Facts and things: Advanced ESL learners' use of discourse-organising nouns

Marie Kristin Tåqvist, Karlstad University

Abstract

This study examines the use of discourse-organising nouns (DONs), such as *fact*, *issue*, and *thing*, in a corpus of Swedish advanced students' academic writing in second language (L2) English, and in what ways texts produced by the L2 students resemble or differ from those produced by advanced native-speaker students and from expert writing in this respect. The L2 student writing was found to approximate the L1 student writing and the expert writing in several ways, including overall frequency of DON usage, but with less variety and more frequent occurrences of semantically vague types and types expressing attitude and involvement.

Keywords: discourse-organising nouns, cohesion, register, academic writing, L2 learners, ICLE, LOCNESS, ICE

1. Introduction and background

The importance of discourse features in ESL research and writing classrooms is well attested. A number of scholars have found that discourse features—or cohesion and coherence—have an important role to play, particularly for advanced students (e.g., Jafarpur 1991; Chiang 1999, 2003). One aspect of cohesion that is particularly pertinent in relation to academic writing is what will here be referred to as *discourse-organising nouns* (DONs).[1] DONs are semantically unspecific, abstract nouns, such as *argument*, *fact*, *issue*, *problem*, and *thing*, which rely on their linguistic co-text for part of their meaning and which have an organisational function in the discourse (Tåqvist 2016: 1). Overlapping concepts in the literature include Halliday and Hasan's (1976) *general nouns*, Winter's (1977, 1978) *type 3 vocabulary*, Francis' *anaphoric nouns* and *labels* (1986, 1994), Schmid's (2000, 2007) *shell nouns*, and J.

---

[1] The term *discourse-organising noun* is based on similar terms and concepts employed by other scholars, most notably L. Flowerdew (2001) and McCarthy (1991).

Flowerdew's (e.g., 2003a, 2003b, 2006, 2010, 2015; J. Flowerdew & Forest 2015) *signalling nouns*.

DONs can be further described as nouns which refer anaphorically (that is, they refer back to the preceding text) or cataphorically (that is, they refer to what follows) to the linguistic co-text, either within one and the same sentence or across sentence boundaries. In the context of the present study, the term discourse should therefore be understood as including, but not being restricted to, the intersentential level. The four types of co-reference relationship are illustrated in examples (1)–(4). Example (1) illustrates within-sentence anaphoric reference in a piece of academic writing in L2 English, and example (2) illustrates across-sentence anaphoric reference. Examples (3) and (4) illustrate within-sentence cataphoric reference and across-sentence cataphoric reference, respectively.[2]

1.   I like my subject and that is probably the most important **thing** when choosing a topic; it has to be a subject which interests you. (ICLE-SW: SWUL8027)

2.   Female authors often had to use male pseudonyms in order to have their newspaper articles or novels published. But that was a minor **problem** since most women did not have an opportunity to writ at all. (ICLE-SW: SWUL7025)

3.   Admittedly, there is a risk that our modern society becomes too concentrated on the **fact** that we should be efficient and deal with our duties at a rapid pace. (ICLE-SW: SWUL3013)

4.   Let me briefly point out some **arguments** that speak in favour of the car, and some that don't. To begin with the car enables us to move about more freely and to experience places and wonders which would otherwise have been impossible. […] (ICLE-SW: SWUL7031)

In the examples cited here, the DONs refer to information given in the preceding or following discourse; thereby, they function as signposts or signals, directing the reader and signalling how different parts of the text cohere and are to be understood. Discourse-organising nouns can thus be said to function as cohesive devices contributing to the organisation of discourse, and, in the words of J. Flowerdew and Forest (2015: 48), 'establish[ing] and maintain[ing] continuity in text.'

---

[2] The examples are from the Swedish subsection of the International Corpus of Learner English, one of the three corpora on which this study is based (see Section 2). In all examples, DONs are marked in boldface and their referents are underscored. No amendments have been made to spelling or grammar.

It has been frequently pointed out that such nouns play an important role in the creation of cohesion in academic text (e.g., Hoey 1983; J. Flowerdew 2003a, 2003b, 2006; J. Flowerdew & Forest 2015). J. Flowerdew (2006: 346) notes that such text-forming nouns (*signalling nouns*, in his terminology) are 'particularly frequent in academic language,' and that their use can therefore be seen as 'an important dimension in the development of academic literacy.' Previous research on these related concepts has focused on aspects of usage in both native-speaker and learner writing in various modes and registers. My doctoral thesis (Tåqvist 2016) explored the use of DONs in two student corpora (L2 and L1 English) and one corpus of expert writing. The study investigated a number of aspects of DON usage, including which DONs are used by the different writer groups, how often they are used, the reference patterns in which they are used, and the pre-modifiers with which they are used. The present article draws on this previous study, with a focus on overall frequency and function, variety of usage, and what kinds of DONs are overrepresented in one of the corpora in comparison with the other two. Thus, the overall aim of the study is to find out how DONs are used in Swedish advanced students' academic writing in L2 English, and in what ways the L2 students' writing is similar to or differs from that of L1 students and from expert writers in this respect.

2. Material and methods

This is a trilateral study, investigating DON usage in academic writing produced by three different writer groups—two corpora of student writing (L2 and L1 English) and one corpus of expert academic writing. This material is described below, starting with the student material.

The two student corpora used are, first, the Swedish subsection of the International Corpus of English (ICLE-SW), and second, the Louvain Corpus of Native English Essays (LOCNESS). The two corpora were compiled by a team led by Sylviane Granger at the Université Catholique de Louvain in the early 1990s and are often used together in comparative studies on learner and native-speaker language (e.g., Aijmer 2002, 2005; Ädel 2003, 2006; Boström Aronsson 2005; Mondor 2008; Tåqvist 2016). The two student corpora can be said to represent advanced-level student writing on the basis of external criteria: the students are undergraduate university (or A-level) students with their major (or minor) in English. The

L2 student writers can also be defined as advanced-level students on the basis of internal criteria: a random sample of 20 essays from the ICLE-SW subcorpus were rated by a professional rater applying the *Common European Framework* descriptors for writing, with the result that all sample essays were found to meet the criteria for advanced-level writing.

The corpus of expert writing consists of the academic or 'learned' subsections of three national subcorpora belonging to the International Corpus of English (ICE), compiled as part of a project led by Sidney Greenbaum. Specifically, the national subcorpora used in the present study are ICE-Ireland, ICE-New Zealand, and ICE-USA. These subcorpora were chosen, first, to represent a geographical spread, and second, to represent countries in which English is a first language. The academic subsections of ICE consist of peer-reviewed, published research from the humanities, the social sciences, the natural sciences, and technology, in equal proportions. Each text in the expert corpus is therefore on a separate topic. The reason for using scientific research articles as expert corpus in the present study over against, say, a corpus of press editorials, was that the former may be said to approximate the target register in the academic writing classroom more closely than the latter and therefore to be more pedagogically useful (but see, e.g., Rørvik 2013 for a different approach).

