Use of Signaling Nouns in Published and Korean Student Academic Writing in Applied Linguistics

Sun-Young Oh
(Seoul National University)


The present study investigates the use of signaling nouns (SNs) in published and Korean graduate student academic writing in applied linguistics. A set of 35 nouns was examined for their frequency as SNs, while the six most frequent SNs in published writing were subjected to detailed analyses of realization patterns. The results indicated that the nouns function as SNs in fifteen percent of the time of their total use and that students overall used a greater number of SNs than did published authors. Despite considerable overlap in the most frequent SNs, there was evidence that students rely on a narrower range of SNs than published writers. Published authors were differentiated from students in employing more anaphoric SNs, which contributed to cohesiveness and organization of text through the effective encapsulation of the preceding stretch of discourse. This study challenges the previous claim that the frequency of SNs positively correlates with the writing proficiency. Some pedagogical implications are drawn for academic writing instruction.

**Key words**: signaling nouns, academic writing, lexical cohesion, English for Academic Purposes (EAP)

1. INTRODUCTION

Signaling nouns (SNs hereafter) refer to abstract nouns which signal that their specific meaning may be found in the context of their use (Flowerdew, 2003). In other words, although these nouns have an invariable meaning of their own, their exact or full meaning needs to be made clear in a specific context. Typical examples include *change, difficulty, fact, issue, problem, process, reason*, and others. The signaling function of SNs may work either anaphorically or cataphorically across clause boundaries, or within the clause. The following examples illustrate each of these cases (the SNs are boldfaced and their specifications are underlined).
In these history textbooks, information is often presented as a list of events that assumes the reader can make the necessary connections between them, without elaboration through more detail, comment, or example. Eighth-grade students noted this difficulty during our interviews with them.

(2) This runs counter to the results of previous studies, almost all of which have reported that pre-task planning leads to more fluent language use.

In example (1) the noun difficulty refers back to the whole preceding sentence, whereas in example (2), the SN results refers forward to the underlined clause that follows. Unlike these cases where the SNs are realized in adjacent clauses, the next two examples contain SNs that are detailed within their own noun phrase or clause.

(3) Beatriz viewed reading as a process of knowing how to pronounce words, as confirmed during the think-aloud task in English.

(4) The problem is that in WC sessions, “modeling” can run very close to “telling.”

In examples (3) and (4), the full meanings of process and problem are made clear in the post-modifying of-clause and the complement that-clause, respectively (See Appendix A for other types of in-clause SN realizations).

SNs (or their sub-varieties) have been discussed under a number of different labels, including type 3 vocabulary (Winter, 1977), unspecific nouns (Winter 1982), metalanguage nouns (Winter, 1992), anaphoric nouns (Francis, 1986), carrier nouns (Ivanič, 1991), advance/retrospective labels (Francis, 1988, 1989, 1994), enumerative nouns (Hinkel, 2001; Tadros, 1994), and shell nouns (Aktas & Cortes, 2008; Hunston & Francis, 1999; Schmid, 2000). What these nouns have in common, irrespective of the different labels, is their essentially unspecific nature and their reliance on the context in which they occur for the specific meaning. Since the current study mainly adopts Flowerdew’s (2010) framework, his term signaling nouns will be used.

One main reason underlying many linguists’ interest in SNs is their role as cohesive devices. Given that creating cohesive (and coherent) texts is a necessary (though not sufficient) condition for successful written communication of any kind, linguistic items that can achieve cohesion have been a popular subject of study in ESL/EFL writing since the 1980s (Carrell, 1982; Connor, 1984; Hinkel, 2001; Scarcella, 1984). Despite substantial research on the topic of cohesion, however, SNs and their cohesive roles have not received

---

1 Flowerdew borrowed the term “signaling” from Hoey (1993) who argued that “(l)exical signals are the author’s/speaker’s explicit signaling of the intended organization,” which constitutes a primary means “whereby a reader/listener ‘decodes’ a discourse correctly” (p. 63).
much attention in the field of ESL/EFL until very recently (for a rare exception, see Francis, 1988). This is regrettable considering that SNs are very likely to present challenges for ESL/EFL learners. As some researchers have pointed out (Charles, 2003; Flowerdew, 2010), the use of SNs typically involves nominalizations and constitutes a form of grammatical metaphor (Halliday, 1994), which characterizes a higher level of discourse and is thus acquired late by children (Halliday & Matthiessen, 2004). Furthermore, SNs are cognitively complex in that they refer to abstract entities and their realization may not be straightforward to locate. The problems that SNs pose for non-native speakers have been attested to by some studies (Flowerdew, 2006, 2010; Francis, 1988).

Recent interest in SNs tends to reflect the importance of this class of nouns in academic discourse. It has been noted that SNs represent one of the characteristic features of written academic texts (Cullip, 2000; Flowerdew, 2003; Ivanič, 1991; Nation, 1990; Tadros, 1994; Thurstun & Candlin, 1998). This may be attributed, in large part, to their capacity to present (either new or previously stated) information which is elaborated in further detail (Tadros, 1994). They can “capture an event, a process, an action, a portion of on-going discourse” (Ivanič, 1991, p. 113) and are “particularly useful for learners entering the academic discourse community” (Ivanič, 1991, p. 96). Previous corpus-based studies have shown that SNs are indeed pervasive in academic research articles where they fulfill significant functions as cohesive devices (Aktas & Cortes, 2008; Flowerdew, 2003). In addition, some divergences have been found in the use of SNs (e.g., in terms of the frequency, range, and/or functions of SNs used) between published writers and student writers (Aktas & Cortes, 2008), or between native speakers and non-native learners (Flowerdew, 2006, 2010; Francis, 1988).

