BEING AUTONOMOUS EFL LEARNERS BY THE USE OF A CORPUS

A Case Study of Mental Verbs across Genres in the Learning of Lexis and Grammar

Prihantoro

Universitas Diponegoro

prihantoro2001@yahoo.com

Abstract

In this paper, I will show how a corpus may take EFL (English as Foreign Language) learners a step closer to be autonomous. A group of randomly selected EFL learners was asked to navigate the Corpus of Contemporary American English (COCA) with focus on mental verbs. This type of verb is selected, as in general, mental verbs (e.g, to think, to confirm, to agree, etc) are commonly known to take human agent. This study is aimed to show how mental verbs are actually used in diverse English texts. The finding of this study has shown that mental verbs might have non-animate agent such as report, finding, document, result etc, especially in academic section. This shows that successful language learning requires proper usage as shown by linguists following the works under the scope of Systemic Functional Grammar and Neo-Firthian School. By using corpora, a teacher may shift from the sole source of information to a facilitator, and in turn, may take learners a step closer to be autonomous.

Keywords: Corpus Linguistics, autonomous learners, English as a Foreign Language, mental verbs

INTRODUCTION

This paper is written under the scope of EFL (English as a Foreign Language) learners. I believe that this should be made clear first, as I obtained my data from native speakers of Indonesian in Indonesia, where in Indonesia, the status of English would best fit ‘foreign’ rather than ‘second’ language. Although by the end, learners would eventually acquire English, but the route of how learners acquire English does matter. In Indonesia, English is studied in classroom and under formal context, rather than learned under genuinely natural English-speaking-country-like environments. Of course, there are some exceptions such as expatriate families, or specific situations (in family, business meetings, tourism spots, etc) where English is learned, not studied.

 One of the challenges of learning English in this situation is the difficulty to obtain a genuine language situation. Not all textbooks and EFL material are genuine; even assuming that they are, simplification or modification is sometimes necessary and unavoidable to catalyze successful foreign language learning. Fortunately, Internet and computer technology allow learners to access authentic materials. News script, magazines, documentary movies in English is available.

 Corpus is one alternative to help learners acquire English. It might help learners validate a certain word choice or structure preference without the presence of a teacher. In turn, this will take them a step closer to be autonomous learners. This will be very useful, as intuition of a non-native speaker will certainly differ than that of a native one.

LITERATURE REVIEW

Is it possible for a non-native speaker to construct a valid claim based on a corpus? The answer is yes. Consider the finding of Frej & Nam (2014), who are not native speakers of English, about the usage of intensifier *too* and *very* in BNC (British National Corpus). They have managed to show that when too co-occurs with evaluative adjective, it will turn the sentiment polarity of the adjective to negative as it reflects something unexpected (compare very big and too big). For *very*, it is more flexible; larger context is required to determine its polarity.

Corpus approach can be useful when properly applied under pedagogical scope. In this field, Sylviane Granger is a well-recognized name. Some of her works include practical approach on how to use corpus and computer in language learning as in Granger (2003), Granger, Hung, & Petch-Tyson (2001) and Maunier & Granger (2008). In the field of academic writing, the work of Coxhead (2000) was refined by Davies (2008) by using COCA (Corpus of Contemporary American English) data.

A group of linguists in Birmingham is called Neo-Firthian linguists (Mc Enery & Hardie, 2012), as they follow Firth’s functional sentence perspective (FSP). They develop Firth’s FSP by applying them to corpora. One that is considered the pioneer is John Sinclair who is a leading figure in the creation of Collin corpus-based dictionary. Sinclair (1991) believes that a corpus might sometimes give surprising evidence; it can be complementary or even contrary to the knowledge of a native speaker. In Reading Concordance, Sinclair (2003) shows how because and some other English words tend to have negative semantic prosody; where at that time, this information was not present in the dictionary.

As for Mental verb, it is one of the semantic categories of English verbs proposed by Biber, et al (2003:106). They believe that verbs, which fall to this category, refer to mental state and activities. Under functional grammar approach, Halliday & Matthiassen use a more generic term, which is ‘mental clause’ (Halliday & Matthiassen, 2004:201). The use of more generic term here makes sense as other POS might also be classified into this category. Take an example from Biber, et al (2003), to think. The noun form thought can also be classified as mental words although it is a noun. In traditional lexical semantic or any introductory textbook to linguistics, as in (Fromkin et al, 2011) these words are the properties of human agents. Halliday & Matthiassen (2004:201) refer this the agent as ‘senser’.

