A CASE STUDY CORPUS FOR ACADEMIC ENGLISH WRITING BY NNS AUTHORS

Alejandro Curado Fuentes
Universidad de Extremadura

1. Introduction

The concern with academic English writing as performed by NNS (non-native speakers of English) has led to extensive literature and research over the past three decades. The origins were probably best represented in the British Council's realization of the term EAP (English for Academic Purposes), in 1975, to refer to "interdisciplinary studies in relation to existing practices and institutions" (Brumfit, 1984: 17), or to "exchange of knowledge [...] according to specific features of specialised subject fields" (Baunmann, 1994: 1). Different perspectives ever since on the matter have been adopted, such as academic vocabulary knowledge (e.g., Martin, 1976), or writing in the genre conventions (e.g., Swales, 1995). Various schools or associations have also formed as a result, for instance, in Britain, BALEAP, from which significant studies have developed (e.g., corpus data-driven learning, academic phraseology, discourse analysis in the academic setting, etc —see, among others, work by Johns, 1993; Howarth, 1998; Lockett, 1999, etc—).

EAP often stands to the test in the achievement of foreign undergraduate and graduate writing proficiency for specialized fields. The focus is on university compositions or essays where L2 learners ought to go through the re-writing procedures of content clarification, structure revision, lexical-grammatical revision, and so forth, and where aspects of register and genre conventions play significant reference roles. Less consideration seems to be given, by comparison, to NNS research writing for publication aims, although just to give two examples, Burrough-Boenisch (2003; 2005) examine proof-reading procedures in this line, observing, among other aspects, the important position of rhetorical organization in the reviewing process. As far as my knowledge goes, however, no work has been done on corpus material with the aim of analyzing L2 writing in the last phase prior to publication.

In this paper, the aim is to examine corpus-based analysis with both NS and NNS material. The texts available for the corpus analysis are those authored by Spanish writers in Computer Science in their final versions; however, they are accessed prior to the journal editors' last review. The corpus examination has been done by comparing the texts with NS material from a selection of the BNC (British National Corpus) Sampler (Burnard and McEnery, 1999). The chief objective in the process has been to identify both similarity and divergence in terms of the significant lexical items used, especially academic lexical items and / or rhetorical-lexical items. Word co-occurrence and use probability in the contrasted contexts determine academic competence, since the mastery of specific lexical patterns should indicate specialized writing (cf. Hoey, 2005). Based on the literature described below, an attempt at assessing two general hypotheses on NNS writing is also included.

2. Theoretical framework

NNS academic writing involves some specific traits and concerns with both the product and process of composition. Jafarpur (1996: 89), for instance, observes L2 writers' performance in

comparison with native writing command, and measures the degree of NNS lexical-grammatical knowledge in terms of the exact word test (i.e., lexical precision), in which NS writers tend to score higher. However, Jafarpur (1996: 91) also observes that NNS writing need not be uniformly worse than NS, and that L2 writers tend to have better content knowledge than linguistic command. This fact, claimed by some authors (e.g., Storch and Tapper, 1997) as characteristic of NNS writers, seems to hold for the NNS prioritization of first content and then grammar and lexis, while the NS focus seems to be first on structure and then content.

The scholarship consulted also generally agrees that L2 writers can do academic work just as well (or as bad, if done poorly) as native writers. The main qualitative difference lies in the production of particular linguistic-discursive features by the NNS group, who, as Burrough-Boenisch (2003) observes, may differ in relation to lexical choice, grammar, and cohesion / structure. Thonus (2004) claims that the key in the reviewing process is to make L2 authors aware of the overall need for content and structure revision, since they may be more concerned with lexis and grammar, and pay less attention, in turn, to the overall design of the paragraphs and sentences.

Some intrinsic features perceived in L2 writing seem to derive from the use of higher frequency rates for specific lexical / rhetorical items. Hinkel (1997: 361), working with NNS writers from China, Japan, Korea and Indonesia, discovers their greater reliance on some types of hedges, pronouns, and other features that clearly stand out when compared with NS writing. In some cases, the comparison with a reference academic corpus (e.g., Biber, 1988) points out deviant usage in NNS performance, such as the predominance of indefinite pronouns, seldom used in academic writing (Hinkel, 1997: 378). Other case studies demonstrate the abuse of certain adverbial features by NNS writers, especially amplifiers, emphatic and manner adverbs, the result of the influence of L2 informal conversation on their writing (Hinkel, 2003: 1065). Nonetheless, NNS and NS writers may achieve similar percentages of lexical and rhetorical items, such as hedges, in their compositions (e.g., Burrough-Boenisch, 2005: 29), but the distinction in the reviewing process is made in terms of lexical choice, which becomes less varied and / or precise in the case of the NNS authors (e.g., the distinction between *appear to be* and *seem to be* is less clear in NNS writing, according to Burrough-Boenisch [2005: 33]).