The present study attempts to contribute to the above line of research by investigating how SNs are used in the corpora of published research articles and Korean graduate students’ research papers in the academic discipline of applied linguistics. It is motivated by the realization that the information is as yet unavailable on the use of SNs by Korean students of English for Academic Purposes (EAP), specifically in comparison with published writers’ usage. EAP writing pedagogy would be enlightened by specifics about how the target learner group employs (and/or does not employ) SNs in their academic texts. The study focuses upon the overall frequencies of target SNs and the relative frequencies of different functions and realization patterns of the most frequent SNs, as well as their functions in discourse. Two corpora will be compared and contrasted with a view to revealing convergences and/or divergences between the two groups of writers. The present study is expected to shed new light on the use of SNs by published authors and Korean graduate students in the field of applied linguistics. By exploring how SNs function in published articles compared to student writing, this study will also provide information that
could help students write more cohesive and successful research papers. It is hoped that the findings of the present study will usefully inform EAP writing instruction and materials development.

2. LITERATURE REVIEW

2.1. Studies of Lexical Signaling Devices

Lexical items with signaling function have been studied under different terms for the past few decades. Halliday and Hasan (1976) is probably the first who paid special attention to what they call general nouns, or “a small set of nouns having generalized reference” (p. 274) which indicate humans, places, or facts (e.g., *people, thing, place, question*). They note that a general noun combined with a specific determiner (e.g., *the man*) is highly comparable to a pronoun (e.g., *he*), in that they both need some preceding clause to be fully interpreted and thereby serve cohesive functions. Unlike Halliday and Hasan (1976), Winter (1977) and Tadros (1985) emphasized the predictive, i.e., cataphoric, function of what they respectively called Vocabulary 3 and enumeration items, whose specific meanings are presented in the following linguistic context. It is Francis (1986) who achieved balance in focusing on the referential direction, maintaining interest both in the anaphoric and cataphoric functions of what she called anaphoric nouns. In later studies she employed separate terms for the two, i.e., advance and retrospective labels (Francis, 1989, 1994). These nouns, she argues, fulfill a key organizing function in discourse by encapsulating previous or ensuing stretches of text and incorporating it into a writer’s ongoing argument.

Whereas the researchers discussed above were concerned only with inter-clausal uses of lexical signals, Ivanič (1991) included intra-clausal relations in analyzing so-called carrier nouns, which can carry a specific meaning that they adopt in the particular context in addition to their dictionary meaning. She characterizes carrier nouns as abstract nouns which are common in the plural and can have a noun clause or nominalization as a complement. Below *purpose* is realized by its complement *to*-clause within the same clause:

(5) The **purpose** of the following section is to provide an **elementary account of the magnetic properties of ferrites**. (Ivanič, 1991, p. 101)

Ivanič’s insight that carrier nouns may derive meaning not only across clauses but also within clauses is important from a cognitive and pedagogical viewpoint (Flowerdew, 2003,
In processing such lexical signaling items in discourse, readers or listeners should be able to seek out their specific meaning within the clause where used as well as within (an) adjacent clause(s). Further refinement of the ways in which SNs are realized was carried out by Winter (1992), who pointed out that it may take a form of post-modification to the noun itself. Problem in extract (6), for example, is given a specific meaning by the post-modifying that-clause:

(6) The P.W.R. is the world-wide “standard” reactor. While this brings advantages on knowhow, it also brings the problem that the U.K. may be influenced by the effects of foreign experience, particularly safety standards. (Winter, 1992, p. 157)

Recent research in SNs has begun to exploit large electronic corpora using corpus linguistic tools and techniques. In his book on English abstract nouns acting as “conceptual shells,” Schmid (2000) operationally defined shell nouns as nouns which can be used in the following two types of constructions (p. 3):

(a) Determiner + (Premodifier) + Noun + postnominal that-clause, wh-clause, or to-infinitive
   (e.g., The (deplorable) fact that I have no money.)
(b) Determiner + (Premodifier) + Noun + be + complement that-clause, wh-clause, or to-infinitive
   (e.g., The (big) problem was that I had no money.)

From computerized searches of these grammatical patterns, Schmid has identified 670 shell nouns in the 225 million-word British section of the COBUILD corpus, discussing their various uses in great detail. He highlights the important yet often-neglected fact that what makes a noun a shell noun is not any of its inherent features but its use in actual texts. The functions which characterize uses of nouns as shell nouns are, according to him, the following three: “the semantic function of characterizing,” “the cognitive function of temporary concept-formation,” and “the textual function of linking” (p. 14, emphasis in original). The corpus linguistic methodology allowed this study to present a great amount of diverse examples of shell nouns and their frequency information. Automatic collection and processing of the corpus data, on the other hand, was only compatible with four lexico-grammatical patterns (especially intra-clause patterns), which resulted in relative neglect of inter-clause relations and total exclusion of other possible patterns, thereby producing an incomplete picture of shell noun usage.
2.2. Studies of Signaling Noun Use in Specialized Texts

In recent years, studies have investigated the use of SNs in specialized texts, specifically, learner writing and academic texts. Noting the potentially problematic nature of SNs for language learners, Flowerdew (2006) carried out a study of learners’ SN use with a 111,000 word corpus of argumentative essays written by Cantonese English learners. He found that the average number of SNs used per essay and the numbers of SN errors showed positive and negative correlations, respectively, with grades awarded to the essays. It was thus concluded that the effective use of SNs is “a developmental phenomenon which correlates with overall writing ability” (Flowerdew, 2006, p. 360). In a follow-up study, Flowerdew (2010) compared learners and native speakers using two comparable corpora. The focus of comparison was the overall frequency and variety of SNs, frequency of different lexical realization types (e.g., across-clause vs. in-clause, cataphoric vs. anaphoric), and frequency of different in-clause realization patterns. The results confirmed his hypothesis that L2 writers would employ SNs less frequently than native speakers. In particular, anaphoric across-clause SNs and in-clause SNs were twice more frequent in the L1 corpus than in the L2. While many of the most frequent SNs were found in common between both corpora, the range of different SN types was much greater in the L1 corpus. Based on these results, Flowerdew suggested that the frequency and range of use of SNs should increase in learner writing.