There are some key differences in the design of the three corpora to do with task and production circumstances. Specifically, the two student corpora consist of essays on set topics, and the majority of L2 student essays were written as timed essays with no access to secondary sources. In contrast, each text in the expert corpus is a published paper on a separate topic. These differences in task and production circumstances likely affected the outcome of the study to some extent (see further Section 4). Furthermore, the three corpora differ in terms of size, length of texts, and number of texts, with the L2 student corpus containing the shortest texts and the expert corpus containing the longest. An overview is presented in Table 1.

Table 1. Summary overview of the three corpora

	L2 students (ICLE-SW)	L1 students (LOCNESS)	Expert writing (ICE)
No. of words	200,033	324,304	253,733
No. of texts	355	436	120
Average text length	563	743	2,114

The difference across corpora in terms of the length and number of texts means that there are fewer texts, and also fewer writers represented, in the expert material as compared to the L2 and L1 student material. There is also a difference across the corpora in terms of overall size. Such differences are hard to avoid when comparing pre-existing corpus material.[3] However, it does mean that compensatory measures have to be taken. In order to compensate for the difference in corpus size, frequencies were normalised to a basis of 200,000 words (roughly the size of the smallest corpus), though in the presentation of results both raw and and normalised frequencies will be reported.

The analyses carried out are both qualitative and quantitative, the latter including both descriptive and inferential statistical analyses. The descriptive analyses include frequency counts and percentages. The inferential statistical analyses used are significance tests (Chi-square). In this study, a *p*-value of below .05 is considered significant. In addition, effect size is measured using Cramér's V test. Throughout the study, many significant differences were found across the corpora, but effect sizes were found to be consistently small by Cohen's conventions for interpreting effect size (see Aron, Aron, & Coups 2005: 192–193). This is to be expected in an investigation of a relatively infrequent phenomenon in a large data set (DONs accounting for approximately 1% of all words in the corpus material), and the results reported here should be understood in this light.

As for the DONs included in the study, a selection was made of 93 nouns with the potential to function as discourse organisers in text. These lexical items were selected in a two-step process involving (a) automated corpus retrieval based on structural criteria, and (b) manual retrieval based on functional equivalency (for a detailed account, see Tåqvist 2016: 63–69). In step 1, corpus linguistic software was used to identify and automatically retrieve nouns occurring in a set of predetermined patterns, previously shown by Schmid (2000) to be frequently used with shell nouns (in his terminology). The material used in this step was the academic subsection of the Corpus of Contemporary American English, or COCA

---

[3] ICLE-SW and LOCNESS are frequently used together in comparative studies on learner language, despite their difference in size, but the expert corpus used here was compiled anew. As the academic subsection of each regional subcorpus contains 40 texts of about 80,000 words each, three such regional subcorpora were deemed sufficient for inclusion here.

(Davies, 2008). In step 2, a manual text analysis was carried out to identify nouns with the potential to function as discourse organisers, on the basis of the definition of DONs used in this study (see Section 1). The text analysis was performed on a random selection of 20 full-length texts from the L2 student material and 20 from the L1 student material. The selected items (see Appendix) were regarded as potential DONs. With the help of the Concord tool in the software programme WordSmith Tools (Scott 2012), the corpus material was then searched for occurrences (tokens) of these potential DONs (types). Each token was evaluated in context and classified as either a DON or a non-DON. The difference is illustrated in examples (5) and (6).

5. My **point** is that in order to be able to keep or improve her (social) position without acting immorally an Austen heroine has to be very sensitive to social nuances, a sensitivity bordering on (and in one or two cases exceeding) the limit of snobbery. (ICLE-SW: SWUL8052)

6. I would like to quote Thoreau at this **point** and say that Although they make money and succeed in making a position for themselves, they are not happy […]. (ICLE-SW: SWUL4023)

The noun *point* in example (5) meets the criteria for DONs, with part of its semantic content located in the following discourse (underlined). In contrast, the same noun in example (6), although it can be said to contribute to textual organisation and cohesion, does not have a referent in the linguistic co-text. Therefore, this and similar instances were classified as non-DONs and excluded from the study. In the discussion below, the term *DON* is used to refer both to potential DONs (i.e., types), and to instantiated DONs (i.e., tokens).

3. Results
This section presents results of the quantitative and qualitative analyses, with a focus on frequency and variety of use (in Section 3.1) and what kinds of DONs can be said to characterise each of the three corpora (in Section 3.2). A discussion follows in Section 4.

3.1 Frequency and variety of use

As stated, the corpora were searched for all occurrences of each search word. Table 2 presents the total number of DONs (i.e., tokens) and the number of different DONs (i.e., types) in the three corpora.

Table 2. Frequency of DONs in the three corpora: types and tokens

	L2 students (ICLE-SW)	L1 students (LOCNESS)	Expert writing (ICE)
Absolute freq (tokens)	1,994	2,770	2,345
Freq per 200,000 words	1,994	1,709	1,848
No. of different DONs (types)	77	86	91

As shown in Table 2, the L2 students use DONs slightly more frequently (in normalised frequencies) than either of the other writer groups. However, although the difference across the corpora was found to be statistically significant (χ^2 28.41, df 2, p.000), it was so small as to be almost negligible (φ_c .006).[4] Thus, the two groups of student writers were found to approximate the corpus of expert writing in terms of overall frequency of DON usage, with the phenomenon of DONs accounting for approximately 1% of all words in all three corpora.

These figures are particularly noteworthy in the light of previous research on related concepts. Specifically, J. Flowerdew (2010) investigated an L2-writer corpus of argumentative texts by undergraduate students with Cantonese as their L1 and found it to be characterised by *less* frequent use of so-called signalling nouns (in his terminology) in comparison with native-speaker writing. Furthermore, in a later study, J. Flowerdew and Forest (2015) found that signalling nouns are more frequent in formal academic genres than in less formal ones. In view of the fact that L2 student writing—even at fairly advanced levels—can be characterised as informal and spoken-like (e.g., Ädel 2003, 2006; Aijmer 2001, 2002; Altenberg 1997; Altenberg & Tapper 1998; Gilquin & Paquot 2007, 2008; Granger 1998, 2007; Lorenz 1999), the similar frequencies of DON usage in the three corpora of the present study were somewhat unexpected.

There are also notable differences across the corpora, and one such has to do with variety of usage. As shown in Table 2, there are fewer types in

---

[4] The Chi-square test was run on the absolute number of DONs (tokens) in proportion to the total number of words in each corpus.

the L2 student writing than in the other writer groups. Out of the 93 search words included in the study, 16 were never used as DONs in the L2 student material. The corresponding figure in the L1 student corpus and the corpus of expert writing is seven and two, respectively. This difference is statistically significant (χ^2 13.27, df 2, *p* .001, φ_c .218), with the skew of the distribution most pronounced between the L2 student writers and the expert writers. This finding is particularly noteworthy in light of the fact that the list of search words included in the study was largely drawn from the student material. It can be concluded, then, that the L2 student writing stands out as being the least varied, in terms of DON usage, of the three corpora investigated. This finding was expected, in the light of previous research which has also found the vocabulary production of L2 students to be less varied than that of L1 students (J. Flowerdew 2010; see also Hasselgren 1994).