“In a clause of ‘mental’ process, there is always one participant who is human;

this is the Senser, introduced above: the one that ‘senses’ — feels, thinks,

wants or perceives, ….. Expressed in grammatical terms, the participant

that is engaged in the mental process is one that is referred to pronominally

as he or she, not as it.”

Do the findings in this paper support the view of Halliday & Matthiassen that senser possess [+HUMAN] feature? Is it possible for the so-called mental verbs to have inanimate agent that is expressed by it? This paper seeks answer to those questions. The methodology section will deal with the corpus data and method used to retrieve the data. That section also shows how learners are involved in decision-making and data interpretation.

METHODOLOGY

 A group of advanced EFL learners in Undip was randomly selected to participate in this research. From 40 randomly selected students, 20 confirmed positively. As these students had no corpus foundation at all, they were given a short practical training on how to navigate the Corpus of Contemporary American English (COCA). Students used a list of mental verbs from Biber, et al (2003). The list contains 20 most frequent mental verbs in English. From these 20 verbs, five were randomly selected. Students used these keywords in writing queries to perform automatic retrievals across text types. The objective of the search was to retrieve agents/sensers responsible for the mental verbs. The findings were then classified based on semantic compositionality and further be validated by relevant literatures that has been commented previously.

 As for COCA (Davies, 2008) itself, there are reasons why I prefer this corpus to be used in this research. First, in terms of quantity, the size of the corpus is considerably large. Second, the interface is quite friendly to its users, and it requires no computing skills to navigate COCA. Therefore, as long as the users are familiar with internet and computer, using COCA is relatively less complicated. Third, the interface is not complicated, however very functional. Besides displaying the result in the form concordance, it can also visualize it based on the text type distribution. It can also display the result in the form of historical order. For this reason, this corpus is claimed to be the first reliable monitor corpus of English (Davies, 2010).

FINDINGS AND DISCUSSION

The findings and discussion section here is divided into two sub-sections. The first sub-sections deals with student findings on COCA. The second sub-section is dedicated to understand how students reflected their hands-on experience on using the corpus.

The Non-Compositionality

The query was designed to retrieve mental verbs. We here used square brackets [LEMMA] to surround the query strings. The intent was to retrieve all possible forms from a single lemma (e.g suggest, suggests, suggested, suggesting). The lemma must be specified as a verb, otherwise, derivational forms (suggestive, suggestion and etc) might have been included too. Therefore, each query word was additionally marked by .[v\*]. The result was shown in the form of concordance and text type chart. Students, alongside with the teacher, interpret the data. This will test Halliday & Matthiassen’s (2004) view that the senser is always human.

Table 1. Mental Verb – Agent Composition Across Text Types (Top 100 Concordance Lines)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Mental Verbs | spoken | fiction | magazine | newspaper | academic |
| *believe**find**read**consider**expect* | ----- | --+-- | --+-+ | ----+ | +---+ |

Table 1 shows mental verbs and their distribution across text types. In COCA, there are six text types, one is spoken, and the others (fiction, magazine, newspaper, and academic) are written. Marking on the table is made based on the analysis of top 100 concordance lines for each lemma. Within this short span, we managed to identify the presence of non-human agents (+) that accompany mental verbs in all text types except for spoken section. Now, consider the sample keywords in context (concordance) as shown by table 2:

Table 2. Sample Concordance Line

|  |  |
| --- | --- |
| Academic | *The lambda form creates an anonymous function, which in this case expects zero arguments.**to realize that the New Orleans community believed in the " strong, self-contained hero* *Nixon administration debated cutting off aid but did not because it believed it needed to keep* *The second hypothesis believes that psoriasis is an immune-mediated disorder**and Jupiter Research expects online dating sites to record over $640 million by 2007**Its editor Leslie Fong afterwards said in an interview that the paper expected the Fay story to*  |
| Magazine | *By 1999, the market-research firm Mercury Research expects AGP graphics systems* *what to do with. Tax me more. " A sign held by twins read, " You would know our dad* |
| Newspaper | *San Jose industry-watcher VLSI Research expects the industry's sales to rise a thumping*  |
| Fiction | *a magic monkey paw had read my wish for a more mommishy mom and it had broken Liza* |