In the use of such words thus a main difference rises between NNS and NS writing performance. This academic language, defined as early as Jones and Sinclair (1974:16) as "any word or group of words [...] with a pattern of collocation, or regular co-occurrence with other items", establishes patterns where collocations—or colligations, if the pattern involves a related grammatical aspect—tend to characterize academic discourse (Cowie, 1998: 6). An example would be the aforementioned group of hedges or the percentage rate of passive versus active statements in a quantitative corpus analysis that takes these items into account in order to determine academic register or dimensions (e.g., Biber, 1988; Biber, Conrad and Reppen, 1998).

In agreement with Hoey (2005: 182), being competent in a given academic discipline is closely related to having "mastery of collocations, colligations and semantic associations of the vocabulary [...] of the domain-specific and genre-specific primings". Lexical choice competence, i.e., being able to submit in writing those items that adjust to the contextual / co-textual space, is of prime importance. The context, whether specialized or restricted to a given genre, should influence writers to use certain constructions in their text. An example given by Hoey (2005:48-49) is the noun *consequence*, which appears as the head of nominal groups in 98 percent of its occurrences within nominal groups, according to Hoey's corpus measurement of journalistic texts. Obviously enough, any reader of such newspaper texts,

would be inclined to expect this lexical use, and would likely write the construction according to the pattern spurred in such contexts (e.g., *the consequence of*, *as a consequence of*, etc).

Academic writing need not be exclusively characterized by the linguistic-discursive traits used by NS writers. In a growing global community for academic and professional exchanges of knowledge in English, in agreement with some authors (e.g., Baker, 2004), the corpus analysis done with discourse traits can "only provide statistic snap-shots that give the appearance of stability but are bound to the context of the data set" (Baker, 2004: 10). The production of linguistic-discursive items in academic writing may act as a mechanism to describe the linguistic system, compared by Halliday (1991: 32) with a weather system where "each day's weather affects the climate, however infinitesimally, either maintaining the status quo or helping to tip the balance towards climatic change". As a consequence of this changing nature of discourse, given the differences in writing styles and NNS variations, among other factors, the corpus analysis should not be taken as a prescriptive tool, but rather as a fundamentally descriptive scope in which academic writing is compared according to the statistically significant linguistic-discursive use made by both NS and NNS writers.

The corpus analysis below should thus focus on the lexical features or items that significantly differ (or match) from one corpus to another. The aim is to analyze the type of academic discourse traits prevailing in the texts, and whether such aspects may prove to characterize particular elements in NNS writing or imply existing NNS variation. In this sense, two hypotheses may be tested: 1. NNS lexical-grammatical patterns are less varied than NS; 2. NNS have different hedges from NS. The literature is concerned with indirectness in NNS academic writing (e.g., Hinkel, 1997; Burrough-Boenisch, 2005), while some lexical items in certain discourse positions (e.g., adverbials) trigger chief academic stance attributes (cf. Hinkel, 2003; Biber, Csomay, Jones and Keck, 2004).

The investigation of some of these lexical features should therefore aim to resolve some of the NNS text peculiarities. The reference corpus, a selection of academic texts from the BNC Sampler (Burnard and McEnery, 1999), should provide valuable contrastive data, not only for the determination of specific academic traits in the lexical items observed, but also for the exploration of related factors, such as possible disciplinary variation, or even genre-based differences—e.g., North (2005) identifies variation in the Humanities in terms of theme and rheme sequencing, used as early as undergraduate essays—. Furthermore, because of the tendency among NNS to rely on discourse markers, hedges, and paraphrasing, especially among Spanish writers both in their mother language and L2 (e.g., Lahuerta Martínez, 2004; Morales, Cassany and González-Peña, 2007; Cademártori, Parodi and Venegas, 2007), the lexical investigation should take functional items as the pillars for scrutiny.

3. Corpus method

The L2 writing corpus is a collection of nine journal articles before final edition and publication in Computer Science journals. These articles were written by groups of authors from University of Extremadura who ranged in number between 4 and 5 people.

The total number of words or tokens is 25,931, with 6,856 distinct words or types, and a STTR (Standardised Token-to-Type Ratio, as measured in a rate of every 1,000 words) of 48.03 words. This lexical density is high if compared to other written registers (e.g., journalistic texts), even to the reference writing corpus used from the BNC Sampler. Figure 1 displays such features from the NNS corpus in contrast with those from the reference corpus of NS writing.