Another way of looking at variety of usage has to do with the proportion of high-frequency DONs in each corpus. In this study, the L2 student writing was found to contain by far the most tokens of high-frequency DONs, where high-frequency DONs are defined as the 20 most frequent types in each corpus. The figures are displayed in Table 3.

Table 3. Frequency of the 20 most frequent DONs in all three corpora

	L2 students (ICLE-SW)	L1 students (LOCNESS)	Expert writing (ICE)
Abs freq	1,502	1,997	1,351
Freq per 200,000	1,502	1,229	1,064

A Chi-square test on the absolute frequency of the top-20 DONs in proportion to the absolute frequency of all DONs in each corpus revealed a statistically significant difference across the corpora (χ^2 187.34, df 2, *p* .000, φ_c .162), with the skew of the distribution most pronounced between the L2 student writing and the expert writing. In terms of percentages, 75% of all instances of DONs in the L2 student material are among the 20 most frequent types, compared to 72% in the L1 student writing and 58% in the corpus of expert writing. Thus, the L2 students stand out in that they rely on a small number of types to a greater extent than is the case in the other writer groups, particularly the expert writers (see Hasselgren 1994 for a similar finding). These results are illustrated in Figure 1.
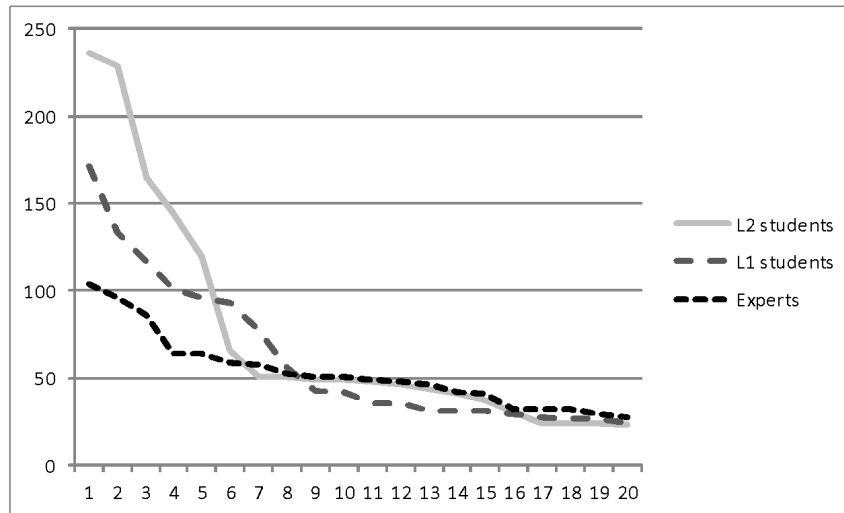
Figure 1. Frequency per 200,000 words of the 20 most frequent DONs

Another perspective on the matter of high-frequency DONs has to do with *which* types are high-frequency types. Table 4 shows the 20 most frequent types in each corpus.[5]

As is shown in Table 4, there is a notable degree of overlap across the corpora in terms of which types are high-frequency types. One striking similarity is the high frequency of *problem* throughout the material: it is the second most frequent DON in the L2 student writing and the most frequent DON in the L1 student writing and the expert writing. Thus, this is the most frequent DON overall, not just in the L2 student writing (see, e.g., J. Flowerdew 2010 for a similar finding). Other DONs occurring in all three top-20 frequency lists are *fact, factor, idea, issue, need, question, reason,* and *result.* That is, a total of nine DONs are common to all three lists. But there are also differences across the corpora. A total of four DONs (*opinion, situation, solution,* and *thing*) are among the top-20 DONs in the two student corpora but not in the expert corpus; three DONs (*argument, concept,* and *view*) are among the top-20 DONs in the L1 student writing and the expert writing; and three DONs (*approach, change,* and *possibility*) are among the top-20 DONs in the L2 student

---

[5] Bold typeface indicates occurrence in all three lists, italics indicates occurrence in two lists, and unmarked typeface indicates occurrence in one list only.

writing and the expert writing. In addition, four DONs (*dream*, *feeling*, *matter*, and *task*) are unique to the L2 students; four DONs (*point*, *statement*, *theme*, and *theory*) are unique to the L1 students; and five DONs (*attempt*, *finding*, *method*, *process*, and *sense*) are unique to the expert writers. These differences will be further explored in Section 3.2.

Table 4. The 20 most frequent DONs in all three corpora

	L2 students (ICLE-SW)		L1 students (LOCNESS)		Expert writing (ICE)	
	DON	Freq per 200,000	DON	Freq per 200,000	DON	Freq per 200,000
#1	*thing*	236	**problem**	171	**problem**	104
#2	**problem**	229	**fact**	133	**result**	96
#3	**fact**	165	**idea**	117	process	86
#4	**question**	145	**issue**	101	*approach*	64
#5	**reason**	120	*argument*	96	**factor**	64
#6	*opinion*	66	**reason**	93	method	59
#7	**issue**	51	**question**	77	**issue**	58
#8	*situation*	51	*thing*	56	**fact**	52
#9	**idea**	49	**result**	43	*concept*	51
#10	*possibility*	49	point	42	**need**	51
#11	**need**	48	*situation*	36	**reason**	49
#12	**result**	46	**need**	36	**question**	48
#13	*solution*	44	**factor**	31	*change*	46
#14	*change*	41	*opinion*	31	sense	42
#15	matter	37	*view*	31	attempt	41
#16	*approach*	30	solution	29	*argument*	32
#17	dream	24	theme	28	*possibility*	32
#18	**factor**	24	statement	27	*view*	32
#19	task	24	theory	27	finding	29
#20	feeling	23	*concept*	24	**idea**	28
	TOTAL	1,502	TOTAL	1,229	TOTAL	1,064

3.2 Overrepresentation of DONs in the three corpora

This section gives an account of what DONs – and what kinds of DONs – are overrepresented in each corpus, in comparison with the other two, and attempts to state the profile of each corpus on the basis of these findings. The Chi-square and Cramér's V tests were carried out on the absolute frequency with which each type is instantiated in proportion to the total number of tokens in each corpus (see Appendix). The purpose of the tests was to find out which DONs, if any, are significantly more frequent in one

corpus than in the other two. The following subsections present the findings for each corpus, starting with the L2 student writing.