Two subsequent studies discussed the use of SNs in academic discourse. Flowerdew (2003) presented the frequency and range of SN types which he identified in two academic corpora of different modalities. Confirming their higher frequency in a formal written genre, SNs occurred much more frequently in textbooks than in college lectures. The range of SNs identified was wide, but many of them occurred with relatively low frequencies, stressing the importance of establishing a valid criterion for selection for pedagogy. While this study provides a useful framework for learners and teachers in EAP, its findings may not be generalized to all EAP instruction, given the relatively small size of the corpora (less than 100,000 words), and more crucially, the specific discipline (i.e., biology) targeted in the study.

With a particular interest in the cohesive role of SNs in academic writing, Aktas and Cortes (2008) analyzed the use of SNs (or shell nouns in their term) in research papers written by published authors and international graduate students from various disciplines. Adopting Schmid’s (2000) methodological framework, they comparatively examined the usage of 35 pre-selected SNs in two corpora. The results showed that overall SNs were more frequent in the student papers than in the published articles. This rather unexpected finding, however, was not considered significant by the researchers because some SNs (e.g., event, category) were frequent only in a few student papers. Instead, some differences in
the functional patterns of SNs were highlighted. The most prominent finding concerned the use of *fact:* in the published articles this SN tended to be followed by a complementing noun clause referring forward in the service of a characterizing function, whereas in the student writing it followed a demonstrative determiner *this* referring backward and serving a linking function. This study represents the first attempt to look into EAP students’ SN uses compared to those of published authors. It has, however, various limitations, some with serious consequences. First, all the 35 nouns were not examined to determine whether they were functioning as shell nouns. Given that potential SNs are not always used as true SNs, this is likely to have produced a distorted picture of overall SN frequencies. Second, the in-clause and within-clause functions of SNs were not differentiated, nor were the locations considered where SNs were realized in the text. Third, attempts were made to find associations between three functions of SNs proposed by Schmid (2000) and particular lexico-grammatical patterns used, but this effort was ill-advised in view of Schmid’s suggestion that all three of these functions are present in the use of any SN. Another drawback of this study was the small size of the student corpus, which was acknowledged as a possible reason for some of the unsatisfactory or unexpected results. Moreover, this small corpus consisted of papers from six different academic fields, some of which (e.g., economics) were represented only by two texts.

As seen above, earlier studies have significantly increased our knowledge of SN uses by ESL/EFL learners or novice writers. Previous research, however, needs to be complemented by studies which overcome their limitations and offer greater clarity. Given the prevalence and significance of SNs in academic writing, the present study explores how they are used in research papers by published writers and Korean graduate students. Recognizing that each academic discipline may have different writing practices or conventions, this study chose to target only one discipline, applied linguistics, instead of many. By collecting writings from students who share a first language and a field of study, the present investigation is expected to provide more useful and directly applicable pedagogical recommendations for this specific group of learners. The research questions that this study seeks to find answers to are as follows:

1. What are the overall frequencies of the target SNs in the corpora of published academic research articles and Korean graduate students’ research papers in applied linguistics?
2. What are the relative frequencies of across-functions (anaphoric and cataphoric) and in-clause functions of the most frequent SNs?
3. What are the relative frequencies of the different realization patterns for the in-clause function[s] of the most frequent SNs?
3. DATA AND METHOD

3.1. Data

Two corpora were compiled and used as data for this study. The Korean Graduate Student Article Corpus (KGSAC) consists of research papers written by Korean graduate students majoring in English education, a core subfield of applied linguistics. The papers were produced as a major requirement for the courses that the students took in the graduate program at a university in South Korea in the years between 2008 and 2011. They report empirical research that the students planned and conducted throughout the term of the course. The students may be assumed to have an advanced level of English proficiency, in that they all satisfied the requirement of a minimum score of 750 on the TEPS (Test of English Proficiency, developed by Seoul National University). Once all the necessary measures were taken (see below), the resulting student paper corpus consisted of 34 term papers, which amount to approximately 160,000 words.

The Applied Linguistics Published Article Corpus (ALPAC) is a collection of research articles published in international journals in the discipline of applied linguistics. Five academic journals were selected for this purpose, i.e., *Applied Linguistics, TESOL Quarterly, Journal of Second Language Writing, English for Specific Purposes*, and *Journal of English for Academic Purposes*. The selection was based on the results of an informal survey, which indicated that the target students consider these journals as leading academic journals in their field and rely on them as an important source for references. The journals were accessible online at the moment of building the corpus. A total of 70 articles totaling over 500,000 words were chosen for inclusion in the corpus by considering similarities in research topics with the student papers. The publication years ranged from 2001 to 2010.

The two corpora were comparable in terms of research methods and rhetorical structures of the papers included. As a means of counteracting the potential effects of individual writing practices, only one paper per writer was included in each corpus. The standard practice for building academic corpora was implemented, eliminating all the sections not belonging to the body of the papers or not produced by the writers themselves (e.g., tables, figures, references, appendices, or direct quotes). Table 1 shows basic information on the two corpora.

---

2 The TEPS grading system specifies that scores ranging from 701 to 800 are an indicator of an advanced level of communicative competence, 801 to 900, a near-native level, and 901 to 990, a native-level (http://www.teps.or.kr/Teps/info/teps_grade.aspx).
TABLE 1

<table>
<thead>
<tr>
<th>Corpus</th>
<th>Number of texts</th>
<th>Number of words</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALPAC</td>
<td>34</td>
<td>524,627</td>
</tr>
<tr>
<td>KGSAC</td>
<td>70</td>
<td>162,219</td>
</tr>
</tbody>
</table>

3.2. Analytical Procedures

The current study followed Aktas and Cortes (2008) in adopting a list of 35 enumerative nouns from Hinkel (2004) as the target SNs (see Table 2 for a complete list).\(^3\) Among the various existing lists of abstract nouns that are equivalent to (or are some sub-categories of) SNs (e.g., Francis et al., 1998; Schmid, 2000; Winter, 1977), Hinkel’s was chosen because she was specifically interested in academic writing. It should be borne in mind, however, that any such list cannot be considered complete because new SNs can be created through foreign borrowings, metaphors, or nominalization (Flowerdew, 2002, 2010; Francis, 1986).

