A CORPUS BASED STUDY ON THE USE OF 'ETC' IN ACADEMIC WRITING

Studi Berbasis Korpus pada Penggunaan 'etc' dalam Penulisan Akademik

SRI WIDIASTUTI

Sultan Ageng Tirtayasa University

Jl. Raya Jakarta Km. 4 Pakupatan Serang

081932249760

widi@untirta.ac.id

Abstrak: Studi sebelumnya dari kata-kata yang digunakan dalam penulisan akademik telah menghasilkan daftar kosakata bahasa Inggris akademik. Penelitian ini bertujuan untuk menganalisis penggunaan kata *'etc' atau 'dll'*, singkatan untuk *'et cetera'* yang bermakna 'dan lain-lain', yang tidak termasuk dalam daftar itu. Penelitian ini menggunakan secara khusus metodologi yang merupakan pengukuran yang akurat dari penggunaan kata dalam menulis akademik, yaitu data dari *Corpus of Contemporary American English (COCA)* dan *British National Corpus (BNC)*. Berdasarkan temuan, dapat disimpulkan bahwa meskipun pedoman penulisan akademik di beberapa *writing centers* di universitas di Amerika Serikat menyatakan untuk menghindari penggunaan *'etc'* dan menggantikannya dengan *'and so forth'* dalam menulis akademik, data corpus menunjukkan bahwa *'etc'* lebih digunakan dalam menulis akademik dan lebih populer dibandingkan dengan penggunaan *'and so forth'*.

Kata Kunci: 'etc', penulisan akademik, *COCA, BNC*

Abstract: *Previous studies of words used in academic writing have resulted in the vocabulary list of academic English. This study aims at analyzing the use of the word 'etc', the abbreviation for 'et cetera', which does not belong to one of that list. It especially concerned with methodological issues relating to the accurate measurement of the usage of that word in academic writing, in this case the data from the Corpus of Contemporary American English (COCA) and the British National Corpus (BNC). On the basis of findings, it can be concluded that despite the academic writing guidelines in several university's writing centers in United States stating to avoid the use of 'etc' and substitute it with 'and so forth' in academic writing, the corpus data show that 'etc' used mostly in academic writing compared to the other types of writing and more popular than the use of 'and so forth'.*

Keywords: 'etc', academic writing, COCA, BNC

INTRODUCTION

In English-major programs at the tertiawry level, written work is an integral part of its education. Those written assignments and terms papers the students are required to do is called an academic writing. It differs from other kinds of writing such as personal, literary, journalistic or business writing. Additionally, in academic writing, the reader is a lecturer, professor or instructor (Oshima & Hogue, 1999). Academic writing is arguably the most important language skill to English tertiary students whose grades are largely determined by their performance in written assignments, academic reports, term examinations and graduation theses (Nga, 2009).

Nevertheless, academic writing is the kind of writing demanding the acquisition of several linguistic practices which are embedded in complex sets of discourses, identities, and social meanings. This complexity makes the learners of English have

difficulty to compose academic work in forms acceptable to English audiences (Paltridge & Starfield, 2007; Swales, 1990). Thus, several studies have revealed the language used in academic writing that can help the students to compose their academic writing. Those studies result in the list of words that can be used in academic writing (Swales, 1990; Bailey, 2003; Swales & Feak, 2004).

Despite so many studies on the language used in academic writing, little research has examined the use of abbreviation 'etc' which tends to be avoided in the guidelines for academic writing in several university's writing centers in United States (see the list in the discussion section). According to the APA, Latin abbreviations (e.g., i.e., etc., et al.) should only appear in parenthetical material; otherwise, English translations of these abbreviations should be used instead.

Thus, this study is aiming to analyze and compare the use of 'etc' and its English translation, 'and so forth'. It especially concerned with methodological issues relating to the accurate measurement of the usage of those words in academic writing, in this case the data from COCA and BNC corpus. The corpus study will reveal three research problems, i.e. the use of 'etc' in COCA and BNC, the use of 'etc' in corpus versus the guidelines in academic writing, and the use of 'etc' versus 'and so forth' in academic writing.

LITERATURE REVIEW

The Use of 'etc' in Academic Writing
The abbreviation 'etc' for 'et cetera' is used after a list to show that there are other things that you could have mentioned (Oxford Advanced Learner's Dictionary 8th Edition, 2010).

