

# Uncovering Informal Expressions Used in English NNS Thesis Writing

Hao-Jan Howard Chen (National Taiwan Normal University, Taiwan)

Hyland and Jiang (2016) explored the perception that since informality has now invaded a large range of written and spoken domains of discourse, academic writing has also followed this trend to accept more informal expressions. They, however, found that only a small increase in the use of some informal features and academics in general still hold high standards toward formality in academic writing.

Given that the informal expressions are widely used in a wide range of written and spoken discourses, many L2 learners were found to use various informal expressions in formal academic writing, as pointed out by Gilquin and Paquot (2008). However, few empirical studies systematically investigated these great challenges faced by L2 writers. To better understand L2 learners' real difficulties in using these various informal features, this study compared two NS and NNS academic corpora to uncover informal expressions used by NNS writers. The NNS academic corpus is consisted of 210 Taiwanese graduate students' theses in computer science. The NS corpus includes 171 papers from well-known international journals of computer science. By comparing these two sets of data, it is possible to pinpoint the differences between novice academic writers and professional writers in computer science.

There are various possible methods we can further compare the NNS and professional academic corpora. In this study, the research focus would be on the unusual high-frequency informal expressions used by NNS academic writers. It is expected that some of these high-frequency informal expressions can be identified from the learner corpus. Keyword and key clusters identification method was found useful by several researchers (Mahlberg, 2007; Fischer-Starcke, 2009). In this study, WordSmith 7 created by Mike Scott was adopted and the keyword and key clusters extraction functions were used. The word list and the clusters (2-word, 3-word, and 4-word clusters) of these two corpora were first compiled. Then these lists produced by professional and NNS academic writers were compared using the keyword tool of WordSmith.

First, by running the keyword analysis, several informal items were uncovered from the NNS academic corpus. These include connector 'besides' and several informal verbs such as 'get', 'wants', 'means'. Some more examples are shown in Table 1.

Key word	Freq. (Professional)	RC. Freq. (Learners' Theses)	Log_L
besides	37	344	-242.99
according	512	1188	-187.62
get	262	726	-164.87
wants	23	206	-142.90
lot	32	189	-102.05
means	440	838	-75.22

Table 1. Some Informal Words Used by NNS Academic Writers

For the 2-word clusters, some informal phrasal verbs such as 'find out' were found. Some informal verbs such as 'wants' and 'get' were also widely used in this learner corpus. These items were in general informal but many Taiwanese academic writers were not aware of their informality. In addition, in the theses written by graduate students, there are many

examples of first personal plural pronoun 'we'. These include 'we can', 'our system', 'we proposed', 'we will'. Table 2 shows some of the key differences between NS and NNS.

Key word	Freq. (Professional)	RC. Freq. (Learners' Theses)	Log_L
we can	678	1730	-336.09
our system	19	318	-279.58
wants to	15	200	-163.65
get the	51	271	-134.88
find out	10	153	-130.97
lot of	22	173	-112.11
we proposed	19	161	-108.60
it means	11	131	-102.87
a lot	32	184	-97.40
we will	429	872	-96.71
can get	12	123	-90.83

Table 2. Some 2-word Informal Expressions Used by NNS Academic Writers

For 3-word clusters, some informal expressions such as 'a lot of', 'reason is that', 'and so on', and 'deal with the' were found. More examples can be found in Table 3. Similar to what we found in the 2-word clusters, there are many expressions containing first personal plural pronoun 'we'. It seems that graduate students were giving credits to the whole research team/lab and many also believe that the thesis work should be shared with their thesis advisers.

Key Clusters	Freq. (Professional)	RC. Freq. (Learners' Theses)	Log_L
a lot of	22	169	-108.09
we have to	25	148	-80.05
we can see	57	194	-60.54
reason is that	14	97	-58.34
we can find	13	93	-57.05
to get the	20	99	-46.41
we propose a	74	199	-42.79
and so on	55	152	-34.35
we can use	32	107	-32.60
we will discuss	11	62	-32.33
we set the	15	70	-31.12
deal with the	21	82	-30.33
we can observe	15	66	-27.72

Table 3. Some 3-word Informal Expressions Used by NNS Academic Writers

For 4-word clusters, some informal expressions such as 'to find out the', 'it means that we', 'reason is that we', 'that is to say', and 'a lot of time' were found. More examples can be found in Table 4. Similar to what we found in the other groups of word clusters, there are many expressions containing first personal plural pronoun 'we'. As pointed out in the previous section, students chose to give credits to the whole research team/lab or their thesis advisers. Thus, many chose to use 'we' in their thesis writing.

Key Clusters	Freq. (Learners' Theses)	RC. Freq. (Professional)	Log_L
in this these we	244	0	304.29
in this chapter we	160	0	199.54
to find out the	53	0	66.10
we will introduce the	40	0	49.88
we can get the	39	0	48.64
thesis we propose a	36	0	44.90
it means that the	35	0	43.65
we can find that	34	0	42.40
reason is that the	33	0	41.15
we can obtain the	31	0	38.66
that is to say	30	0	37.41
we will discuss the	27	0	33.67
a lot of time	23	0	28.68
there are a lot of	21	0	26.19

Table 4. Some 4-word Informal Expressions Used by NNS Academic Writers

Based on the preliminary findings, it is clear that many L2 novice writers were not fully aware of register differences, and they tended to use some informal expressions in their formal thesis writing. It is thus important to prepare some learning materials for these EAP writers. Academic writing instructors can then use these materials and exercises targeting on informality issues to draw L2 learners' attention to these informal items.

## References

- Fischer-Starcke, B. (2009). Keywords and frequent phrases of Jane Austen's *Pride and Prejudice*: A corpus-stylistic analysis. *International Journal of Corpus Linguistic* 14(4), 492-523.
- Gilquin, G. & Paquot, M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction* 1 (1), 41-61.
- Hyland, K. & Jiang, F. K. (2017). Is academic writing becoming more informal?. *English for Specific Purposes* 45, 40-51.
- Mahlberg, M. (2007). Clusters, key clusters and local textual functions in Dickens. *Corpora* 2(1), 1-31.