Several steps were taken for the data analysis. First, all tokens of 35 nouns were retrieved from each corpora using a concordancing program (WordSmith, version 5.0). Non-nominal uses of some nouns (e.g., approach, result, process) were identified and eliminated manually. In the next step, each instance of all the nouns was examined in its context to decide whether it was used as a SN, and if so, its specific realization pattern was identified.\(^4\) Criteria for determining the SN status of a noun were developed through a close reading of the relevant literature, and revised and/or complemented as necessitated by the data during the analytical process. The non-SN uses were classifiable into one of the seven types: exophoric reference, homophoric reference, lexical equivalent, proper noun, constant meaning, pre-modification, and prefabrication (see Appendix B for details). For

---

\(^3\) Noting that “enumerative nouns are far more common in academic than other types of writing,” Hinkel (2004) provides a list of 34 “highly prevalent items” (p. 284). To these nouns Aktas and Cortes (2008) apparently added one more item, effect, although they did not explicitly mention this fact.

\(^4\) The process of identifying non-SN uses and categorizing SN tokens turned out to be a formidable task mainly due to the sheer number of occurrences of these nouns, which amounted to 11,685. Initially efforts were made to investigate all of these tokens one by one, but it soon became apparent that the job was logistically impractical or impossible. The revised analytic plan entailed randomly extracting and analyzing 100 tokens for those nouns whose frequency outnumbers one hundred, and then extending the results for the entire tokens. This method was only applied to the ALPAC, however, which was larger in size and yielded a greater number of tokens; all the instances from the KGSAC were analyzed without random sampling. A total of 5,169 occurrences of nouns were thus analyzed manually (2,897 from the ALPAC, and 2,262 from the KGSAC).
uses that were classified as SNs, each token was further coded for its realization patterns. They were first categorized as either across-clause or in-clause functions. SNs coded as the former type were further classified as either anaphoric or cataphoric references, and the latter as the four major sub-categories, i.e., complementing clauses (e.g., that-clause, wh-clause, to-clause), post-modifying clauses, post-modifying prepositional phrases, and others (e.g., apposition, deverbal nouns). (Each of these categories is illustrated in Appendix A.)

Once the coding of tokens was completed, the following phases of quantitative analysis were performed on the coded data. First, the frequencies of the SNs were calculated for each noun type in each corpus and normalized to 100,000 words for a valid comparison between the corpora. Based on the counts of non-SN uses, the proportion of SN uses to total noun tokens was also computed. Second, for the most frequently-occurring six SNs in the ALPAC, the frequencies of the major SN functions (i.e., across-clause vs. in-clause, with the former being divided into anaphoric and cataphoric uses) were calculated, and their relative percentages were also computed based on the total number of occurrences of the given noun. Third, normalized frequencies were calculated for the realization patterns of the in-clause function of these six SNs. The results of the above three steps were compared between the two corpora, the aim being to detect any similarities and/or differences between published authors and Korean graduate students in the use of SNs. Whenever necessary, chi-square tests were conducted to evaluate the statistical significance of frequency differences. The last stage in the data analysis involved scrutinizing SN uses in their full context in order to provide accounts for and/or to complement the primary results of the quantitative analyses.

4. RESULTS AND DISCUSSION

4.1. Overall Frequency of SNs

Table 2 presents the frequency of the 35 target SNs in each corpus on both raw and normalized counts. The frequency of all noun uses is also provided along with the proportion of SN uses to all noun uses. (The nouns are ordered according to frequency in the ALPAC.) Table 2 provides further information of interest. It indicates that the average percentage of SNs to all noun uses is similar between the two corpora (i.e., 15% in the ALPAC and 17% in the KGSAC). In other words, the target nouns function as SNs in about fifteen percent of the time of their entire use. Some of these nouns, however, including fact, reason, tendency, and trend, seem to be more dedicated to the role of SNs than others in that SN uses account for over half of their total occurrences.
Table 2 also shows that the total frequency of occurrences of SNs is somewhat lower in the ALPAC than in the KGSAC (276 vs. 300 on normalized counts). In other words, Korean graduate students used a greater number of SNs, when viewed in aggregate, than published authors. This difference in frequency was statistically significant ($\chi^2 = 80.882$, $df = 31$, $p < 0.01$). This finding corroborates that of Aktas and Cortes (2008), who compared the academic writing of published writers and international graduate students from various disciplines. In fact, the frequency difference was much greater in their study, but the
researchers attributed this to the repetitive use of the same shell nouns (e.g., event, category) by one or two student writers. Such highly uneven frequency distributions across writers, however, were not detected in the present study. The finding of this study, on the other hand, is not in line with Flowerdew’s two studies (2006, 2010), where Cantonese-speaking college students employed significantly fewer SNs in argumentative essays than their native speaker peers. The present result challenges Flowerdew’s argument that the frequency of SNs correlates positively with the writer’s general writing proficiency. It appears that graduate students, who have developed at least an implicit understanding of the prevalence and necessity of SNs in academic texts, attempt to employ them in their writing to an even greater degree than what is typical in their academic discourse community. Yet simple differences in total frequencies of SNs may matter less than discrepancies in their actual uses or realization patterns, a point raised by previous researchers as well.

In addition, the difference in overall frequency is not necessarily mirrored in the frequency of the individual SNs. A closer look at the table reveals that 18 SNs indeed occurred more frequently in the KGSAC, but 12 showed the opposite trend, with the remaining five occurring with the same frequencies in the two corpora. The results of the chi-square tests on the individual SNs indicated that approach and issue are significantly less frequent, and tendency, characteristics, feature, problem, and result are significantly more frequent, in the student writings than in the published writings (p < .01). The difference between the two corpora in typical uses of some of these nouns will be discussed in the next section.