Having searched the guidelines of the use of 'etc' in academic writing books, the writer did not find a book clearly mentioning whether the use of 'etc' in academic writing is allowed or avoided. Most sources are from the writing centers in several universities in United States and United Kingdom. However, the writing centers in universities in United Kingdom do not have a detail explanation of the use 'etc' in academic writing as it is found in the writing centers in several universities in United States. It is also found that there are three writing centers in universities in United States explicitly stating to avoid the use of 'etc' in academic writing as shown in the following.

Walden University
According to the APA, Latin abbreviations (e.g., i.e., etc., et al.) should only appear in parenthetical material; otherwise, English translations of these abbreviations should be used instead. The only exception is *et al.*, which can be used both in normal and parenthetical material.

Common Latin Abbreviations

Latin Abbreviation	English Translation
i.e.	that is
e.g.	for example
etc.	**and so forth**
vs.	versus
cf.	compare
et al.	and others

University of Montana
Latin Abbreviations for Academic Writing
Despite the fact that Latin is technically a dead language (i.e. no longer spoken and changing), it has heavily influenced English; many Latin phrases and abbreviations are

still found in modern texts and used in everyday conversation.

Using Latin Abbreviations
Because they are often misunderstood and misused, Latin abbreviations should be used sparingly, if at all, in formal writing. Most sources agree that writers can make use of abbreviations in footnotes and references but should otherwise avoid them.

In most cases, the English translation of the abbreviation can be substituted with no detrimental consequences. If an abbreviation must be used, writers must be certain they choose the correct abbreviation. A good way to ensure that the abbreviation chosen is actually the one the writer wants is to substitute the English translation of that abbreviation and see if the sentence still makes sense.

Substitution Example
L: My teacher asked me to hand in my paper after the end of Spring Break (i.e. March 17).
E: My teacher asked me to hand in my paper after the end of Spring Break (that is, March 17).

Empire State College
Standard abbreviations in writing, which are often abbreviations of Latin phrases such as:
a. i.e. (*id est*) - that is
b. cf. (*confer*) - compare
c. e.g. (*exempli gratia*) - for example
d. et al. (*et alii*) - and others
e. **etc. (*et cetera*)-and other things**
f. N.B. (*nota bene*) - note well
Note: Such abbreviations are usually used in tables, source citations, and comments inside parentheses in formal writing. Avoid their use in academic or formal writing.

The Corpus of Contemporary American English (COCA) and the British National Corpus (BNC)
The COCA and BNC are corpora of English that are freely-available online. Following are the comparison of the two corpora in terms of corpus size, genre coverage, and how up-to-date they are.

Corpus size
The Corpus of Contemporary American English (410+ million words) is more than four times as large as the British National Corpus (100 million words). As a result, it often provides data for lower-frequency constructions that are not available from the BNC. In terms of concrete examples, following are two types of phenomena -- collocates and syntax.

Collocates/semantics
The following table shows the number of different collocates that occur at least 3-5 times with the given node words. Notice that with a word like *nibble,* the word itself only occurs 4-5 times as often in COCA as the BNC (1194 to 244; to be expected from a corpus four times the size). But in terms of collocates, there are 14 times as many in COCA that occur 5 times or more as there are in the BNC. For low frequency words like these, there is often a real difference between a 100 million word corpus and a 410 million word corpus.

Word (PoS)	COCA freq	BNC freq	collocate PoS / span	COCA	BNC
click (noun)	3145	445	adj 2L / 0R	**27** loud, audible, double, sharp	**5** double, sharp, loud
nibble (verb)	1194	244	noun 0L / 3R	**29** edges, grass, ear, lip	**2** ear, bait
serenely (adv)	308	83	verb 4L / 4R	**23** smile, float, gaze, glide	**2** said, smiled
crumbled (adj)	446	27	noun 0L / 3R	**32** cheese, bacon, bread, cornbread	**0** ---

Syntax

Consider the following three examples.

1. [like] **for** [p*] to [v*] (*I'd really like for you to stay*) There are **5** tokens in the BNC, but **352** tokens in COCA. With the BNC there aren't enough examples to see if this is a feature of informal or formal English, but the data from COCA show that it is clearly a feature of spoken English. The data also shows that it is increasing slowly over time, when compared as a ratio to the construction [like **--** him to V].

2. Is it *excel **in** V-ing*, or *excel **at** V-ing* ? (*she excels in/at playing the piano*) Granted, this is a very narrow issue, but it is precisely the thing that translators and non-native speakers are interested in. With the BNC there are **5** tokens with ***at*** and **6** with ***in*** -- probably not enough to say which is more common. In COCA, however, there are **136** with ***at*** and **47** with ***in***. This is enough to begin to see which genres prefer one or the other, as well as which subordinate clause verbs occur with each. Such granularity is not possible with the BNC.