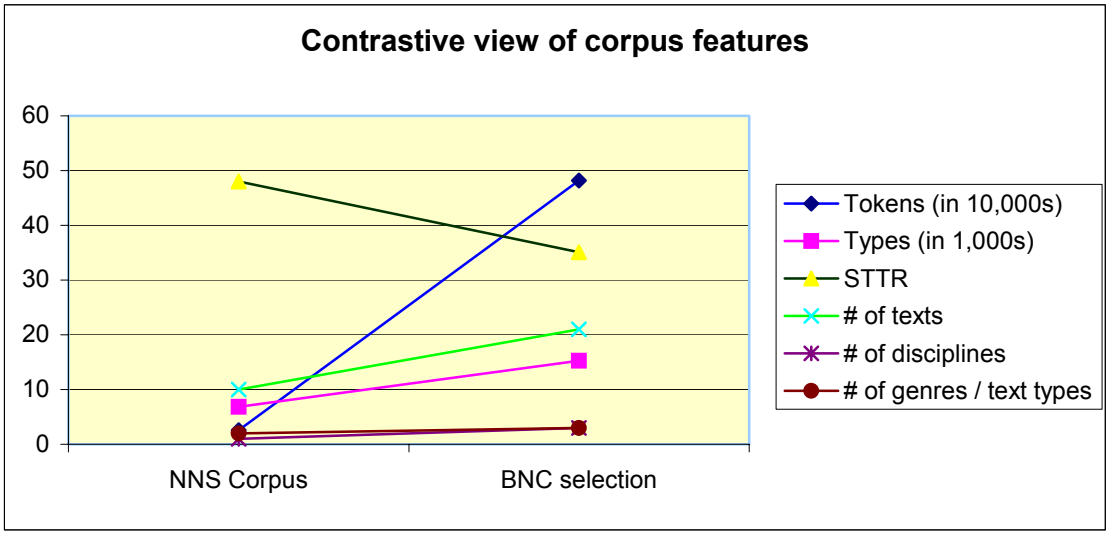

Figure 1: Contrastive view of corpora features
STTR = Standardised Token-to-type Ratio

The reference corpus is larger than the NNS collection for contrastive purposes, and covers some various disciplines or subject areas (12 texts from Arts and Sciences, five in Social Sciences, and four in Arts and Humanities), as well as some genres or text types (magazine articles, journal articles and reports). The approach to the corpus starts with the comparison of some function / grammatical words. Lexical frequency is taken as a key measurement reference in this corpus examination, and so, the most frequent grammatical words among the first 100 are selected. Then, their position is checked in the BNC selection word list to see their similarity or differentiation in terms of frequency. This first stage also includes some words from the top 100 used in the reference corpus (e.g., the possessive *our* in position 92, while it appears in position 102 in the target NNS texts).

The second step in the corpus analysis deals with the core contrastive exploration of lexical items as attributes of academic stance (cf. Biber et al., 2004). The academic register is at least partly checked by means of collocational strength degrees—i.e., in terms of content, grammatical-discursive position, semantic space, and text-item relationships (cf. Hoey, 2005)—. Finally, a third step in the study corresponds to the classification of the findings according to possible textual and contextual variables.

4. Results

The two corpora, target and reference, present the same three function words at the top of their frequency-based word lists: *the*, *of*, *and*. The coincidence is not surprising, as both collections contain written academic material, and many grammatical words that characterize this register are the same ones. Ranked positions for these words in the frequency lists can be contrasted. At number four, for instance, the NNS corpus has the preposition *in*, while the BNC selection has *to* (these words are actually reversed at position five). This contrastive arrangement serves as a basis for the examination of frequency similarities and divergence. The analysis of most grammatical words at such high positions on the lists is bound to produce significant results, while an in-depth view of other words may be less useful. An

example would be the article *the* at first position, that, because of its dependent use on other grammatical node words, is not taken as a reference.

Table 1 shows the words used as reference for the corpus analysis after secondary items (mostly articles and some conjunctions) and content words have been removed. It also shows the ranking differences (with signs “>” and “<”) and similarities (“=”) of the words on the lists. The arrows “>” and “<” indicate whether the word is used more frequently within the NNS corpus (>) or in the BNC selection (<). In the case of the double arrows (“>>” and “<<”), these mean there is a difference of more than 10 word list positions for that word. Needless to say, the comparison of frequencies is not based on absolute numerical figures, but on relative word frequency within each corpus. The distinction is considered important for the second stage of the analysis, where similar or divergent word use may be compared and paired up with this frequency examination.

WORD	NNS Corpus for Case Study	BNC selection
IN	>	
TO		<
FOR	>	
AS	>	
THAT	>	
IS		<
ON	>	
BY	>	
WITH	>	
BE		<
THIS	>	
ARE	>	
OR	>	
WE	>>	
HAVE	>	
CAN	>>	
AT	>	
USE	>>	
FROM		<
WHICH	=	
BUT	>	
ALSO	>>	
NOT		<
WILL	>	
SHOULD	>>	
INTO	>>	
ONE	>	
BOTH	>>	
EACH	>>	
SO	>>	
SOME		<<
MAY		<<
SUCH		<<

I	<<
IF	<<
OUR	<

Table 1: Contrastive view of words according to frequency within corpora

As shown in Table 1, most words are prepositions, non-qualifying adjectives, pronouns and conjunctions. In addition, auxiliary / modal verbs (i.e., de-lexicalized), discourse markers, and procedural / structure words (e.g., the noun *use*) are included, since they also operate structurally, and convey cohesion and "structure the text" (Hutchinson and Waters, 1981: 65). All these words constitute node or head words with which to observe lexical item behavior within and across the two corpora.