One can also refer to Table 2 to compare the top 10 most common SNs in the two corpora. While six nouns (result, fact, process, reason, purpose, and effect) occur in the top 10 of both corpora, approach, issue, task, and aspect are unique to the ALPAC, and tendency, problem, characteristics, and feature, to the KGSAC. The top SN (i.e., result) remains the same across the corpora but it occurs much more frequently in the KGSAC than in the ALPAC (56 vs. 36). Compared to the ALPAC, on the other hand, the KGSAC has a greater number of SNs equal to or less than two in frequency (10 vs. 17). These facts may suggest that SNs are more evenly distributed in the ALPAC, which is comparable to Flowerdew’s (2006) finding that nonnative students show a more limited range of SN uses than their native peer writers.

The most frequent six SNs in the ALPAC (result, fact, process, approach, issue, and reason) were selected for examination and comparison between the two corpora in their detailed realization patterns. The criterion for selecting SNs for detailed analysis (i.e., top six in the published articles) was adopted from Flowerdew (2003) and Aktas and Cortes (2008), who regard published authors’ SN usage as the model for graduate students to emulate. A chi-square test performed on these six nouns pointed to a highly significant
difference between the two corpora ($\chi^2 = 23.531, df = 5, p < 0.001$).

### 4.2. Realization Patterns of the Selected SNs

Table 3 presents the (normalized) frequencies and the corresponding percentiles of the six selected SNs occurring in three different realization patterns, i.e., anaphoric, cataphoric, and in-clause functions. These data are also graphically represented in Figure 1. In the table, the anaphoric and cataphoric patterns are combined into a separate category (i.e., “across clauses”) in order to make two-way comparisons between across- and in-clause patterns straightforward.

<table>
<thead>
<tr>
<th>SN</th>
<th>Corpus</th>
<th>anaphoric</th>
<th>Cataphoric</th>
<th>Across clauses</th>
<th>In clauses</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>result</td>
<td>A</td>
<td>21.8</td>
<td>11.4</td>
<td>33.2</td>
<td>2.8</td>
<td>36.0</td>
</tr>
<tr>
<td></td>
<td>(60.5%)</td>
<td>(31.6%)</td>
<td>(92.1%)</td>
<td>(7.9%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>35.5</td>
<td>14.2</td>
<td>49.7</td>
<td>6.3</td>
<td>56.0</td>
</tr>
<tr>
<td></td>
<td>(63.4%)</td>
<td>(25.4%)</td>
<td>(88.7%)</td>
<td>(11.3%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>fact</td>
<td>A</td>
<td>0.8</td>
<td>0.0</td>
<td>0.8</td>
<td>26.2</td>
<td>27.0</td>
</tr>
<tr>
<td></td>
<td>(2.9%)</td>
<td>(0%)</td>
<td>(2.9%)</td>
<td>(97.1%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>3.9</td>
<td>0.0</td>
<td>3.9</td>
<td>33.1</td>
<td>37.0</td>
</tr>
<tr>
<td></td>
<td>(10.6%)</td>
<td>(0%)</td>
<td>(10.6%)</td>
<td>(89.4%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>process</td>
<td>A</td>
<td>8.0</td>
<td>0.0</td>
<td>8.0</td>
<td>17.0</td>
<td>25.0</td>
</tr>
<tr>
<td></td>
<td>(32.1%)</td>
<td>(0%)</td>
<td>(32.1%)</td>
<td>(67.9%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>4.7</td>
<td>0.0</td>
<td>4.7</td>
<td>10.3</td>
<td>15.0</td>
</tr>
<tr>
<td></td>
<td>(31.6%)</td>
<td>(0%)</td>
<td>(31.6%)</td>
<td>(68.4%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>approach</td>
<td>A</td>
<td>16.5</td>
<td>1.4</td>
<td>17.9</td>
<td>4.1</td>
<td>22.0</td>
</tr>
<tr>
<td></td>
<td>(75.0%)</td>
<td>(6.3%)</td>
<td>(81.3%)</td>
<td>(18.8%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>3.9</td>
<td>2.3</td>
<td>6.2</td>
<td>0.8</td>
<td>7.0</td>
</tr>
<tr>
<td></td>
<td>(55.6%)</td>
<td>(33.3%)</td>
<td>(88.9%)</td>
<td>(11.1%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>issue</td>
<td>A</td>
<td>8.5</td>
<td>1.5</td>
<td>10.0</td>
<td>7.7</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>(47.8%)</td>
<td>(8.7%)</td>
<td>(56.5%)</td>
<td>(43.5%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.0</td>
<td>1.7</td>
<td>1.7</td>
<td>4.3</td>
<td>6.0</td>
</tr>
<tr>
<td></td>
<td>(0%)</td>
<td>(28.6%)</td>
<td>(28.6%)</td>
<td>(71.4%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td>reason</td>
<td>A</td>
<td>7.0</td>
<td>7.6</td>
<td>14.6</td>
<td>4.4</td>
<td>19.0</td>
</tr>
<tr>
<td></td>
<td>(36.7%)</td>
<td>(40.0%)</td>
<td>(76.7%)</td>
<td>(23.3%)</td>
<td>(100%)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>4.6</td>
<td>5.4</td>
<td>10.0</td>
<td>7.0</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>(27.3%)</td>
<td>(31.8%)</td>
<td>(59.1%)</td>
<td>(40.9%)</td>
<td>(100%)</td>
<td></td>
</tr>
</tbody>
</table>
In comparing the proportion between across- and in-clause patterns, the SN that stands out the most is *fact*: in both corpora, this SN is predominantly realized within clauses (97.1% and 87.4% in published articles and student writings, respectively). *Process* shares this tendency, only to a lesser extent (67.9%, 68.4%). In contrast, *result*, *approach*, and *reason* are principally realized across clauses, listed in order of decreasing degree (92.1%, 88.7%; 81.3%, 88.9%; 76.7%, 59.1%). As can be seen in Table 3, there was not much difference in these trends between the two corpora. The only notable divergence concerns *issue*, which is an SN that published authors employ much more often (3.3 times) than Korean graduate students. In contrast to the published writers who used in-clause and across-clause patterns with similar percentages (43.5% vs. 56.5%), the students showed a stronger preference for the in-clause pattern with this noun (71.4%). The followings are typical examples, the first two taken from published articles, and the remaining two from student papers:

(7) Thus, it is possible that this group was weaker to start with as to grammatical accuracy. This is a key issue to keep in mind in interpreting the findings of this study.

