

*Facts, ideas, questions, problems, and issues in advanced learners' English*¹

Hilde Hasselgård, University of Oslo

Abstract

Recurrent word combinations containing the nouns *fact*, *idea*, *question*, *problem* and *issue* are explored in three corpora of advanced learner English and a corpus of native speaker English, focusing on the comparison between Norwegian learners and native speakers. Native speakers use the nouns in recurrent word combinations more frequently than learners. Norwegian learners underuse *idea* and *issue*, whose use in English cannot be easily related to any structure in their L1. They also underuse combinations that reflect extended noun phrases, e.g. *the NOUN of/that*, and favour simple phrases such as *this NOUN* and *the NOUN is*.

1 Introduction

The present study explores the use of a small set of abstract nouns in advanced learner English, namely *fact*, *idea*, *question*, *problem*, and *issue*. A particular point of interest is the phraseology of these words. Abstract nouns such as *fact* and *question* acquire much of their meaning from the context; “Words mean things in the context of other words” (Ellis 2008: 1), because “the complete meaning of a word is always contextual” (Firth 1957: 7). The focus of this study will thus be on recurrent word combinations containing one of the nouns *fact*, *idea*, *question*, *problem* and *issue*. These nouns, though somewhat randomly chosen, have in common that they can be used as *shell nouns* (Hunston & Francis 1999, Schmid 2000), i.e. “they have, to varying degrees, the potential for being used as conceptual shells for complex, proposition-like pieces of information” (Schmid 2000: 4). An example is *the fact that*, where *fact* refers cataphorically to the projected *that*-clause and labels its content as ‘fact’. The shell noun function is associated with lexical cohesion, though often using different terms, e.g. ‘signalling nouns’ (Flowerdew 2006), and ‘labels’ (Francis 1994). The use of shell

¹ I am grateful to the following colleagues, who have read and commented on this paper at various stages of completion and thereby helped improve the present version: Cecilia Alvstad, Kjersti Bale, Signe O. Ebeling, Maria F. Krave and the *NJES* reviewer.

nouns thus has a textual function. At the same time as the labelling of something as 'fact', as against e.g. 'idea', involves some degree of evaluation (cf. Schmid 2000: 8), thus also assuming an interpersonal function. Finally, the words may have a primarily referential function, as when *question* refers to a question that has been asked, or *idea* is used in the sense of "a thought that you have about how to do something or how to deal with something" (Macmillan). The textual and interpersonal uses of these nouns may belong to relatively advanced language mastery, and are thus of particular interest in a study of learner language.

Previous studies (e.g. Nesselhauf 2005, Paquot 2010) have shown that learners do not always use collocations in native-like fashion, even if their language may be grammatically correct (see also Pawley & Syder 1983). The main questions to be explored here are the following: How do Norwegian learners use the nouns *fact*, *question*, *issue*, *problem*, and *idea* compared to native speakers and to other learner groups? Do learners and native speakers use the same recurrent word combinations? Do the learners use the word combinations in appropriate contexts and with appropriate discourse functions?

2 Material and method

The investigation is based on the International Corpus of Learner English (ICLE) and the Louvain Corpus of Native English Essays (LOCNESS). Three subcorpora of ICLE have been used, viz. those where the learners have Norwegian (ICLE-NO), German (ICLE-GE) or French (ICLE-FR) as their first language. The three learner groups were chosen to represent both Germanic and Romance language backgrounds. The essays in the ICLE subcorpora are all written by university students of English, and most of them are argumentative. The LOCNESS essays are more varied, representing more genres (though mainly expository and argumentative) and a wider range of topics and being written by both university and secondary school students. Supplementary data have been drawn from the British National Corpus (BNC) and the English-Norwegian Parallel Corpus (ENPC).²

² For more information on the corpora, see the websites cited at the end of this paper.

The ICLE subcorpora have been accessed from the ICLeV2 CD-ROM. To identify recurrent word combinations, the selected subcorpora were downloaded for analysis with the corpus tool AntConc.³ The ‘cluster’ function of this tool allows searches for word combinations of any length containing a specified word. The length of the cluster was set to 2-4 since Altenberg’s investigation (1998: 102) showed that most recurrent word combinations lie within this band. Longer recurrent word combinations will be discussed as extended patterns of 2-4-word clusters. The units studied are thus not collocations in the statistical sense of the word or phraseological units in the sense of Gläser (1998: 127 f.), but simply combinations of words that recur in identical form (Altenberg 1998: 101) and may therefore be viewed as “routinized and more or less prefabricated expressions” (ibid.: 120).