3.2.1 Overrepresentation of DONs in the L2 students' writing
In the L2 student corpus, a total of fourteen types were found to be significantly more frequent than in the other two, namely *answer*, *dream*, *fact*, *matter*, *necessity*, *opinion*, *possibility*, *problem*, *question*, *reason*, *solution*, *thing*, *truth*, and *wish*. An additional thirteen DONs were found to be more frequent in the L2 student corpus but for these DONs, the cut-off point for statistical significance was not reached. These DONs are *conclusion*, *conviction*, *danger*, *doubt*, *fear*, *hope*, *impression*, *meaning*, *message*, *reality*, *situation*, *suggestion*, and *topic*.

On the basis of these results, a number of observations were made about DON usage in L2 student writing. First, L2 student writing is characterised by a high proportion of DONs expressing attitude and involvement. Specifically, the DONs *danger*, *doubt*, *dream*, *fear*, *hope*, *necessity*, *opinion*, *question*, and *wish* have such a function. Although not all of these are significantly more frequent in the L2 student corpus, the aggregated effect is that of writing which can be characterised as emotive and personal. Example (7) illustrates such usage in the L2 student corpus.

7.   My **hope** is of course <u>that the future development of Eastern Europe will be peaceful and that the thaw between the Superpowers will continue</u>. (ICLE-SW: SWUL7007)

The choice of DON in this example creates an impression of writing which is involved, as opposed to detached, an impression which is further strengthened by the use of other linguistic features which are atypical of formal academic writing, specifically, the first-person possessive determiner [*m*]*y* but also the modal auxiliary *will*. The frequent use of emotive and attitudinal DONs, and the use of them in close proximity to interpersonal markers and involvement features, is thus characteristic of the L2 students' writing.

Second, the L2 student writing is also characterised by a relatively high proportion of instances of the DONs *answer* and *question*, often used in close succession or in close proximity to actual direct or indirect questions posed by the writer. Such usage is illustrated in example (8).

8.   Being all alone at Christmas is not generally accepted in our society. The **question** is: how can we improve the situation? I believe that there is one **answer** to that **question**. Namely that we need to focus on the reason why we celebrate Christmas. (ICLE-SW: SWUG2065)

As illustrated, the L2 student writing is characterised by frequent use of the *question–answer* pairing, in which the writer poses a question and provides the answer in a mock dialogue with the reader, often explicitly labelling both question and answer as such throughout the sequence. As in example (7), the use of these DONs is coupled with first-person references. These results are in line with observations previously made by Ädel (2003), using the same student data as in the present study. Ädel found questions functioning as 'markers of a dialogic style' to be more frequent in L2 than L1 student writing (2003: 175). Although there are differences between Ädel's study and the present one – most importantly, Ädel investigated actual direct questions as represented by the number of question marks whereas the present study investigates particular uses of the lexical items *question* and *answer* – the similar findings are striking. Another central finding in this study is that the *question–answer* pairing is less frequent in the L1 student writing and in the expert writing: there, *answer* is more often used synonymously with *solution* and is used in connection with a discussion of some problem. Such usage is illustrated in examples (9) and (10). The examples are from the L1 student material and the expert material, respectively.

9.   Since there is so much crime, especially murder, in our world today, something needs to be done about it. Education may be the best **answer** to our problem, but some people feel that the death penalty can stop increased crime. (LOCNESS: USARG)

10.   MMP may not be the **answer**, but the pressure for proportional representation in some form seems likely to increase as time goes on. (ICE-NZ: W2A011)

Third, the L2 students make relatively frequent use of DONs expressing uncertainty, specifically, the DONs *impression*, *possibility*, and *suggestion*. Of these, only *possibility* is significantly more frequent. Example (11) illustrates the use of *possibility* in the L2 student corpus.

11.   Today, if somebody asked me about my impression of the world I would probably say that it is more of a hell than a paradise; although I would add that perhaps there is a **possibility**, small as it might seem, for humanity to turn it into a paradise some day. (ICLE-SW: SWUL8007)

The impression of uncertainty created by the choice of DON in this example is further strengthened by the use of modal auxiliaries (*would*, *might*), and modal adverbials (*probably*, *perhaps*). That *possibility* is often used in close proximity with modal auxiliary verbs is also seen in the use of *possibility* in the L1 student writing (example 12) and the expert writing (example 13):

12.   Exactly how far the negotiations on <u>a single monetary union</u> will go is still debatable
          but it must be seen as a real **possibility**. (LOCNESS: BRSUR3)

13.   The **possibility** <u>that her husband may be injured or killed </u>not only places an extra
          stress on the military wife but can also be used by her husband and the military as
          a form of emotional blackmail. (ICE-NZ: W2A017)

The modal auxiliaries in these examples are *will* and *must* (in 12) and *may* (in 13). Thus, a wide range of modal meanings is possible.

Finally, the L2 student writing is characterised by a significantly larger proportion of DONs which are semantically vague or flexible, notably the DONs *thing* and *fact*. The high frequency of *thing* in the L2 student writing is particularly noteworthy, with *thing* being over four times more frequent in the L2 student corpus than in the L1 student corpus, and almost ten times more frequent than in the expert corpus (see Appendix). These differences may be due at least in part to the tendency of the L2 student writers to overuse semantically flexible, general-purpose nouns in lieu of more specific items. These findings corroborate previous studies by other scholars. For instance, in a study on the writing of Norwegian advanced learners, Hasselgren (1994) found that they overuse so-called lexical teddy bears (Hasselgren's term), that is, words that are general in meaning and that they feel safe with. Ringbom (1998) came to similar conclusions in a comparative study on advanced learner language and native-speaker language; he found the learner language investigated to be vague, with particularly frequent use of the nouns *people* and *thing*. And finally, in a book-length study by J. Flowerdew and Forest (2015) on signalling noun usage in academic language, the authors found the frequency of *thing* to be affected by mode (i.e., *thing* was found to be significantly more frequent in spoken than in written language) but not by discipline (i.e., *thing* was found to be equally frequent in the natural and the social sciences).

Examples (14)–(16) illustrate the use of *thing* in the L2 student corpus. Modifiers and determiner have been italicised.

14.   I like my subject and that is probably *the most important* **thing** when choosing a topic; it has to be a subject which interests you. (ICLE-SW: SWUL8027)

15.   This vicious circle is *a common* **thing** in most countries where different cultures meet and try to live together. (ICLE-SW: SWUG2011)

16.   *Another* **thing** that has made people less creative and more passive is the revolution of the computer. (ICLE-SW: SWUL3020)

These examples are interesting for a number of reasons. First, the use of *thing* in these examples is both correct and idiomatic. Second, the use of *thing* here is also imprecise and informal. Other more specific DONs which may have been used in these contexts include *factor* (examples 14 and 15), *phenomenon* or *problem* (example 15), and *event* (example 16). Third, example (14) also contains first- and second-person reference (*I, you*), which is atypical of formal academic writing, and example (16) contains the equally vague noun *people*, also frequently used in L2 student writing (see Ringbom 1998). And fourth, each instance of *thing* is preceded by a pre-modifier (in 14 and 15) or a determiner (in 16), items which contribute to attitudinal meaning (in 14), to propositional meaning (in 15), and to organisational meaning (in 16). A central function of *thing* is therefore that of being used as a peg for information expressed elsewhere in the nominal group (see also Francis 1986, 1994).

