***
implies how they are aware of the resources they are introduced to and how they can make decision when to use them in an appropriate context.

*3.2.2 Constructing their own corpora*

As soon as the students were introduced to the data driven approach, they started collecting journal papers to make their own corpus. At the end of the course, the students were asked how many papers they have started with and how many they have collected so far. They were also asked whether they are still adding papers to their corpus or not. Two of the 3 participants who interviewed at the end of the course said that they had started with about 5 papers and they ended up with 10 and 15. One of the participant had started with 10 papers and had added 20 papers till the end of the course. It's worth noting that two of the students said that they are still adding papers to their corpora. They showed their responsibility of building their own corpus as they see the value of their corpus as one said "***When I finish reading a paper I add it to my corpus***".

Table 2 Use of the corpus by the participants

| | Number of papers in their corpus at the beginning | Number of papers in their corpus at the end | Still Adding? | Future use |
|---|---|---|---|---|
| Participant 1 | 5 | 10 | No | Yes |
| Participant 2 | 5 | 15 | Yes | Yes |
| Participant 3 | 10 | 20 | Yes | Yes |

In another response to the question 'why are you still adding more papers to your corpus?', one of the students said S2:"***I am collecting for my thesis***" and the other one S3:"***I want to use my corpus to write my paper***". The results as illustrated in Table 2., shows the students see the purpose of using their corpus and realize the importance of it. It seems the results can be compared to that of Dreyer, Bangeni and Nel (2005) where the student have gained an understanding of their own needs and interests by creating opportunities and experiences that encourage students' conscience decision-making and self-reliance. Moreover, providing students opportunities to think about their needs and objectives (Dickenson, 1994) is an aspect of philosophy or approach to independent language learning.

*3.2.3 Discovering new instances*

The data driven learning approach gave the students the opportunity to discover new words and patterns in a way that they have had never experienced it before. The data driven learning approach not only provided the students with chances to discover new patterns in academic writing but also the way the concordance lines were shown helped the students. As an illustration, students said it is easier for them to remember the propositions of some words as they can they see them in practice using the concordance lines. In another observation, one of the students said ***"she used to use only two words "present and propose" as a reporting verb in her review literature section of her papers and she said now she can search something like "'et al' and 'author*'" in her corpus to come up with other possible verbs. She was very excited to discover it" (observation notes, 5th week).*** Consequently, this discovery helped her find words to expand her knowledge about vocabulary. This was achieved at the point when the students had expanded their corpora and they could discover more instances of their specific searches for their own needs.

Besides, there were times when the students could discover new searches to learn new things by their own. As an example, as an expert from the observation shows below, in the session on transition words, the students found a way how they can learn the way they should use punctuation using this approach.

***"Students said they only knew the punctuation for the most common positions of transition (when they appear that the beginning of the sentence) but they learned the other ways too."***

It's worth nothing that this was not set as a goal for the tasks in that session, but the students come up with a discovery which showed they liked learning it for their needs. Moreover, in the interview at the end of the course, in response to the question in what situations the subject will use his or her own corpus, one of the students said:

S2: ***"Sometimes I want to say "this paper propose" and I will see what follows [next], or "this paper presents". If I search on my corpus I will see the overview picture of my field***."

The results from the interview shows that the students are able to formulate their own discoveries and use subsequently learn new things helping them to become independent in developing their academic writing.

## 4. Conclusion

The present study shows providing students with opportunities through the use of data driven learning tasks and particularly DIY corpus-building can bring about potential capabilities for developing independent language learning. Data driven learning approach provides the students the freedom of making choices and when it is complied with a DIY corpus-building, it gets much closer to the students' needs and interests. Moreover, making a corpus by the students themselves provokes interest and enthusiasms among them and a sense of responsibility for their own learning.

## Acknowledgement

We would like to thank Chris Osment, and Dougal Graham at the School of Liberal Arts, King Mongkut's University of Technology Thonburi, and Maggie Charles at Oxford University Language Centre for granting permission to use some of their materials for this study.

**References**

Allan, R. 2006. Data-driven learning and vocabulary: Investigating the use of concordances with advanced learners of English. *Centre for Language and Communication Studies, Occasional Paper,* 66. Dublin: Trinity College Dublin.

Anderson, T., and Garrison, D.R. (1998). Learning in a networked world: New roles and responsibilities. In C. Gibson (Ed.), *Distance Learners in Higher Education*. (p. 97-112). Madison, WI.: Atwood Publishing.

Aston, G. (1997). Involving learners in developing learning methods: exploiting text corpora in self-access. In P. Benson and P. Voller (eds) *Autonomy and Independence in Language Learning*. London: Longman, pp. 204-214.

Benson, P. & Voller, P. (1997). *Autonomy and Independence in Language Learning* London: Longman.