The words can then be observed in their use according to different statistical factors: First, according to absolute frequencies and distribution when the items present high rates. For example, with the preposition *in***,** the purpose marker *in order to* appears extensively in both corpora: 56 times in seven texts in the NNS corpus, and 80 times in 12 BNC texts. Such a construction would be an example of use similarity. Secondly, the comparison measures lexical item proportion in relation to the use of a collocating content word. For example, the expression *shown in + Figure / Table* yields a percentage rate in the NNS texts of 20 percent, which means that out of the 10 occurrences of the participle *shown*, two are *shown + in Figure / Table*. In contrast, the BNC selection includes this item in a proportion rising to 39.2 percent, i.e., 42 instances in 107 occurrences of *shown* are the mentioned expression.

The collocation *shown in + Figure / Table* illustrates proportional linguistic divergence between the two corpora. Divergence is estimated at a proportion of 15 percent or more. This threshold is arbitrary, but is calculated as a cut-off point on account of the small sizes of the corpora, especially the NNS one. The expression *shown in + Figure / Table* demonstrates variation in terms of the NS reference, i.e., NNS writing does not show the significant use of the given expression, which is more widely used in the BNC texts (thus, an NNS gap is observed). The absolute cases would be those in which any item used significantly in one set of texts has no representation in the other (e.g., with the preposition *to* the construction *to ensure that* appears 36 times in the BNC texts but none in the NNS corpus).

Thirdly, there are those lexical items with higher proportion percentages in NNS than in NS texts. An example with *in* is *in the case of*, with 33.3 percent in relation to *case* (15 times out of 45 occurrences of *case*), while the percentage is 16.9 percent (26 out of 153) in the BNC collection. These items tend to be considered, at least a priori, characteristic of NNS use, since they demonstrate less proportional significance in the BNC reference, where some may even be inexistent (e.g., derived from *in*, the item *in this sense* appears eight times in six papers, but none in the NS texts).

Figure 2 illustrates numerically the main findings of the analysis according to the three possibilities: Similar use, according to frequency and distribution; NNS gap, if the item is used below the percentage threshold in the NNS corpus; NNS use, if the item is used below the threshold in the NS selection.

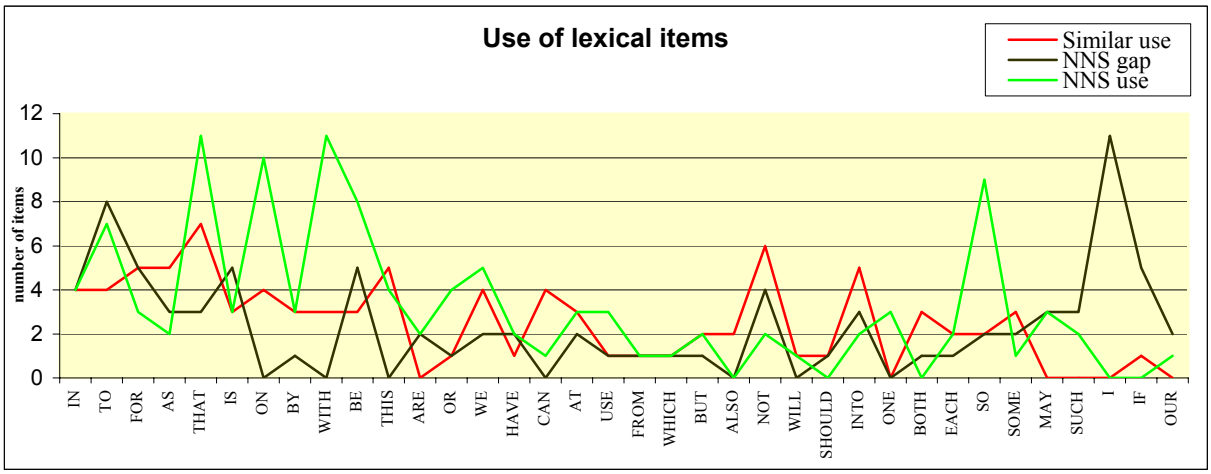

Figure 2: Contrastive view of lexical items in the corpora

"Number of items" in Figure 2 refers to the lexical items that involve not only collocations and colligations, but also semantic associations and textual collocations (cf. Hoey, 2005). Some examples of the four types of lexical use within the three different categories can be seen in Table 2. Collocation refers to the statistically significant co-occurrence of two or more words in the texts. The example with the preposition *to*, *appear** + *to be* is a collocation used similarly, as, in both corpora, its use is around 20 percent (20 for NNS and 20.4 for NS) in relation to the lemma *appear** (i.e., it includes all derived forms—*appears*, *appeared*, etc—). Colligation may involve frequency in the use of a given word class, such as nouns followed by the preposition *to* without the indication of purpose or reported speech (e.g., *access to*, *person to follow*, *of interest to readers*, *taught secrets to*, etc, but not *used the brush to*, or *asked the student to*). These items indicate an NNS gap, as they have a much smaller presence in NNS texts. Colligation may also include any given item in which a grammatical aspect is related to the collocation (e.g., *be* + *asked to*, conjugated in the present tense in NNS texts but not in NS).