The agents responsible for the mental verbs in the examples shown by table 2 are all inanimate agents. In isolated context, they are not semantically compositional. How can a function or research expect something? How might a paw possibly read something? This violates the view that semantic feature of a verb should have been equal to its agent, as in traditional lexical semantics as also expressed also by classic Chomsky’s example *colorless green ideas sleep furiously* (cited by Fromkin et al, 2011). These mental verbs and their agents, however, are used in wider context. A paw can never read. It is an inanimate noun. But in fiction, sometimes animals or plants can talk. People can understand this well and tolerate the non-compositionality because of the genre or text types.

If we notice carefully, there is always extension of selectional restriction from non-human to human, and for this case, when the surface form of the agent is non-human. Consider the non-human agents from academic texts function, community, administration, hypothesis, research and paper. Community is a collection of people as well as administration. They can be considered mass nouns that involve human agents. A hypothesis is surely not a living thing; however, the originator of a hypothesis is a human. This might apply to research (done by researcher/s) or paper (written by person/s).

We can say that these non-human agents are used figuratively or metaphorically; and when they are transferred to accompany mental verbs, the semantic values of the agents are no longer the same. Halliday & Matthiassen believed that these inanimate nouns have been ‘endowed with consciousness’ (2004:201). In this way, the senser might be non-human, but they must have adapted to the semantic configuration of the mental verbs, which is [+HUMAN]. Reasons for this are many; depersonalization (in academic text), personification (in fiction/literary texts) or text-specific or writer’s personal rhetoric (as in newspaper or magazine).

But if that so, why we cannot apply this paradigmatically in all cases? Do you find the phrase *the hypothesis thinks* common? Even in academic texts where non-human agent phenomenon is common, using such expression would sound weird. Sinclair’s view for this matter is consistent (1991, 2003, and 2006) that lexis is inseparable from grammar. Therefore, to what extend the selectional restriction can flex may differ. Therefore, hypothesis can accept believe but not think. We cannot take the previous view for granted, say that hypothesis in *the hypothesis thinks*, has extended, and the semantic values has adapted to the verb. We therefore cannot fill the paradigmatic slot with anything and hope that any inanimate noun would fit any mental verb. Selectional restriction applies for individual lexical item here, and cannot be generalized from the part of speech merely.

This is parallel to the view of Maunier & Granger (2008) with the open-choice and idiom principle. In Phraseology in Second Language Teaching, they believe that EFL/second language learners may find that expression doubting, as having inanimate agent with mental verb is uncommon (unless it is common in their L1). This is caused by the interference of their L1, so that they find it difficult to accept as happened for Japanese (see Kojima & Kojima, 1978).

Assisting Learners to be Autonomous

When the task was finished, I distributed a blank sheet to the students and requested them to write their reflections on corpus use. As it was an open form, student’s responses varied, but I have managed to extract the recurrent responses.

First, it was something new for the students. They experienced consulting the teacher, dictionary or internet, but corpus use was new. Navigating the corpus was not difficult, but almost all of them mentioned that it took time to get them used to the procedures.

Second, the hardest part was the data interpretation. When they consult a teacher, an internet page or a dictionary, they got a complete description rather than just clues suggested by a corpus. As for the corpus use, they still had to analyze the findings.

A small number of the students expressed their interest in the technology; how to build a corpus, how to design the interface, how to use it on their own data, etc. The interests of these students would fit the field of corpus processing, which focus on technical issues (automatic retrieval method, machine-readable dictionary, collocation scoring, ets). However, for pedagogical purpose, what actually matters the most is data interpretation, and how they present this interpretation on a way that works best to support successful language learning.

As for more students, they believe that using a corpus is like a detective work, where they have to deduce a conclusion from the premises. This, according to the students, was equal to morphology or phonology exercise items (e.g identifying phoneme, As expressed by Granger, Hung, & Petch-Tyson (2001), corpus and computer are tools, and these tools function in a decision support system. Humans must perform the final decision / data interpretation.