More precisely, recurrent word combinations containing the relevant nouns were selected according to the following principles: (i) they should have a minimum frequency of 5 in at least one of the learner corpora or 7 in LOCNESS due to the larger size of the corpus; (ii) they should overlap as little as possible. Thus for instance the bigram *fact that* was excluded because it almost always overlaps with either *the fact that* or *it is a fact*. Some recurrent 4-grams containing more frequent 3-grams have been regarded as collocation patterns of the 3-gram (an example is *to the fact that*, which is discussed as a collocation pattern of *the fact that*). The pattern *a/the* + NOUN was not considered phraseologically interesting and thus excluded.⁴ No normative criteria were applied in selecting the material; the reason why no unidiomatic word combinations occur in the surveys presented below is simply that they did not occur above the frequency threshold of 5, unlike Paquot’s findings (2010: 160 ff) in her study of *conclusion*. The core material consists of uninterrupted sequences, but variations on the most frequent phrases have been searched for and studied separately.

The investigation is both qualitative and quantitative. The patterns and meanings of the most frequent clusters will be studied in some detail with a view to finding differences and similarities between learner and native-speaker usage and identifying any learner problems. The focus on

³ For information on AntConc, see www.antlab.sci.waseda.ac.jp/antconc_index.html.

⁴ See Altenberg (1998: 102 f.) for a similar discussion of inclusion criteria.

the investigation is the comparison of patterns found in ICLE-NO and LOCNESS. The backdrop of patterns in ICLE-FR and ICLE-GE is, however, interesting for distinguishing “the phraseological features common to several categories of learners from the L1-dependent features” (Granger 1998: 159).

3 Some overall frequencies

Table 1 shows the overall frequencies of the investigated words across the corpora. Results from each learner corpus have been compared to LOCNESS correlating raw frequencies with corpus size and using the chi square test ($df = 1$). The use of bold type in Table 1 indicates that the difference between the learner corpus and LOCNESS is statistically significant at $p \leq 0.05$. Figure 1 gives frequencies of the nouns per 100,000 words.

Table 1. Raw frequencies of *fact*, *question*, *issue*, *problem*, and *idea* across corpora

	<i>fact</i>	<i>idea</i>	<i>question</i>	<i>problem</i>	<i>issue</i>	Corpus size ⁵
ICLE-NO	232	54	131	121	58	213,940
ICLE-FR	250	122	149	179	22	206,194
ICLE-GE	233	91	135	144	10	240,917
LOCNESS	306	205	147	271	157	326,089

Table 1 and Figure 1 show that most of the nouns are more frequent in ICLE-FR than in the other learner corpora; *fact*, *question* and *problem* are also more frequent than in LOCNESS.⁶ Compared to LOCNESS, all

⁵ The numbers of words in the ICLE subcorpora differ from those given on the ICLEv2 CD. However, as AntConc was used for analysing LOCNESS, this tool was used to calculate ICLE size too, to ensure that all the subcorpora were counted in the same way.

⁶ In a study of shell nouns in research papers by international graduate students compared to published research papers Aktas & Cortes (2008: 7) found *problem* and *issue* to be more frequent in the student corpus than in the published writing, while *fact* was marginally more frequent in the published writing. However, Aktas & Cortes's figures include only the uses of the nouns that have shell functions.

the learner groups overuse *fact*, though the overuse is significant only in ICLE-FR. Likewise, all the learners use *question* more frequently than native speakers; the overuse is significant in both ICLE-FR and ICLE-NO. *Problem* and *idea* are significantly underused by Norwegian and German learners while French learners use them about as frequently as native speakers. *Issue* is significantly underused by all learner groups. Norwegian learners use it more than the others, but rather less frequently than native speakers.

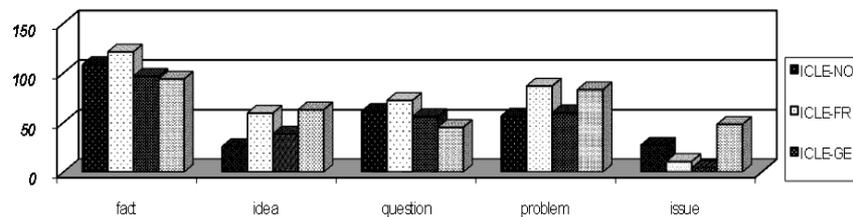


Figure 1. Relative frequencies of *fact*, *idea*, *question*, *problem*, and *issue* across corpora (per 100,000 words)

The general underuse of *issue* may reflect a lack of any direct equivalent in the first languages of the learners, which may also be a source of misuse of this word (see further section 4.5). However, equivalents of the other nouns exist in all three L1 backgrounds concerned, so that differences in usage may be due to phraseological differences between English and the learners' L1. Unfortunately, contrastive phraseological investigations are outside the scope of the present study. However, discrepancies between learners and native speakers may also be due to imperfect mastery of the rhetorical potential of these words in learner English, for example in marking such clause relations as 'problem–solution' (Hoey 1983).