The third type of lexical items refers to those collocations that have statistically significant semantic relationships called semantic associations. With the preposition *to*, the item *related to* + concept or issue is an NNS-use semantic association, whereas the example *to be seeking* + work is used only in NS texts (i.e., represents an NNS gap). Finally, textual collocation refers to a pattern characterized by the positioning of a lexical item at a given point or place in the text. The NNS Gap example *one of the most* + adjective appears at the beginning of sentences at a higher proportion in relation to the pattern *the most* + adjective in NS than NNS texts.[1]

[1] The proportion percentage for co-occurrence significance is not always calculated in relation to a content word, as the textual collocation *one of the most* + adjective, or the colligation noun + *to* demonstrate in Table 2 (in the case of this colligation, the aspect compared is the lack of reported speech or purpose in the item, as explained). Thus, proportion rates can derive from patterns within the lexical items when the proportion would be too low if calculated in terms of one single word. Examples in Table 3 include the subject collocation *if and only if*, which is considered in relation to the cluster *and only*, the genre colligation *I had* + participle, computed in relation to *I had*, or the subject colligation *is to be* + participle, measured in relation to *to be* + *participle*.

WORD USE	Similar Use	NNS Gap	NNS Use
Collocation	*Appear** + *to be* (20 / 20.4%)	*It is possible to* (8 vs. 28.2%)	*We observe that* (14.7 vs. 0%)
Colligation	*The basis for* (Direct Object) (26.3 / 17.6%)	Noun + *to* (no purpose / no reported speech) (1.2 vs. 26.5%)	*Be* + *asked to* (present tense) (61.5 vs. 0%)
Semantic Association	*In the field of* + area (20 / 11.5)	*To be seeking* + work (0 vs. 28%)	*Related to* + concept (76.9 vs. 26%)
Textual Collocation	*As a result of* (beg. paragraphs) (20 vs. 31.5%)	*One of the most* + adj. (beg. sentences) (4.3 vs. 23.2%)	*For this reason,* (beg. sentences) (20 vs. 2.9%)

Table 2: Examples of word use according to lexical item categories

This qualitative contrastive view[2] helps to support the third and final stage of the analysis, in which the items are classified according to potential variables influencing text production. As a preliminary comparison, the data from Table 1 and Figure 2 may provide information about word use expectations within the corpora. Figure 3 graphically displays a correlation between the two data sets; it compares the higher, lower or equal positions of the words in Table 1 with the production of that word in the form of lexical items in Figure 2, i.e., it compares word frequency with the extended use of that word as academic language in the texts.

[2] This classification of the data is by no means exhaustive in the aim to find all the lexical items derived from grammatical words in the texts. The top side of the word lists used is managed as a reference framework for lexical contrastive study so that main linguistic-discursive traits in academic discourse may be depicted. For example, the collocation *need to be* appears significantly in both corpora, but is omitted because it says little about academic stance. For the same reason, differing items between the two sources are discarded, such as the construction *continue to be*, unfound in the NNS texts.

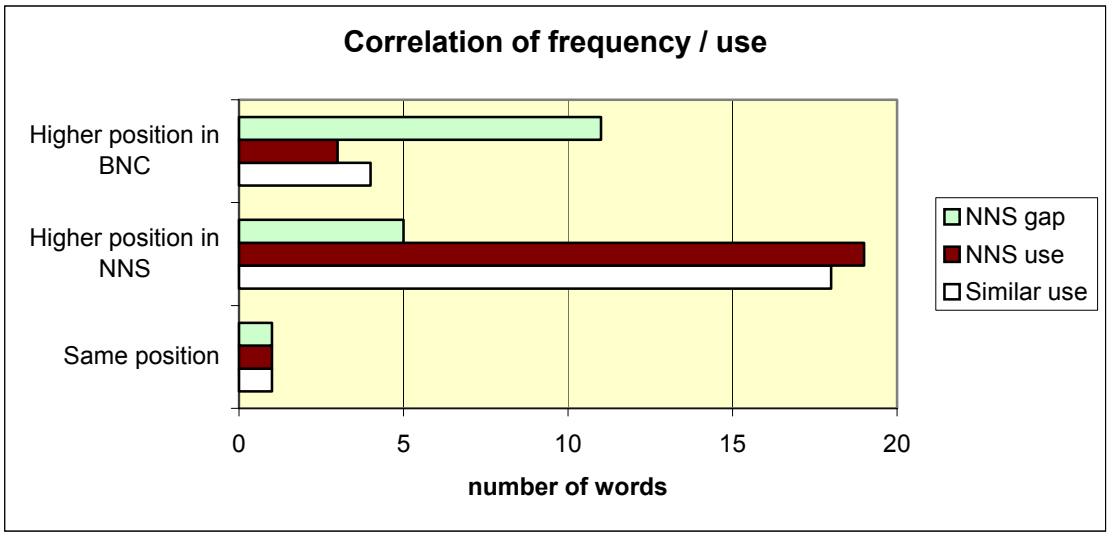

Figure 3: Correlation of word frequency and use in the texts

In Figure 3, "NNS use" and "NNS gap" refer to the higher number of instances found in the NNS and BNC texts respectively. "Number of words" refers to the computation of the words on the lists, which are counted once, except when their position differs by more than 10 spaces (i.e., ">>" or "<<" in Table 1) and their lexical use falls within one single category (i.e., Similar use, NNS gap or NNS use); in those cases, the words are counted twice or multiplied by 2. If the word presents the same amount of items in more than one category in the ">>" or "<<" counts, (e.g., the word *each*—which has the same degrees of NNS use and Similar use—or *may*—same NNS gap as NNS use—see Figure 2), then, the word is counted once. This measurement is arbitrary, but is done to distinguish those words that match in terms of higher relative frequency rates and greater lexical use within the given categories.