3. [have] been being [vvn] (*she had been being watched*)

There are **2** tokens in the BNC (1 spoken, 1 fiction), and this is not enough data to see any possible genre variation. In COCA, on the other hand, there are **14** tokens (10 spoken, 2 fiction, 1 news, 1 academic). This is enough to show that this is a feature of spoken English, and the data also shows that it is increasing since 1990. (By the way, most native speakers of both dialects will cringe at sentences like this, but they *are* in the corpora.)

In summary, while 100 million words is often adequate for studying syntax, for some very low-frequency phenomena, there is a real difference between 100 million words (BNC) and 410 million words (COCA).

Genre balance

The BNC is 10% spoken/90% written, while in COCA the corpus is nearly evenly divided (20% in each genre) between spoken, fiction, popular magazines, newspaper, and academic.

GENRE	**COCA** (millions of words)	**BNC** (millions of words)
Spoken	85	10
Fiction	81	17
Popular magazines	86	16
Newspaper	81	11
Academic	81	16
Other		30

The BNC has a much wider range of spoken sub-genres, while COCA is composed of *unscripted* conversation on TV and radio shows (See notes on the naturalness of these conversations: COCA/Help-Information/Texts/Spoken). Both corpora are very well balanced in terms of sub-genres for the written genres (e.g. Newspaper-Sports, or Academic-Medicine). In addition, because there is a diachronic aspect to COCA (coverage over time), in COCA the distribution of 20% in each of the five genres stays constant from year to year.

Summary

COCA and the BNC complement each other nicely, and they are only large, well-balanced corpora of English that are publicly-available. The BNC has better coverage of informal, everyday conversation, while COCA is much larger and more recent, which has important implications for the quantity and quality of the data overall.

Unless one is inherently interested in only British or American English, there is really no reason to not take advantage of both corpora. This is especially true when both corpora can be used side-by-side, with the same interface. For most types of studies, academic publications and presentations that rely on just the BNC for data from Modern English will look increasingly outdated and insular as time goes on.
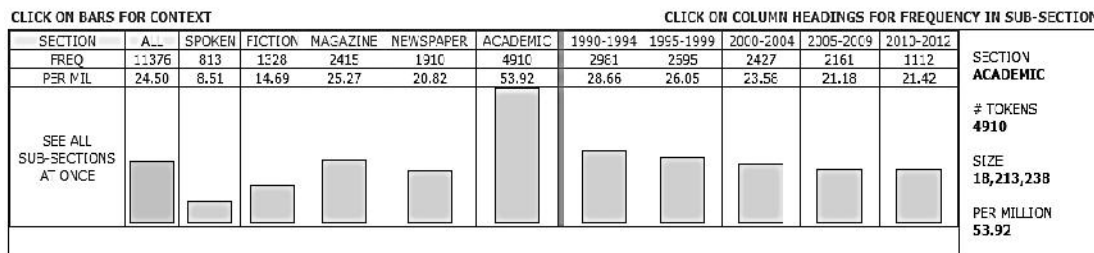
RESULTS AND DISCUSSION

Results

The Use of 'etc' in COCA

Figure 1 shows that the use of 'etc' in American English is mostly found in academic writing with no significant difference in the period of time. It has 11376 hits in which 4910 (53.6%) found in academic writing.

Figure 1
The use of 'etc' in COCA

The Use of 'etc' in BNC

Figure 2 shows that the word 'etc' is mostly found in written text compared to spoken text with a significant difference. It has 5015 hits in which 4961 hits found in academic writing and only 74 hits found in spoken language.

Figure 2
The use of 'etc' in BNC based on the spoken and written text

Spoken or Written:				
Category	No. of words	No. of hits	Dispersion (over files)	Frequency per million words
Written	87,903,571	4,941	1,096/3,140	56.21
Spoken	10,409,858	74	31/908	7.11
total	98,313,429	5,015	1,127/4,048	51.01

Figure 3
The use of 'etc' in BNC based on the derived text type

Derived text type:				
Category	No. of words	No. of hits	Dispersion (over files)	Frequency per million words
Unpublished written material	4,466,673	825	106/251	184.7
Other published written material	17,924,109	1,443	364/710	80.51
Academic prose	15,778,028	983	219/497	62.62
Non-academic prose and biography	24,178,674	1,475	291/744	61
Newspapers	9,412,174	114	62/486	12.11
Other spoken material	6,175,896	70	27/755	11.33
Fiction and verse	16,143,913	96	54/452	5.95
Spoken conversation	4,233,962	4	4/153	0.94
total	98,313,429	5,015	1,127/4,048	51.01

Figure 3 shows that the word 'etc' is mostly found in academic prose and biography (62.62 frequency per million words) compared to non-academic prose and biography (61 frequency per million words) although the difference is little in number.