(8) If, however, learners were able to access explicit language knowledge in this way, we would expect that the results of this test would not correlate with language measures that were “time-pressured” (i.e., likely measures of implicit language
knowledge). This **issue** will be dealt with below.

(9) The aim of the pretest was to examine the participants’ understanding of the definite article use, thus obviating any problem that can be caused by the **difficulty in understanding the test sentences** was an important **issue**.

(10) Accordingly, one of the crucial **issues** in L2 acquisition research has been whether L2 grammars are under control of the principles of UG.

Even when **issue** was realized across clauses in the students’ writing, it never referred to the preceding clause(s), whereas anaphoric reference was the most frequent use of **issue** by the expert writers (47.8%). It thus seems that students are less capable of transforming the content of the previous clause(s), which may otherwise not be taken by the reader as an important subject in its own right, into an “issue” that deserves further discussion or consideration in the subsequent stretch of discourse. Note that this is precisely what happens in examples (7) and (8) above. Such an ability to exploit SNs for anaphoric functions has been argued to be critical for academic writers for the following two reasons: the successful integration of ideas develops an argument that is better organized and thus more persuasive, and, furthermore, the selection of a specific noun allows the writer to take a stance as a competent member of the discipline (Charles, 2003).

A similar type of published authors’ preference for anaphoric use is also observed with **approach** (75%). Like **issue**, this SN is also (3.1 times) more frequent in the published articles than in the student papers. This can be seen in the next example, where a published writer mainly uses **approach** to encapsulate the preceding clause into a single nominal concept, which accords it the status of a pedagogical “approach”:

(11) While it could be argued that students should be given the chance to edit their own work without feedback so that they can develop independent self-editing skills, at this level of instruction such an **approach** might prove ineffective in helping students to reach accuracy levels that would enable them to pass the course and might thus be extremely frustrating to the students.

On the other hand, **result**, the most frequently used SN in both corpora, occurs in significantly higher frequencies (1.5 times) in the student writings. Typically, it is repeated many times within a short stretch of text where the student compares his/her research findings with those of previous studies. Furthermore, this noun often occurred within the prepositional phrase **according to** with its specific meaning provided in the rest of the sentence. Such a use, in contrast, was never found in the published articles.

The preceding discussion pointed to some differing distributions of **issue**, **approach**, and **result** across major realization patterns between the two corpora. This difference was
confirmed to be highly significant ($x^2 = 150.728$, $df = 22$, $p < 0.001$). An examination of a few other nouns also displayed disparate tendencies between the two corpora. Specifically, students tend to show a preference for certain realization patterns, whereas expert writers are more balanced in the use of diverse patterns. For example, the published authors used effect and purpose both for anaphoric and cataphoric functions in similar ratios, but the students only used the former cataphorically, and the latter, anaphorically.

4.3. In-Clause Patterns of the Selected SNs

The in-clause function of SNs may be divided into four sub-categories, which are complement clauses (i.e., that-clause, wh-clause, to-clause), post-modifying clauses, post-modifying prepositional phrases, and others (including apposition and deverbal nouns). Table 4 shows the frequency distribution of these realization patterns for the six SNs in the two corpora.

<table>
<thead>
<tr>
<th>SN</th>
<th>Corpus</th>
<th>Comp. Cls</th>
<th>Postmod. Cls</th>
<th>Postmod. PP</th>
<th>Other</th>
<th>SUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>results</td>
<td>A</td>
<td>0.0</td>
<td>0.9</td>
<td>0.0</td>
<td>1.9</td>
<td>2.8</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>1.6</td>
<td>0.8</td>
<td>0.0</td>
<td>3.9</td>
<td>6.3</td>
</tr>
<tr>
<td>fact</td>
<td>A</td>
<td>0.0</td>
<td>25.7</td>
<td>0.6</td>
<td>0.0</td>
<td>26.2</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.0</td>
<td>33.1</td>
<td>0.0</td>
<td>0.0</td>
<td>33.1</td>
</tr>
<tr>
<td>process</td>
<td>A</td>
<td>0.0</td>
<td>0.0</td>
<td>17.0</td>
<td>0.0</td>
<td>17.0</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.0</td>
<td>0.0</td>
<td>9.5</td>
<td>0.8</td>
<td>10.3</td>
</tr>
<tr>
<td>approach</td>
<td>A</td>
<td>1.4</td>
<td>0.0</td>
<td>1.4</td>
<td>1.4</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.8</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
</tr>
<tr>
<td>issue</td>
<td>A</td>
<td>0.8</td>
<td>0.0</td>
<td>5.4</td>
<td>1.5</td>
<td>7.7</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>1.7</td>
<td>0.0</td>
<td>0.9</td>
<td>1.7</td>
<td>4.3</td>
</tr>
<tr>
<td>reason</td>
<td>A</td>
<td>1.9</td>
<td>0.0</td>
<td>0.0</td>
<td>2.5</td>
<td>4.4</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>3.1</td>
<td>0.8</td>
<td>0.0</td>
<td>3.1</td>
<td>7.0</td>
</tr>
</tbody>
</table>

As noted in Section 5.2, fact is by far the most frequently realized within the same clause both in the student and published writings. Table 4 additionally indicates that this SN almost exclusively occurs in the pattern of the post-modifying that-clause, as illustrated by the following examples.