This is also the predominant use of *thing* in the L1 student writing and the expert writing, as illustrated in examples (17) and (18), respectively.

17.   Scientists today are more concerned with advancing their knowledge, and I consider this *a good* **thing**, because how else would technology be updated. (LOCNESS: alevels8)

18.   While women generally leave paid employment younger than men, most women never retire at all because their major responsibility for housework is never-ending. Ironically, this may be *a good* **thing** for women – it continues giving them a sense of purpose while their male partners often face a life made meaningless without paid work. (ICE-NZ: W2A008)

As illustrated, the use of *thing* here is qualitatively similar to that in the L2 student writing, in that most of the semantic content of the noun group lies in the modifier (*good*) in each example. The main difference across the corpora is therefore one of frequency.

On the basis of these findings, it can be concluded that the L2 student writing contains a higher proportion of DONs expressing attitude and involvement, and a higher proportion of semantically vague or flexible DONs, in comparison with the other writer groups.

3.2.2 Overrepresentation of DONs in the L1 students' writing

In the L1 student corpus, a total of thirteen types were found to be significantly more frequent than in the other corpora, namely *argument*, *attitude*, *belief*, *claim*, *debate*, *discovery*, *idea*, *issue*, *point*, *reasoning*, *statement*, *theme*, and *theory*. An additional five DONs were found to be more frequent in the L1 student corpus but not significantly so. These DONs are *dilemma*, *illusion*, *obstacle*, *option*, and *realisation* or *realization*.

On the basis of these findings, it can be concluded that the L1 student writing is characterised by DONs pertaining to the domain of argumentation. Specifically, the DONs *argument*, *claim*, *debate*, *idea*, *issue*, *point*, *reasoning*, *statement*, and *theory* are all significantly more frequent in the L1 student corpus. The prototypical function of these DONs is to label and evaluate different standpoints and contributions in a debate – either one's own or those of a real or imagined opponent. Such usage is illustrated in example (19).

19.   The **claim** <u>that the death penalty is a superior deterrent</u> is simply weak and unsupported. The fact is that society simply craves violence and capital punishment is one of the only legal means of achieving it. (LOCNESS: USARG)

In this example, the DON *claim* is used to label and evaluate the standpoint of an imagined opponent. This standpoint is then refuted, citing lack of supporting evidence, and the author's own alternative standpoint is presented, labelled a *fact*. The difference between the implied truth-values of *claims* and *facts* may go a long way towards explaining the choice of DON here and in similar places in the L1 student material. Similar usage is illustrated in examples (20) and (21), from the L2 student writing and the expert writing, respectively.

20.   For example, attitudes towards homosecxuality in Society have improved since the Bible's **claim** <u>that homosexuality was a cardinal sin</u>. (ICLE-SW: SWUG2064)

21.   […] I shall later argue that this last **claim** is false, or at best trivial or misleading. (ICE-US: W2A010)

Although similar examples can be found in the L2 student writing and in the expert writing, they are significantly less frequent there. The frequent use of debate nouns therefore indicates a greater tendency on the part of the L1 student writers to identify and evaluate standpoints in a debate, as a rhetorical device typical of argumentative writing. The L1

student essays can therefore be classified as clearly argumentative in nature, both in terms of the set writing task and in terms of DON usage.

3.2.3 Overrepresentation of DONs in the expert writing

In the corpus of expert writing, 30 DONs were identified as being significantly more frequent than in the other two corpora. These DONs are *account, approach, assumption, attempt, change, concept, concern, decision, difference, evidence, factor, feeling, finding, goal, hypothesis, implication, method, objective, observation, principle, probability, process, proposal, requirement, result, risk, sense, step, tendency,* and *view*. The following 17 additional DONs were also found to be more frequent in the expert writing, but for these DONs the difference was not significant: *assertion, contention, desire, expectation, ground, likelihood, manner, need, notion, perception, prediction, premise, proposition, provision, reality, recognition,* and *task*.

Two observations were made on the basis of these findings. First, the expert corpus is characterised by greater variation in the use of DONs than the two student corpora, in that it contains by far the highest number of types with a significantly higher frequency in comparison with the other corpora. There are some differences in the design of the expert corpus on the one hand and the two student corpora on the other, which may at first sight explain this result. These differences relate to the number of texts (i.e., there are fewer but longer texts in the expert corpus; see Table 1) as well as to the number of topics (i.e., unlike the student corpora, each text in the expert corpus is on a separate topic). However, DONs are, by definition, semantically flexible and rarely topic-specific, so the differences in corpus design cannot entirely explain the differences in variety of usage.

The second observation is that twelve of the DONs listed here are potentially research-oriented, namely *evidence, finding, goal, hypothesis, method, objective, observation, premise, process, proposal, result,* and *step*. These DONs can be expected to be particularly frequent in the type of writing represented in the expert corpus. The significant difference in frequency is therefore not surprising. However, the so-called research-oriented DONs also have a more general meaning and use, which predominates in the student writing. *Result* is one such DON. Examples (22)–(24) illustrate the use of *result* in the three corpora.

22. One **result** of the economic crisis is <u>the transfer of responsibility for education from the governmental level to the local authorities, which has brought about several negative implications</u>. (ICLE-SW: SWUG2046)

23. After an examination, he family doctor told her that <u>her ovaries had been damaged</u> and it appeared to be the **result** of an abortion gone bad. (LOCNESS: USARG)

24. In Thailand, the **results** of that study showed <u>some of the same contradictions found among the Maori but not Pakeha, between the large families of older generations, and the small family normative structures of younger couples</u> […]. (ICE-NZ: W2A007)

Only a handful of cases were found in the student writing of *result* being used to refer to the results of a scientific study. Three of these instances are from the L1 student writing, and one is from the L2 student writing and is cited below as example (25).

25. **Results** from research show <u>that learning is more rapid with CAI than with other training methods</u>. (ICLE-SW: SWUL3013)

These results are particularly interesting when seen from the perspective of comparability between the corpus materials on which this study was based. As reported, the two student corpora consist of undergraduate student essays on a number of set topics, a large percentage of which were written under exam conditions (see Ädel 2008 on the effect of writing conditions on the frequency of involvement features in L2 and L1 student writing). In contrast, the expert material consists of reports on original research, in which the design and findings of scientific studies can be expected to be central. It is therefore not surprising that the expert corpus contains more instances of research-oriented nouns than the two student corpora. However, as discourse-organising nouns are semantically flexible, high-frequency, general-purpose nouns with a wide range of applicability this objection has limited validity.