As shown in Figure 3, a great deal of word use similarity appears, especially in the case of those words with higher frequencies in NNS texts. Putting it differently, while the NNS writers seem to use some specific words very frequently, and these form characteristic NNS expressions, in many other cases, their production of lexical items materializes in similar proportions to the ones measured in NS writing. This observation is reinforced by the fact that some of the very frequent NNS vocabulary is used in a great proportion in the BNC texts, whereas the opposite is less common. Notwithstanding, as shown in Figures 2 and 3, there are several words that show characteristic lexical use in both NNS and BNC texts—three words are especially distinctive in NNS texts (*that*, *on*, *with*), and two in the BNC selection (*I, if*)—. It may be said then that, according to the lexical statistics and word use, while many items tend to be similar to NS, many others are likely to be distinctive in NNS writing.

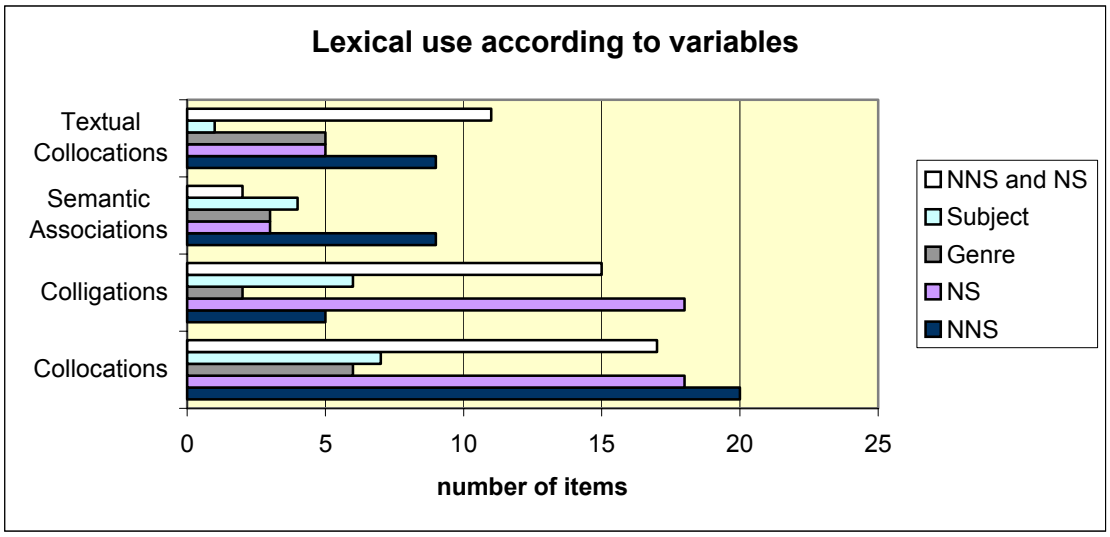

Figure 4: Types of lexical use according to textual and contextual variables

As a final contrastive examination of the data, lexical use is described in academic discourse according to the five variables stated in Figure 4. The lexical items are classified very similarly to how it was done in Table 2 above. The categories "NNS and NS", "NS", and "NNS" are the same respective ones as "Similar use", "NNS Gap", and "NNS use"; the new names are given in this case to refer to contextual factors or variables. As an example, NNS writers reveal the production of 20 collocations that NS authors lack or rely on less, while both NNS and NS convey 15 colligations in similar significant terms. In addition, there are various lexical items in the texts that are neither characteristic of NNS or NS writers, but that are there as a result of genre / text type, or subject / topic influence or constraints. Table 3 displays some examples found of such genre- and subject-related items in the four lexical categories.

Lexical use	*Genre*	*Subject*
Collocation	*Such as* + examples (56% --NNS papers)	*If and only if* (71.4% --BNC: Logic)
Colligation	*I had* + past participle (47% --BNC articles)	*is* + *to be* + past participle (17.8% --NNS: IT and e-learning)
Semantic Association	Be + *applied to* + people (25.6% --NNS paper introductions)	Be / appear + *on the right* + side (26.6% --NNS: IT and e-learning)
Textual Collocation	*There is no* (beg. paragraphs) (34.8% -- BNC articles)	*This form* + *be completed* (beg. paragraphs) (16.4% -- BNC: Computer surveys)

Table 3: Examples of lexical use influenced by genre and subject factors

As can be observed, genre and subject items are classified according to their proportional frequencies and respective relationships with the type of texts and subjects encompassed. For instance, the genre collocation *such as* followed by examples is much more common in the NNS papers than NS texts because of the greater need for exemplification in these journal articles, as demonstrated by the also significant presence of nouns like *example*

and *instance*, verbs like *illustrate* and *present*, the acronym *e.g.*, etc. In contrast, the genre colligation included in Table 3 is typical of the report register in the BNC selection, not found in the NNS types of texts.