---

5 Aktas and Cortes (2008) reported that fact was used by the students in the pattern of “th- + N” (e.g., this fact...) for anaphoric reference. They consider this the most significant difference between student and published writers since the latter group typically used this noun in the pattern of “N + clause” (i.e., the fact that...). Students in the current study also used fact for anaphoric reference more frequently than published writers (10.6% vs. 2.9%), but in general terms, the two groups were not very different in how they use this SN with regard to its realization patterns.
(12) Secondly, the fact that L2 learners did correct 35 per cent of ungrammatical statements is evidence that this test is reconstructive. (ALPAC)

(13) The fact that learners occasionally drew wrong conclusions and struggled when they lack enough linguistic knowledge at their hand also implies that teachers’ optional intervention is necessary. (KGSAC)

Process is parallel to fact in that it frequently occurs in in-clause, especially post-modifying, patterns. Unlike fact, however, process was used only in post-modifying prepositional phrases, not clauses, as shown below:

(14) Candlin and Plum (1999) propose the term induction to describe the mediated but not unproblematic process of educating undergraduate students in the discursive and epistemological practices of the disciplines. (ALPAC)

(15) The authors hypothesized that L2 learners have full access to UG thus to both settings of the Article Choice Parameter, and that L2 learners may show fluctuation between two settings in the process of setting the correct parameter for the target language. (KGSAC)

The post-modifying prepositional phrases are also favored by published writers for the in-clause function of issue as in example (16), but not by the student writers who use this SN significantly less frequently.

(16) Closely linked to the issue of borrowing from sources and appropriation of discourses is the problem of “textual plagiarism” (Pecorari, 2003) in ESL student writing.

It may be concluded that in general, the two corpora do not show striking differences in the in-clause usage of the selected SNs. The chi-square test confirms that the difference is not statistically significant ($\chi^2 = 14.40, df = 16, p = 31.999$). It should be noted, though, that some student writers exhibited idiosyncratic uses of SNs with regard to realization patterns. Subject, for example, was realized in a complement to-clause, a pattern which never occurred in published articles or in the Bank of English corpus (Schmid, 2000) (e.g., it would be an interesting subject to further investigate...).

5. CONCLUSION AND IMPLICATIONS

The present study investigated the frequency and realization patterns of signaling nouns
The present study contributes to the study of SNs in at least two respects. The finding that roughly fifteen percent of potential SNs actually fulfilled SN functions (in academic prose) adds significantly to the existing knowledge on SNs. No research has thus far shown the ratio of SNs to all noun uses of potential SNs. This information has been made available in the current study through systematic analyses of a large amount of data based on explicit criteria. Establishing and presenting overt criteria for the identification of (non-) SN uses itself constitutes a significant step forward in this area of research since few previous studies have done so. Next, by challenging Flowerdew’s (2010) claim that “better writers use more SNs than do less competent ones” (p. 49), this study shows that a higher frequency of SN usage does not in itself lead to better texts. Although SN frequency and writing quality may show positive correlations up to a certain level of proficiency, as suggested by Flowerdew (2010), they seem to entail more complex relations for advanced-level learners as shown in the present study.

What do the findings of the present study imply for the pedagogy of academic writing? First, this research confirms the pervasiveness of SNs in academic discourse and their critical role in creating cohesion, which ultimately contributes to the coherence of a text. While the students appear to have a decent grasp of the basic functions of SNs, they need
to be made aware of and be provided with opportunities to exploit various and typical patterns in which SNs are realized to signal discourse relations in academic texts. Flowerdew’s (2010) recommendation that non-native learners should increase their frequency of SN use, however, is not suitable for this particular group of students. What they need instead appears to be fine-tuning of the ways that they utilize SNs if they are to better approximate expert usage.

While valuing native texts as models that non-native learners are to imitate is sometimes criticized from the perspective of English as a lingua franca, it cannot be denied that following the discursive patterns of published articles can help graduate students who aim to have their research published in international journals to achieve this goal. Given the present finding that a higher frequency does not equal more effective use of SNs, more consideration should be given to improving the use of SNs by function. As seen with issue or approach, students were less successful in exploiting “the function of anaphoric signals of establishing the meaning of a stretch of discourse as given and at the same time providing the starting point for new information” (Flowerdew, 2003, p. 334). Another divergence between student writers and published authors concerned a severely biased preference for one referential direction (i.e., either forward or backward) with some across-clause SNs (e.g., effect, purpose). Such findings identify needs for instructional attention in teaching of SNs to target learners.

Instruction for SNs should always be situated in context so that their important discourse roles could be highlighted. While either an inductive or a deductive approach may be chosen, depending on the learners’ preference or learning style, the former type appears to be fitting (and effective) for advanced-level students. For this purpose, activities used by Francis (1988) can be adopted with some modifications. For example, students are given a sample research paper from their discipline and asked to identify common SNs and how they are realized. Or some target SNs may be deleted and students may be asked to fill in an appropriate SN, with or without alternatives provided. They may discuss with the teachers the consequences of choosing one from among various potential SNs as well as their functions in the text. Students may also attempt to follow a given stretch of discourse with a sentence containing a suitable anaphoric across-clause SN. Such inductive approaches are compatible with the DDL (Data-Driven Learning, Johns, 1991) methodology which exploits corpus-linguistic techniques. Students would then be instructed to explore various forms and functions of SNs by themselves using a concordancing program, thereby learning about the typical associations of SNs and their realization patterns, for example. In environments with no suitable computer facilities, printed-out versions of concordance lists which the teacher generated and edited according to the particular pedagogical purpose may be used instead. Similar instructional activities have also been suggested by Flowerdew (2003).
Despite the theoretical and pedagogical contributions that it makes to the field, the present study is not without limitations, which point to areas for future studies. First, the number of SNs examined here is limited. If all SNs were identified in comparable corpora of published articles and student papers, selection of specific SNs and the range of different SN types could be compared in a more systematic manner. Second, frequencies of SNs in individual texts were not taken into account in this study. Considering text-based frequencies as well as overall frequencies may reveal some internal variations that were not observed here. Since the present study focused specifically on one academic discipline, i.e., applied linguistics, it would also be of interest to investigate the use of SNs in different fields of study and to compare the results across disciplines. Studies dealing with these issues will add to our knowledge of SNs in academic writing and provide further help for academically-bound L2 learners.