Benson, P. (2001). *Teaching and Researching Autonomy in Language.* Applied Linguistics in Action Series. London: Longman.

Boulton, A. (2009). Testing the limits of data-driven learning: Language proficiency and training. ReCALL, 21(1), 37-51.

Carver, D. (1984) Plans, learner strategies and self-direction in language learning. *System* 12 (2), 123-131.

Charles, M. (2012). 'Proper vocabulary and juicy collocations': EAP students evaluate do-it-yourself corpus-building. English for Specific Purposes 31 (2), 93–102.

Dickinson, L. (1994). Preparing learners: Toolkit requirements for preparing/orienting learners. In E. Esch (Ed.), *Self-access and the adult language learner* (39-49). London: CILT.

Dreyer, C., N. Bangeni & C. Nel (2005). A framework for supporting students studying English via a mixed-mode delivery system. In Holmberg et al. (eds.), 92–118.

Flowerdew, L. 2001. The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry & R. Roseberry (Eds.), *Small Corpus Studies and ELT: Theory and practice* (pp. 363-379). Amsterdam: John Benjamins.

Gaskell, D. & Cobb, T. 2004. Can learners use concordance feedback for writing errors? *System,* 32/3, 301-319.

Guillot, M. (1996) Resource-based language learning: pedagogic strategies for Le Monde sur CD-Rom. In E. Broady and M. Kenning (eds) *Promoting Learner Autonomy in University Language Teaching* (pp. 139-158). London: CILT

Holec, H. (1981) *Autonomy in Foreign Language Learning*. Oxford: Pergamon

Hurd, Stella (2005). Autonomy and the distance language learner. In: Holmberg, Boerje; Shelley, Monica and White, Cynthia eds. Distance education and languages: evolution and change. New perspectives on language and education. Clevedon, UK: Multilingual Matters, pp. 1–19.

Hurd, S., & Lewis, T. (2008). *Language Learning Stratgies in Independent Settings.* Great Britain:Cromwell Press Ltd.

Lee, D., & Swales, J. (2006). A corpus-based EAP course for NNS doctoral students: Moving from available specialized corpora to self-compiled corpora.

Little, D. (1991). *Learner Autonomy. 1: Definitions, Issues and Problems*. Dublin: Authentik.

Milton, J. (1997). Providing computerized self-access opportunities for the development of writing skills. In P. Benson and P. Voller (eds) *Autonomy and Independence in Language Learning*. London: Longman, pp. 237-248.

Motteram, G. (1997). Learner autonomy and the Web. In V. Darleguy et al. (eds) *Educational Technology in Language Learning: Theoretical Considerations and Practical Applications*. Lyons: INSA (National Institute of Applied Science), pp. 17-24.

Murphy, L. (2005). Critical reflection and autonomy: a study of distance learners of French, German and Spanish. In B. Holmberg, M. Shelley,&C. J. White (Eds.), *Distance education and languages: evolution and change*(pp. 20-39). Clevedon: Multilingual Matters Ltd.

Nunan, D. (1988). *The learner-centred curriculum.* Cambridge: Cambridge University Press.

Nunan, D. (1988). *The learner-centred curriculum: a study in second language teaching*. Cambridge/ New York/ Melbourne: Cambridge University Press.

Paul, R. (1990). Towards a new measure of success: developing independent learners. *Open Learning: The Journal of Open, Distance and e-Learning, 5*(1), 31-38.

Rubin, J (1975). What the 'good language learner' can teach us, *TESOL Quarterly,* 9, 41-51.

Seidlhofer, B. 2000. Operationalizing intertextuality: Using learner corpora for learning. In L. Burnard & T. McEnery (Eds.), *Rethinking Language Pedagogy from a Corpus Perspective* (pp. 207-223). Frankfurt: Peter Lang

Sheerin, S. 1997. An Exploration of the Relationship between Self-access and Independent Learning. In Benson, P. and Voller, P. (eds.). 1997. Autonomy and Independence in Language Learning. London: Longman.

Tudor, I. (1996). *Learner-centeredness as language education*. Cambridge: Cambridge University Press.

Warschauer, M., Turbee, L., & Roberts, B. (1994). *Computer learning networks and student empowerment*. Honolulu, HI: University of Hawaii, Second Language Teaching & Curriculum Center.

White, C. (2006). Robinson Crusoe and the challenges of supported distance language learning. In T. Lamb & H. Reinders (Eds). *Supporting independent language learning: Issues and interventions*. Frankfurt am Mein: Peter Lang. Pp. 55-72.

Yoon, H. & Hirvela, A. 2004. ESL student attitudes toward corpus use in L2. *Journal of Second Language Writing,* 13/4, 257-283.