The lexical items can thus be interpreted according to where they are used in the corpora. For instance, the genre semantic association in Table 3 illustrates how an expression is used in the NNS corpus because of the need to describe a specific research methodology line. The influence of the genre movement "Methodology" would lead to this type of lexical item. An example with subject-based items would be the colligation *is* + *to be* + participle, used characteristically in the description of future uses of technology. In contrast, a subject collocation like the one included in Table 3 is the product of the influence of topic in some BNC texts (in this case, Mathematical Logic). In all these instances found, we cannot ignore that fact that writers share patterns of lexical use that can be traced to such textual and contextual variables.

5. Discussion and conclusions

The findings from the contrastive analysis may be used for the evaluation of the two hypotheses about NNS writing mentioned above:

1. Do NNS writers produce a more restrained number of lexical items in academic discourse, i.e., are the items more pre-defined or predictable than in NS texts? According to Storch and Tapper (1997), the NS writers would work according to a rhetorical paradigm and then produce the items needed, while NNS authors are more likely to concern themselves with lexical-grammatical items. Obviously enough, because of the restrained type and small size of the NNS corpus used in this study, the answers to these questions cannot be unbiased or absolute. Given the information obtained on characteristic patterns, it is clear that the NNS authors analyzed demonstrate linguistic-discursive competence in terms of their wide use of distinctive, NS, and genre- / -subject related academic items (e.g., Figures 3 and 4). Hypothesis 1 can be confirmed in relation to the number of characteristic NNS items and cannot be contradicted, while the extensive NNS use of academic patterns measured in this case study may provide valuable feedback for further research that investigates this hypothesis in depth.
2. Are hedges used differently by NNS writers, i.e., do these authors produce a more pre-defined / predictable set of hedges? Neff, Ballesteros, Dafouz, Martínez, Rica, Díez and Prieto (2004) claim that some specific modals are overused in NNS texts (e.g., *can*), and sometimes combined in items that are unfound in NS writing (e.g., *we can wonder* to express doubt). Notwithstanding, a corpus of NNS writing like ICLE (International Corpus of Learner English), used by Neff et al. (2004), has little to do with the NNS corpus in this paper, where all the authors demonstrate proficient skills in English writing. Only in the higher frequency of *can* in contrast with the rest of the modals might a slight parallelism occur with the ICLE data, while, according to the measurement with this modal, a common colligation like *can be* + past participle, used at proportionally similar rates by both NNS and NS writers, counterbalances the ICLE findings. Hypothesis 2 may thus be contradicted in this sense, whereas it stands to reason that the NNS analysis in this paper is a case study and needs future contrastive work.

Because of such limitations in size and scope, this paper may serve as a first step or basis for further corpus-based contrastive examination of NNS academic writing aimed at

publication. In particular, the analysis may be projected at the elaboration of stylistic guidelines for NNS writers where they may find clear information on those linguistic-discursive items characteristic of academic stance. The material should thus include both effective NNS and NS writing examples, providing different types of feedback that may ever become handy as possible re-writing options. I, for one, being an NNS writer, can go back now to my own writing of this paper, and find both instances of NNS use and NNS gap that may encourage the modification of some sentences and / or lexical items. One example is *in relation to*, appearing in my paper 4 times—it could be replaced by the NS item *with regard to* or NNS *with respect to*—. Another example may be the substitution of *seem to be* by *appear to be*, etc. The items may be classified according to frequency in NS, NNS texts, genre, and subject / topic.

Then, the material may especially find a target audience among scientific-technical NNS writers. As a casual reviewer of colleagues' papers in Computer Science, I believe that rather than paper reviewing services, a more valuable information that we can give these NNS authors can be in the form of thesaurus-like and / or stylistic / discursive handout-like resources either in paper or digital form for academic language effectiveness. The scope would entail a descriptive, and not prescriptive, approach to the material in a way that, for example, NS colligations may be offered as options to take in the process of writing (e.g., using nouns followed by *to* to express directions in discourse, not just purpose or reported speech—see Table 2—). Likewise, textual considerations may be made as regards genre and / or subject traits; for example, sentences such as *There is no easy answer* or *there is no point in* may provide suitable alternatives for NNS authors in their reflections about a given argument or topic as they may introduce ideas into a new paragraph (i.e., a textual collocation in Table 3). The overall objective is, in agreement with Bhatia, Langton and Long (2004: 205), the strengthening of text processing and production competence by the increase in linguistic discursive awareness of basic generic principles and lexico-grammatical resources.

References

Baker, P. (2004): *Public Discourses of Gay Men*, London, Routledge.