To summarize the reflection, corpus methodology was new for these students. When using this method, learners have to be careful not to be ‘trapped’ on the technology sophistication. This is a negative evaluation for the pedagogical purpose, but for technical issues this might be a plus. These days, the number of linguists working on technical issues has increased as it was in the past (dominated by computer scientist), and the quality of the corpus. Nevertheless, all of the students share one equal reflection. By using a corpus, they have become more appreciative of the academic works. The finding of this research, for example, is simple, but it took complex procedures (and time) on how to arrive at the finding. It is actually analogous of how books or dictionaries provide descriptions.

CONCLUSION

 This paper has shown that it is possible for mental verbs to have non-human agents where direct/indirect human interference is present. We might also argue that depersonalization is acceptable (or even central) in one text type. Therefore, one cannot take this claim for granted and apply this to all text types. The semantic non-compositionality must at least consider the text type where the composition is used and the degree of collocation between the two lexical items. The collocation frequency indicated that speaker of a language may store words separately, or as chunks. The chunks, or multi word expressions might be fixed or semi fixed where certain lexicogrammatical variations are permitted.

 In terms of assisting learners, corpus is extremely helpful. Students retrieve the corpus data, investigate, discuss, analyze and take conclusion. Teacher may shift from a sole source of information to a facilitator. This characterizes autonomous learners. However, due to the highly abstraction skills required to ‘read’ a corpus data, a teacher must assess the students skills beforehand. Navigating a corpus with its complexity might be too challenging for beginners.

Readers must be aware that this paper is written on qualitative description. I definitely agree that quantitative and statistical measures are highly required to make a claim on data validity. For instance, the top 100 concordance for [believe] shows that non-human agents are distributed on academic and newspaper texts. However, it does not mean that they do not exist in other text type. When the concordance span is increased to 1000, we might find the presence of non-human agent on all text type with significant differences. Therefore, a suggestion for further research is to include deep computing and serious statistical measure to claim the validity of this research.

# References

Biber, D., Johansonn, S., Leech, G., Conrad, S., & Finegan, E. (2003). *Longman Grammar of Spoken and Written English.* Edinburgh: Pearson Education.

Coxhead, A. (2000). A New Academic Wordlist. *TESOL Quarterly 34 (2)* , 213-238.

Davies, M. (2008). *American Corpus*. Retrieved August 9, 2014, from The corpus of contemporary American English (COCA): http://www.americancorpus.org

Davies, M. (2010). The Corpus of Contemporary American English as the First Reliable Monitor Corpus of English. *Language and Literary Computing* , 447-464.

Frej, Y., & Nam, J. (2014). Study on the Sentiment Polarity Types of Collocations for too and very. *Journal of Philology (1)* , 23-32.

Fromkin, V., Rodman, R., & Hymes, N. (An Introduction to Language 9th Edition). *2011.* New York: Wadsworth Cengage Learning.

Granger, S. (2003). Error-tagged Learner Corpora and CALL: A Promising Synergy. *CALICO, 20 (3)* , 465-480.

Granger, S., Hung, J., & Petch-Tyson, S. (2001). *Computer Learner Corpora, Second Language Corpora and Foreign Language Teaching.* Philadelphia-Amsterdam: John Benjamin Publishing.

Halliday, M., & Matthiassen, M. (2004). *An Introduction to Functional Grammar (Revised Edition).* Oxford: Oxford University Press.

Kojima, S., & Kojima, K. (1978). S(Inanimate Subject)VO: A Syntactical Problem in EST Writing for Japanese. In M. Trimble, K. Trimble, & K. Drobnic, *ESP: Science and Technology* (pp. 198-226). Oregon: Oregon State University.

Maunier, F., & Granger, S. (2008). *Phraseology in Foreign Language Teaching.* Philadelphia-Amsterdam: John Benjamin Publishing.

Mc Enery, T., & Hardie, A. (2012). *Corpus Linguistics: Method, Theory and Practice.* Cambridge: Cambridge University Press.

Sinclair, J. (1991). *Corpus, Concordance, Collocation.* Oxford: Oxford University Press.

Sinclair, J. (2003). *Reading Concordance.* London: Pearson & Longman.

Sinclair, J. (2006). *Trust The Text: Language, Corpus, and Discourse.* London & New York: Blackwell.