4. Discussion and conclusion
As the above analysis shows, there are similarities across the corpora, but there are also key differences in terms of what DONs are used and how often they are used. Overall, the L2 students more often use a few high-frequency DONs, in comparison with the other writer groups, but they use fewer *different* DONs. Another key finding is that although the L2 students' usage was found to be idiomatic in many ways, it was also found to be colloquial, emotive, and vague, and therefore atypical of formal

academic writing. The L2 student writing is characterised by relatively high frequencies of semantically vague or flexible nouns (such as *thing* and *fact*), which can serve as general-purpose items, as well as by items expressing attitude and involvement (such as *opinion* and *question*). The L2 student writers are also those making the most use of interpersonal markers in their use of DONs. The L1 student writing is characterised by relatively high frequencies of debate-oriented nouns (such as *argument* and *issue*), which are typical of argumentative writing and can serve to label and evaluate standpoints in a (real or imagined) debate. In both student corpora, DONs are frequently used in the construction of a mock dialogue. And finally, the expert writing is characterised by the greatest degree of variety in their use of DONs as well as by relatively high frequencies of research-oriented DONs (such as *evidence* and *finding*).

Thus, the findings of the present study give us interesting information about the nature of L2 student writing, as represented in the corpus material used here. A large number of studies (e.g., Ädel 2003, 2006; Aijmer 2001, 2002; Altenberg 1997; Altenberg & Tapper 1998; Gilquin & Paquot 2007, 2008; Granger 1998, 2007; Lorenz 1999) have found that L2 student writing—even at fairly advanced levels—can be characterised as informal and spoken-like in many respects. The present study corroborates and complements these findings, in that DON usage was found to be an area in which L2 student usage is often idiomatic but colloquial, imprecise, and informal. Such characteristics of L2 student writing may have unintended and infelicitous effects on the finished product, not only in terms of variety and exactness but also in terms of style and register appropriacy.

A related finding is that many of the characteristics of the L2 student writing were also found in the L1 student writing, though to a lesser extent; the differences between the L2 and L1 student writing were therefore often smaller than those between the L2 student writing and the expert writing (for congruent findings, see Bolton et al. 2002; Gilquin & Paquot 2008). In short, being a native speaker of the target *language* (here, English) does not necessarily entail mastery of the target *register* (here, academic writing). In this respect, the two student groups are on a relatively even footing; they are both made up of novice academic writers. This is an important finding, particularly in view of the fact that many comparative studies on learner writing have been limited to L2 and L1 student material (e.g., Aijmer 2001, 2002; Ädel 2003, 2006; Boström Aronsson 2005). The

trilateral design of the present study, with its inclusion of expert writing, has thus enabled a fuller investigation of the nature of L2 and L1 student writing.

The differences found across the corpora may be due to a number of possible factors. Previous studies identify, for instance, teaching effects, text-type effects, developmental factors, and L1 transfer as factors which may impact L2 students' writing in various ways (see, e.g., Ljung 1990, 1991; Altenberg 2002; Gilquin & Paquot 2007, 2008). In addition, Altenberg (1997: 130) has suggested that student writers may have 'a general lack of register awareness', which may impact their writing. To further complicate matters, a study by Ädel (2008) found that L2 students' writing is affected by genre and production circumstances. Ädel found a statistically significant correlation between, on the one hand, the relative frequency of certain involvement features (e.g., first-person pronouns and questions), and, on the other hand, the variables *task setting* (here, whether essays were timed or untimed) and *intertextuality* (here, whether students had access to secondary sources while writing). These findings are in line with previous studies which have also found production circumstances to have a significant impact on language at many linguistic levels (e.g., Chafe 1982; Chafe & Danielewicz 1987; Biber 2006). In the present study, differences in genre and production circumstances were not investigated systematically; however, it seems fair to say that the time contraints under which the L2 students produced their texts (66% of the ICLE-SW essays are timed), as well as their lack of access to secondary sources, likely affected the outcome. The extent to which these factors have had an impact on the L2 student material in the present study remains an open question, but as J. Flowerdew (2006: 360) points out, the reasons for students' infelicitous and non-nativelike usage are likely to be complex (see also Ellis & Barkhuizen 2005: 66).

What this study has not investigated is how the insights presented here might be implemented in the academic writing classroom. However, the pervasiveness of these nouns in academic discourse (see, e.g., J. Flowerdew 2003a, 2003b, 2006; Hoey 1983: 63) and the characteristics of student writing reported in the present study suggest that student writers may profit from being taught the meanings and functions of these nouns, and their use in academic discourse. A number of scholars have addressed this issue. Aktas and Cortes (2008: 13) suggest aspects of shell noun usage which may more profitably be addressed than others. Specifically, they

argue that students may profit primarily from being taught the syntactic frames in which such nouns occur (the syntagmatic perspective), rather than being taught the nouns themselves as vocabulary items (the paradigmatic perspective). Tåqvist (2016) also found evidence to suggest that student writers have difficulties relating to the lexico-grammatical patterns in which the nouns occur; however, it is clear from the present study that students encounter a wide variety of difficulties relating to DON usage, including variety and accuracy, collocation, and style. This finding accords well with Granger's (1993) suggestion that advanced learner writers are likely to struggle primarily with vocabulary and style. In an earlier study, Francis (1988: 227–338) suggested that lexical cohesion may profitably be taught in the context of a genre-based pedagogy (see also, e.g., J. Flowerdew 2015: 31). In view of the finding that student usage is often idiomatic but colloquial and non-academic, such a genre-based pedagogical approach to the teaching of discourse-organising nouns may indeed prove to be useful.

References
Primary sources
The International Corpus of Learner English (ICLE)
http://uclouvain.be/en-cecl-icle.html
The Louvain Corpus of Native English Essays (LOCNESS)
http://uclouvain.be/en-cecl-locness.html
The International Corpus of English (ICE)
http://ice-corpora.net/ice

Secondary sources
Ädel, A. (2003) *The use of metadiscourse in argumentative texts by advanced learners and native speakers of English.* (Doctoral dissertation), University of Gothenburg, Göteborg.
Ädel, A. (2006) *Metadiscourse in L1 and L2 English.* Amsterdam: John Benjamins.
Ädel, A. (2008) Involvement features in writing: Do time and interaction trump register awareness? In S. Gilquin, S. Papp, & M. Díez-Bedmar (Eds), *Linking up contrastive and learner corpus research* (pp. 35–53). Amsterdam: Rodopi.

Aijmer, K. (2001) *I think* as a marker of discourse style in argumentative student writing. In K. Aijmer (Ed), *A wealth of English: Studies in honour of Göran Kjellmer* (pp. 247–257). Göteborg: Acta Univesitatis Gothoburgensis.

Aijmer, K. (2002) Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung, & S. Petch-Tyson (Eds), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 55 – 74). Amsterdam: John Benjamins.

Aijmer, K. (2005) Why is *of course* so frequent in Swedish learners' argumentative texts? In S. Granath, J. Miliander, & E. Wennö (Eds), *The power of words: Studies in honour of Moira Linnarud* (pp. 1–9). Karlstad: Karlstad University Press.