Baunmann, K.D. (1994): "The significance of the statistical method for an interdisciplinary analysis of professionalism of texts", in M. Brekke, Ø. Andersen, T. Dahl and J. Miking (eds.) *Applications and Implications of Current LSP Research*, Bergen, University of Bergen Publications, 1-13.

Bhatia, V.K., N.M. Langton and J. Long (2004): "Legal discourse: Opportunities and threats for Corpus Linguistics", in U. Connor and T.A. Upton (eds.) *Discourse in the Professions: Perspectives from Corpus Linguistics*, Amsterdam, John Benjamins, 203-234.

Biber, D. (1988): *Variation across Speech and Writing*, Cambridge, Cambridge University Press.

Biber, D., S. Conrad and R. Reppen (1998): *Corpus Linguistics. Investigating language structure and use*, Cambridge, Cambridge University Press.

Biber, D., E. Csomay, J.K. Jones and C. Keck (2004): "A corpus linguistic investigation of vocabulary-based discourse units in university registers", in U. Connor and T.A. Upton (eds.) *Applied Corpus Linguistics. A Multidimensional Perspective*, Amsterdam, Rodopi, 53-72.

Brumfit, C.J. (1984): *Communicative Methodology in Language Teaching*, Cambridge, Cambridge University Press.
Burnard, L. and T. McEnery (1999): *The BNC Sampler*, Oxford, Oxford University Press.
Burrough-Boenisch, J. (2003): "Shapers of published NNS research articles", *Journal of Second Language Writing*, 12, 223–243.
Burrough-Boenisch, J. (2005): "NS and NNS scientists' amendments of Dutch scientific English and their impact on hedging", *English for Specific Purposes*, 24, 25–39.
Cademártori, Y., G. Parodi and R. Venegas (2007): "El discurso escrito y especializado: Las nominalizaciones en manuales técnicos", in G. Parodi (ed.) *Lingüística de Corpus y Discursos Especializados: Puntos de Mira*, Valparaíso, Chile, Ediciones Universitarias de Valparaíso, 79-96.
Cowie, A.P. (1998): "Introduction", in A.P. Cowie (ed.) *Phraseology. Theory, Analysis and Applications*, Oxford, Clarendon Press, 1-20.
Halliday, M.A.K. (1991): "Corpus studies and probabilistic grammar", in K. Aijmer and B. Altenberg (eds.) *English Corpus Linguistics*, London, Longman, 30-43.
Hinkel, E. (1997): "Indirectness in L1 and L2 academic writing", *Journal of Pragmatics*, 27, 361-386.
Hinkel, E. (2003): "Adverbial markers and tone in L1 and L2 students' writing", *Journal of Pragmatics*, 35, 1049–1068.
Hoey, M. (2005): *Lexical Priming. A New Theory of Words and Language*, London, Routledge.
Howarth, P. (1998): "Phraseology and second language proficiency", *Applied Linguistics*, 19*,* 24-44.
Hutchinson, T. and A. Waters (1981): "Performance and competence in ESP", *Applied Linguistics*, 2*,* 56-69.
Jafarpur, A. (1996): "Native speaker performance validity: In vain or for gain?", *System*, 24*,* 83-95.
Johns, T. (1993): "Data-driven learning: An update", *TELL & CALL*, 3*,* 23-32.
Jones, S., and J. Sinclair (1974): "English lexical collocations", *Cahiers de Lexicologie*, 24*,* 15-61.
Lahuerta Martínez, A.C. (2004): "Discourse markers in the expository writing of Spanish university students", *Ibérica*, 8*,* 63-80.
Lockett, A. (1999): "From the general to the specific: What the EAP tutor should know about academic discourse", in H. Bool and P. Luford (eds.) *Academic Standards and Expectations. The Role of EAP*, Nottingham, Nottingham University Press, 46-58.
Martin, A.V. (1976): "Teaching academic vocabulary to foreign graduate students", *TESOL Quarterly*, 10*,* 91-97.
Morales, O.A., D. Cassany and C. González-Peña (2007): "La atenuación en artículos de revisión odontológicos en español: Estudio exploratorio", *Ibérica*, 14*,* 33-58.
Neff, J., F. Ballesteros, E. Dafouz, F. Martínez, J.P. Rica, M. Díez and R. Prieto (2004): "Formulating writer stance: A contrastive study of EFL learner corpora", in U. Connor and T.A. Upton (eds.) *Applied Corpus Linguistics. A Multidimensional Perspective*, Amsterdam, Rodopi, 73-90.
North, S. (2005): "Disciplinary variation in the use of theme in undergraduate essays", *Applied Linguistics*, 26*,* 431-452.
Storch, N. and J. Tapper (1997): "Student annotations: What NNS and NS university students say about their own writing", *Journal of Second Language Writing*, 6*,* 245-264.
Swales, J. (1995): "The role of the textbook in EAP writing research", *TESOL Quarterly*, 14*,* 3-18.

Thonus, T. (2004): "What are the differences? Tutor interactions with first- and second-language writers", *Journal of Second Language Writing*, 13*,* 227–242.