Discussion

The Use of 'etc' in COCA and BNC

If we compare the data from COCA (see Figure 1) and BNC (see Figure 2), it can be seen that the hits of 'etc' in COCA is much bigger than the hits of it in BNC. Figure 1 shows that the use of 'etc' in American English has 11376 hits. Compared to Figure 1, Figure 2 shows that the word 'etc' has only 5015 hits. It is because the two has different size. COCA has 410+ million words, which is more than four times as large as the BNC which has 100 million words. As a result, it often provides data for lower-frequency constructions that are not available from the BNC.

Beside the size, we can compare the data from COCA and BNC based on their genre balanced. The BNC is 10% spoken/90% written, while in COCA the corpus is nearly evenly divided (20% in each genre) between spoken, fiction, popular magazines, newspaper, and academic. As a result, it can be seen in Figure 2

that the word 'etc' is mostly found in written text compared to spoken text with a significant difference. It has 5015 hits in which 4961 hits found in academic writing and only 74 hits found in spoken language.

Regarding the derived text type, Figure 1 shows that the use of 'etc' in American English is mostly found in academic writing with no significant difference in the period of time. It has 11376 hits in which 4910 (53.6%) found in academic writing. In addition, Figure 3 shows that the word 'etc' is mostly found in academic prose and biography (62.62 frequency per million words) compared to non-academic prose and biography (61 frequency per million words) although the difference is little in number.

All the data show that the word 'etc' used mostly in written academic text. Moreover, in American English, the use of 'etc' is mostly found in academic writing with no significant difference in the period of time.

The Use of 'etc' in Corpus versus the Guidelines in Academic Writing

Contrary to the results of the use of 'etc' in COCA and BNC, there are three writing centers in universities in United States explicitly stating to avoid the use of 'etc' in academic writing, i.e. Walden University,

University of Montana, and Empire State College.

The writing center of Walden University stated that according to the APA, Latin abbreviations (e.g., i.e., etc., et al.) should only appear in parenthetical material; otherwise, English translations of these abbreviations should be used instead. The only exception is *et al.*, which can be used both in normal and parenthetical material. For *etc*, the students can write 'and so forth' instead of 'etc'.

In the one hand, the writing center of University of Montana also stated that because they are often misunderstood and misused, Latin abbreviations should be used sparingly, if at all, in formal writing. Most sources agree that writers can make use of abbreviations in footnotes and references but should otherwise avoid them. In most cases, the English translation of the abbreviation can be substituted with no detrimental consequences. If an abbreviation must be used, writers must be certain they choose the correct abbreviation. A good way to ensure that the abbreviation chosen is actually the one the writer wants is to substitute the

English translation of that abbreviation and see if the sentence still makes sense. So, instead of using 'etc', the students as the writers can substitute it by 'and so forth'.

Different from the two writing centers mentioned before, Empire State College clearly mentioned that the standard abbreviations in writing, which are often abbreviations of Latin phrases such as, i.e. (*id est*) - that is, cf. (*confer*) – compare, e.g. (*exempli gratia*) - for example, et al. (*et alii*) - and others, etc. (*et cetera*)- and other things, and N.B. (*nota bene*) - note well, are usually used in tables, source citations, and comments inside parentheses in formal writing. The students should avoid their use in academic or formal writing.

The Use of 'etc' versus 'and so forth' in Academic Writing

As it is suggested by the writing center of Walden University and University of Montana, 'etc' should be substituted with the English translation 'and so forth'. To give a comparison between the use of 'etc' and 'and so forth' in academic writing, following figures are the results of the use of 'and so forth' in COCA and BNC.