Aktas, R. & V. Cortes. (2008) Shell nouns as cohesive devices in published and ESL student writing. *Journal of English for Academic Purposes, 7*(1), 3–14.

Altenberg, B. (1997) Exploring the Swedish component of the International Corpus of Learner English. In B. Lewandowska-Tomaszcyk & P. Melia (Eds), *International conference on practical applications in language corpora, Łódź, Poland, 10–14 April 1997: Proceedings* (pp. 119–132). Łódź: Łódź University Press.

Altenberg, B. (2002) The status of English in Sweden. In S. Granger, E. Dagneaux, F. Meunier, & M. Paquot (Eds), *The International Corpus of Learner English. Handbook and CD-ROM* (pp. 174–179). Louvain-la-Neuve: Universitaires de Louvain.

Altenberg, B. & M. Tapper. (1998) The use of adverbial connectors in advanced Swedish learners' written English. In S. Granger (Ed), *Learner English on computer* (pp. 80–93). London: Longman.

Aron, A., E. Aron, & E. Coups. (2005) *Statistics for the behavioral and social sciences: A brief course* (3rd edn). London: Prentice Hall International.

Biber, D. (2006) *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Bolton, K., G. Nelson, & J. Hung. (2002) A corpus-based study of connectors in student writing: Research from the International Corpus of English in Hong Kong (ICE-HK). *International Journal of Corpus Linguistics, 7*(2), 165–182.

Boström Aronsson, M. (2005) *Themes in Swedish advanced learners' written English.* (Doctoral dissertation), University of Gothenburg, Göteborg.

Chafe, W. (1982) Integration and involvement in speaking, writing and oral literature. In D. Tannen (Ed), *Spoken and written language: Exploring orality and literacy* (pp. 35–53). Norwood, NJ: Ablex.

Chafe, W. & J. Danielewicz. (1987) Properties of spoken and written language. In R. Horowitz & J. Samuels (Eds), *Comprehending oral and written language* (pp. 83 – 113). San Diego: Academic Press.

Chiang, S. (1999) Assessing grammatical and textual features in L2 writing samples: The case of French as a Foreign Language. *The Modern Language Journal*, *83*(2), 219–232.

Chiang, S. (2003) The importance of cohesive conditions to perceptions of writing quality at the early stages of foreign language learning. *System*, *31*(2), 471–484.

Davies, M. (2008–) *The Corpus of Contemporary American English: 450 million words, 1900–present.* Available online at http://corpus.byu.edu/coca/.

Ellis, R. & G. Barkhuizen. (2005) *Analysing learner language*. Oxford: Oxford University Press.

Flowerdew, J. (2003a) Register-specificity of signalling nouns in discourse. In C. Meyer and P. Leistyna (Eds), *Corpus analysis: Language structure and language use* (pp. 35–46). Amsterdam: Rodopi.

Flowerdew, J. (2003b) Signalling nouns in discourse. *English for Specific Purposes*, *22*(1), 329–346.

Flowerdew, J. (2006) Use of signalling nouns in a learner corpus. *International Journal of Corpus Linguistics*, *11*(3), 345–362.

Flowerdew, J. (2010) Use of signalling nouns across L1 and L2 writer corpora. *International Journal of Corpus Linguistics*, *15*(1), 36–55.

Flowerdew, J. (2015) Revisiting metadiscourse: Conceptual and methodological issues concerning signalling nouns. *Ibérica*(29), 15–34.

Flowerdew, J. & R. Forest. (2015) *Signalling nouns in English.* Cambridge: Cambridge University Press.

Flowerdew, L. (2001) The exploitation of small learner corpora in EAP materials design. In M. Ghadessy, A. Henry, & R. L. Roseberry (Eds),

Small corpus studies and ELT (pp. 363–379). Philadelphia, PA: John Benjamins.

Francis, G. (1986) *Anaphoric nouns.* Birmingham: University of Birmingham.

Francis, G. (1988) The teaching of techniques of lexical cohesion in an ESL setting. In V. Bickley (Ed.), *Languages in education in a bilingual or multi-lingual setting* (pp. 325–338). Hong Kong: Institute of Language in Education.

Francis, G. (1994) Labelling discourse: An aspect of nominal-group lexical cohesion. In M. Coulthard (Ed.), *Advances in written text analysis* (pp. 83–101). London: Routledge.

Gilquin, G. & M. Paquot. (2007) Spoken features in learner academic writing: Identification, explanation and solution. Paper presented at The Fourth Corpus Linguistics Conference CL2007, Birmingham.

Gilquin, G. & M. Paquot (2008) Too chatty: Learner academic writing and register variation. *English Text Construction*, *1*(1), 41–61.

Granger, S. (1993) The international corpus of learner English. In J. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English language corpora: Design, analysis and exploitation* (pp. 51–71). Amsterdam: Rodopi.

Granger, S. (1998) The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Ed), *Learner English on computer* (pp. 3–18). London: Longman.

Granger, S. (2007) The computer learner corpus: A versatile new source of data for SLA research. In. W. Teubert & R. Krishnamurthy (Eds), *Corpus linguistics* (Vol. 2, pp. 166–182). London: Routledge.

Halliday, M. & R. Hasan. (1976) Cohesion in English. London: Longman.

Hasselgren, A. (1994) Lexical teddy bears and advanced learners: A study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, *4*(2), 237–260.

Hoey, M. (1983) On the surface of discourse. London: George Allen & Unwin.

Jafarpur, A. (1991) Cohesiveness as a basis for evaluating compositions. *System*, *19*(4), 459–465.

Ljung, M. (1990) *A study of TEFL vocabulary*. Stockholm: Almqvist & Wiksell.

Ljung, M. (1991) Swedish TEFL meets reality. In S. Johansson & A-B. Stenström (Eds), *English computer corpora: Selected papers and research guide* (pp. 245–256). Berlin: Mouton de Gruyter.

Lorenz, G. (1999) Learning to cohere: Causal links in native vs. non-native argumentative writing. In W. Bublitz, U. Lenk, & E. Ventola (Eds), *Coherence in spoken and written discourse: How to create it and how to describe it. Selected papers from the International Workshop on Coherence, Augsburg, 24–27 April 1997* (pp. 55–75). Amsterdam: John Benjamins.

McCarthy, M. (1991) *Discourse analysis for language teachers.* Cambridge: Cambridge University Press.

Mondor, M. (2008) *Figuring it out: A corpus-based comparison of the verb-particle construction in argumentative writing by Swedish advanced learners and native speakers of English.* (Doctoral dissertation), University of Gothenburg, Göteborg.

Ringbom, H. (1998) Vocabulary frequencies in advanced learner English: A cross-linguistic approach. In S. Granger (Ed), *Learner English on computer* (pp. 41–52). London: Longman.