REFERENCES

167 Use of Signaling Nouns in Published and Korean Student Academic Writing in Applied Linguistics

Hong Kong: Institute of Language in Education.


APPENDIX A

Examples of Various Realization Patterns of SNs

I. Across Clauses

(1) Anaphoric

*While it could be argued that students should be given the chance to edit their own work without feedback so that they can develop independent self-editing skills, at this level of instruction such an approach might prove ineffective in helping students to reach accuracy levels that would enable them to pass the course and might thus be extremely frustrating to the students.*

(2) Cataphoric

*One might hypothesize that more formal lectures would involve a degree of advance editing and so would have more complex edge effects than the informal lectures. The results, however, were inconsistent; the hypothesis was true for the Race (formal) and Strategy (informal) lectures, but not the EMU (formal) lecture.*

II. Within Clauses

(1) Complementing Clauses

a) *that*-clause

*One reason for the heavy use of directives in the hard disciplines may be that this is one of the few rhetorical devices that scientists use with any regularity to explicitly engage*
(1) Post-modifying Clauses
We have seen that despite the best intentions of these teachers and their desire to respond positively and effectively, the effect of their mitigation was often to make the meaning unclear to the students, sometimes creating confusion and misunderstandings.

(2) Post-modifying Prepositional Phrases
Another factor that may prompt discipline problems is how students perceive activities or tasks.

(3) Post-modifying Prepositional Phrases
What may further limit the value of a corpus-based study is the difficulty of finding a corpus that is truly representative of the language use that is the focus of an investigation.

(4) Other
a) Deverbal noun
I wrote the course with a focus on the learners I was working with, one aspect of which was the incorporation of their L1, and once I had the course in a workable form I wanted to learn what learners thought of it.

b) Apposition
All three corpora are made up of contemporary, everyday, semiformal spoken American English, an important characteristic given that idioms are one of the most time-sensitive aspects of language.

---

6 Deverbal nouns, which refer to nouns derived from verbs, are listed in Ivanič (1991) as a structure in which SNs can be lexically realized within the clause.
APPENDIX B
Types of Non-SN Uses of Potential SNs

(1) Exophoric Reference

The full meaning of a noun can only be located in general background knowledge, not in the text. Below the writer does not mention what exactly the issues are, nor does it matter much in this context.

*The concepts of culture, perspectivity, and positionality ran through Pam’s course. She wanted to present issues from multiple perspectives and through different kinds of texts and media.*

(2) Homophoric Reference

The referent is locatable based on “a set of current or specific circumstances with which readers or listeners are familiar” (Flowerdew, 2002, p. 151). Consider results in the following sentence, which begins the conclusion section of an article.

*The results of this study suggest that affective factors, including readers’ views of their home language and beliefs about reading, may play an important role in reading.*

Having read the previous sections where the research results are presented and discussed, readers will have little problem in inferring what results mean in this context without any linguistic specifications. Both exophoric and homophoric references are excluded from the scope of SNs in most previous studies, which restrict the contexts for realization of SNs to linguistic ones (see Ivanič (1991) for one exception).

(3) Lexical Equivalent

The referent of a potential SN is a single noun phrase, as in the case of issue referring to plagiarism in the next example.

*Admonishments about plagiarism are common in style guides and academic handbooks, and most universities tend to treat the issue as a punishable offence.*

Following Francis (1988), such examples were considered “lexical equivalents” and not counted as SNs.
(4) Proper Noun

Some potential SNs occur in “a proper noun phrase referring to a term common in the literature” (Aktas & Cortes, 2008, p. 8), which is the case with approaches below.

New methodologies arising at the time, such as the Natural and Audio-Lingual **Approaches**, continued to centre on teaching the spoken language and to exclude the **L1**.

Following Aktas and Cortes, these uses were not taken into account as SN uses.

(5) Constant Meaning

Potential SNs may sometimes not have a specific meaning that depends on the context of use. For instance, **difficulty** below does not have a contextualized meaning beyond its dictionary meaning, i.e., “the fact of not being easy, or of being hard to do or understand” (*Cambridge American English Dictionary*, 2011).

**Causative verbs, which play an important rhetorical role in certain types of problem-solution texts, have been shown to be a major area of difficulty for students in several aspects.**

(6) Pre-modification

The noun phrase preceding a potential SN as a modifier may specify its full meaning. The NP **data collection** in the next example details the noun **process** in the form of a pre-modification.

**Thus, 72 ESL students and 8 teachers were involved in the data collection process.**

There has been little discussion of pre-modification in the literature on SNs. After careful consideration and discussion with Flowerdew (p.c., July 6, 2012), they were removed from the SN uses on the grounds that the nouns in this environment do not function in the same way as the typical SNs do.

(7) Prefabrications

Common frozen phrases which are formed on the basis of potential SNs were removed from the SN category because the nouns have become an essential part of the phrases and thus do not seem to operate individually as SNs. Some frequent examples include **have an effect on**, **have a problem with**, **have experience ...ing**, etc.
Applicable levels: Tertiary

Sun-Young Oh
Department of English Education
Seoul National University
1 Gwanak-ro, Gwanak-gu
Seoul 151-742, Korea
Phone: 02-880-7675
Fax: 02-880-7671
Email: sunoh@snu.ac.kr

Received in December 2013
Reviewed in January 2014
Revised version received in February 2014