Figure 4
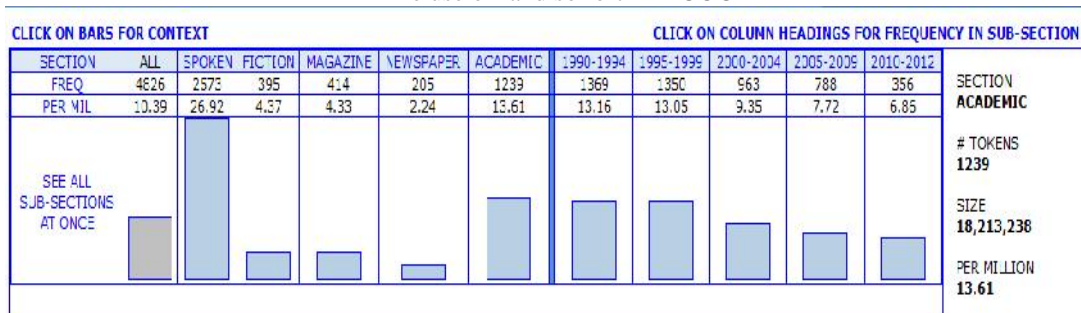The use of 'and so forth' in COCA

	ALL	SPOKEN	FICTION	MAGAZINE	NEWSPAPER	ACADEMIC	1990-1994	1995-1999	2000-2004	2005-2009	2010-2012
FREQ	4826	2573	395	414	205	1239	1369	1350	963	788	356
PER MIL	10.39	26.92	4.37	4.33	2.24	13.61	13.16	13.05	9.35	7.72	6.85

SECTION **ACADEMIC**

TOKENS **1239**

SIZE **18,213,238**

PER MILLION **13.61**

Figure 5
The use of 'and so forth' in BNC based on the derived text type

Derived text type:				
Category	No. of words	No. of hits	Dispersion (over files)	Frequency per million words
Other spoken material	6,175,896	178	76/755	28.82
Academic prose	15,778,028	125	67/497	7.92
Non-academic prose and biography	24,178,674	97	52/744	4.01
Other published written material	17,924,109	70	40/710	3.91
Spoken conversation	4,233,962	16	9/153	3.78
Fiction and verse	16,143,913	43	29/452	2.66
Unpublished written material	4,466,673	9	4/251	2.01
Newspapers	9,412,174	9	9/435	0.96
total	98,313,429	547	286/4,048	5.56

Table 1
The use of 'etc' versus 'and so forth' in academic writing

	'etc'	'and so forth'
COCA (Academic)	4910 hits	1239 hits
BNC (Academic)	62.62 frequency per million words	7.92 frequency per million words

Table 1 shows that despite the guidelines of submitting the Latin abbreviation 'etc' into the English translation 'and so forth', the use of 'etc' is more popular in academic writing compared to 'and so forth'.

CONCLUSION

The corpus study reveals three findings, i.e. the use of 'etc' in COCA and BNC, the use of 'etc' in corpus versus the guidelines in academic writing, and the use of 'etc' versus 'and so forth' in academic writing.

First, the data from COCA and BNC show that the word 'etc' is used mostly in written academic text. Moreover, in American English, the use of 'etc' is mostly found in academic writing with no significant difference in the period of time.

Second, contrary to the results of the use of 'etc' in COCA and BNC, there are three writing centers in universities in United States explicitly stating to avoid the use of 'etc' in academic writing, i.e. Walden University, University of Montana, and Empire State College. It is stated that instead of using 'etc', the students as the writers can substitute it by 'and so forth'.

Third, despite the guidelines of submitting the Latin abbreviation 'etc' into the English translation 'and so forth', the use of 'etc' is more popular in academic writing compared to 'and so forth'.

On the basis of findings above, it can be concluded that despite the academic writing guidelines in several university's writing centers in United States stating to avoid the use of 'etc' and substitute it with 'and so forth' in academic writing, the corpus data show that 'etc' used mostly in academic writing compared to the other types of writing and more popular than the use of 'and so forth'.

REFERENCES

Bailey, S. (2003). *Academic Writing: A Practical Guide for Students.* London: Routledge Falmet.

J.M., S., & Feak, C. (2004). *Academic Writing for Graduate Students.* Michigan: The University of Michigan Press.

Nga, N. (2009). Academic English at Tertiary Level: What, Why,

How. *VNU Journal of Science, Foreign languages 25*, 112-117.

Oshima, A., & Hogue, A. (1999). *Writing Academic English.* London: Addison Wesley Longman.

Paltridge, B., & Starfield, S. (2007). *Thesis and Dissertation Writing in a Second Language.* New York: Routledge.

Swales, J. M. (1990). *Genre Analysis: English in Academic and Research Setting.* Cambridge: Cambridge University Press.