Rørvik, S. (2013) *Texture in learner language.* (Doctoral dissertation), University of Oslo, Oslo.

Schmid, H-J. (2000) *English abstract nouns as conceptual shells: From corpus to cognition.* Berlin: Mouton de Gruyter.

Schmid, H-J. (2007) Non-compositionality and emergent meaning of lexico-grammatical chunks: A corpus study of noun phrases with sentential complements as constructions. *Zeitschrift für Anglistik und Amerikanistik, 55*(3), 313–340.

Scott, M. (2012) WordSmith Tools (Version 6.0). Liverpool: Lexical Analysis Software.

Tåqvist, M-K. (2016) *"Another thing": Discourse-organising nouns in advanced learner English.* (Doctoral dissertation), Karlstad University, Karlstad.

Winter, E. (1977) A clause-relational approach to English texts: A study of some predictive lexical items in written discourse. *Instructional Science, 6*(1), 1–92.

Winter, E. (1978) A look at the role of certain words in information structure. In K. P. Jones & V. Horsnell (Eds), *Informatics 3* (pp. 85–97. London: ASLIB.

Appendix

Table 5 shows the frequencies of DON usage (tokens) in the three corpora, in absolute and normalised frequencies. The Chi-square (χ^2) and Cramér's V (φ_c) tests were used to test for statistical significance and effect size. The absolute frequency of each DON was tested in proportion to the total number of DONs in each corpus.

Table 5. Frequencies of DONs across the three corpora

DON	L2 students (ICLE-SW)		L1 students (LOCNESS)		Expert writing (ICE)		p^6	φ_c
	Abs freq	Freq per 200k	Abs freq	Freq per 200k	Abs freq	Freq per 200k		
account*[7]	4	4	2	1	10	8	.028	.032
answer*	18	18	11	7	10	8	.041	.030
approach**[8]	30	30	21	13	81	64	.000	.086
argument**	18	18	155	96	40	32	.000	.123
assertion[9]	0	0	2	1	9	7		
assumption**	6	6	6	4	32	25	.000	.067
attempt**	14	14	25	15	52	41	.000	.059
attitude**	10	10	27	17	3	2	.000	.048
belief*	17	17	37	23	13	10	.014	.035
change**	41	41	23	14	58	46	.000	.056
claim**	4	4	31	19	11	9	.000	.049
concept**	16	16	39	24	65	51	.000	.062
concern**	4	4	16	10	27	21	.001	.046
conclusion	22	22	30	19	14	11	.124	.024
contention	0	0	0	0	2	2		
conviction	3	3	1	1	3	2		

TO BE CONTINUED

---

[6] Chi-square test, df 2.

[7] A single asterisk stands for statistical significance at the .05 level.

[8] A double asterisk stands for statistical significance at the .01 level.

[9] Significance testing was not performed when two or more expected values were below 5.

danger	14	14	9	6	9	7	.134	.024
debate*	2	2	13	8	3	2	.015	.034
decision**	3	3	26	16	27	21	.001	.046
desire	4	4	10	6	12	9	.238	.020
difference*	17	17	19	12	35	28	.011	.036
dilemma	5	5	12	7	2	2	.055	.029
discovery**	0	0	22	14	0	0	.000	.070
doubt	2	2	0	0	1	1		
dream**	24	24	1	1	0	0	.000	.090
evidence**	0	0	0	0	33	26	.000	.097
expectation	1	1	1	1	4	3		
fact**	165	165	215	133	66	52	.000	.100
factor**	24	24	51	31	81	64	.000	.063
fear	15	15	12	7	8	6	.133	.024
feeling*	23	23	15	9	30	24	.015	.034
finding**	0	0	1	1	37	29	.000	.100
goal**	0	0	0	0	27	21	.000	.088
ground	2	2	3	2	10	8		
hope	11	11	11	7	5	4	.192	.022
hypothesis**	0	0	1	1	20	16	.000	.072
idea**	49	49	190	117	36	28	.000	.125
illusion	1	1	3	2	1	1		
implication**	1	1	5	3	17	13	.000	.051
impression	9	9	1	1	2	2		
issue**	51	51	164	101	74	58	.000	.076
likelihood	0	0	1	1	14	11		
manner	0	0	3	2	4	3		
matter*	37	37	26	16	30	24	.023	.033
meaning	8	8	2	1	4	3		
message	8	8	8	5	6	5	.670	.011
method**	14	14	27	17	75	59	.000	.087
necessity**	13	13	2	1	3	2	.000	.050
need	48	48	58	36	65	51	.288	.019
notion	12	12	15	9	22	17	.199	.021
objective**	0	0	8	5	19	15	.000	.053
observation**	0	0	3	2	21	17	.000	.068

TO BE CONTINUED

obstacle	4	4	8	5	4	3	.651	.011
opinion**	66	66	51	31	9	7	.000	.086
option	11	11	32	20	24	19	.092	.026
perception	2	2	1	1	12	9		
point**	20	20	68	42	33	26	.000	.048
possibility**	49	49	18	11	40	32	.000	.061
prediction	0	0	1	1	2	2		
premise	2	2	1	1	6	5		
principle**	2	2	10	6	35	28	.000	.073
probability**	0	0	1	1	16	13	.000	.064
problem**	229	229	278	171	132	104	.000	.085
process**	21	21	39	24	109	86	.000	.105
proposal*	2	2	12	7	20	16	.002	.043
proposition	0	0	2	1	4	3		
provision	1	1	3	2	4	3		
question**	145	145	125	77	61	48	.000	.093
reality	7	7	4	2	4	3		
realisation/ realization	1	1	8	5	4	3		
reason**	120	120	151	93	62	49	.000	.069
reasoning*	2	2	17	10	4	3	.003	.041
recognition	1	1	0	0	7	6		
requirement**	1	1	0	0	20	16	.000	.072
result**	46	46	70	43	123	96	.000	.074
risk*	9	9	12	7	27	21	.003	.041
sense**	3	3	10	6	53	42	.000	.098
situation	51	51	59	36	36	28	.057	.028
solution**	44	44	47	29	16	13	.000	.050
statement**	16	16	44	27	14	11	.001	.044
step**	0	0	0	0	23	18	.000	.081
suggestion	10	10	7	4	5	4	.185	.022
task	24	24	22	14	35	28	.061	.028
tendency*	14	14	13	8	25	20	.044	.030
theme*	13	13	46	28	23	18	.004	.040
theory**	7	7	44	27	17	13	.000	.054
thesis	0	0	4	2	2	2		

TO BE CONTINUED

thing**	236	236	91	56	30	24	.000	.199
topic	20	20	30	19	14	11	.158	.023
truth*	12	12	17	10	3	2	.017	.034
view*	16	16	51	31	41	32	.008	.037
viewpoint	1	1	8	5	6	5		
wish**	16	16	1	1	7	6	.000	.054
TOTAL**	1,994	1,994	2,770	1,709	2,345	1,848	.000	.006