4 Discussion of individual words in recurrent word combinations

The present section discusses each noun in turn, exploring the recurrent word combinations they enter into and the discourse functions served by the combinations. Only the most frequent clusters will be given more detailed attention, since a handful of examples cannot reveal patterns of use. Overuse and underuse of patterns have been calculated correlating

the frequency of the word combination with the total frequency of the relevant noun in each corpus.⁷ This has been done in order to study the relative distribution of patterns in the learner corpora independently of the overall frequency of the node noun. The overall distribution of the nouns shown in Table 1 and Figure 1 should, however, be borne in mind.

4.1 Fact

Table 2 shows the patterns for *fact*. The top row gives the total frequency of the word in each corpus, and the last two rows show the number of times *fact* enters into the recurrent combinations and a percentage of the total. Bold type signals a significant difference between the learner corpus and LOCNESS at $p \leq 0.05$ ($df=1$). It is noteworthy that *fact* occurs in recurrent word combinations between 79 and 92% of the times it is used; this gives evidence of the strong constructional tendency of *fact*. The Norwegian learners have the lowest percentage of recurrent word combinations with *fact*.

Table 2. Recurrent word combinations containing *fact* across corpora: raw frequencies and frequencies per 100,000 words.

FACT	ICLE-NO (232)		ICLE-FR (250)		ICLE-GE (233)		LOCNESS (306)	
	raw	relative	raw	relative	raw	relative	raw	Relative
<i>the fact that</i>	115	53.8	95	46.1	97	40.3	162	49.7
<i>in fact</i>	36	16.8	93	45.1	63	26.2	93	28.5
<i>it is (it's) a fact</i>	15	7.0	7	3.4	7	2.9	1	0.3
<i>matter of fact</i>	5	2.3	25	12.1	13	5.4	0	0
<i>the fact is</i>	8	3.7	4	1.9	4	1.7	7	2.1
<i>this fact</i>	4	1.9	6	2.9	12	5.0	8	2.5
total	183	85.5	230	111.5	196	81.4	271	83.1
% in recurrent combinations	78.9		92.0		84.1		88.6	

As expected, *the fact that* tops the list of recurrent combinations with *fact* across the board. It is most frequent in ICLE-NO in terms of relative frequencies and least in ICLE-GE. In terms of the distribution of patterns relative to the frequency of the noun in each subcorpus, *the fact that is*

⁷ Note that the chi square test could only be carried out on the most frequent combinations.

underused in ICLE-FR; this is partly because of the large proportion of *in fact*, as shown below. *It is (it's) a fact* is most frequent in ICLE-NO and least in LOCNESS.

Quantitatively, Norwegian learners differ from native speakers mainly in their underuse of *in fact*. The underuse of *in fact* is significant also in relation to the other learner groups. French learners stand out in their frequent use of *in fact* and *matter of fact* (see further below). German learners have a smaller proportion of *the fact that* than the other groups, as mentioned above, and a higher proportion of *this fact*, though the frequencies are too low to show significant differences.

4.1.1 The fact that

In the expression *the fact that*, *fact* has “some kind of expansion in the surrounding text, indicating what the ... fact is” (Hunston & Francis 1999: 185) and is thus a shell noun. In this expression, *fact* is an advance label, representing the proposition in the *that*-clause as factual. LOCNESS and ICLE-NO are relatively similar as regards the syntactic patterns *the fact that* occurs in. The expression functions as the complement of a preposition in 44% of the cases in LOCNESS and 42% in ICLE-NO; see example (1).⁸ It functions as direct object in 27% vs. 32%, as in example (2), and subject in 28% vs. 24%, see example (3).

- (1) Few of them had any education at all, due to *the fact that* they got children at an early age ... (ICLE-NO)
- (2) Men and women today need to understand and respect *the fact that* they are different. (LOCNESS)
- (3) Here, he has even placed a god “on earth” as it were, as if to prove that they are in fact no greater than us and *the fact that* they can produce miracles, has no bearing on their power over us... (LOCNESS)

The pattern shown in (1) was expected to be overused by Norwegian learners since it is often suggested as a correspondence of the Norwegian construction ‘preposition + infinitive or *that*-clause’ (e.g. Hasselgård et

⁸ All examples are rendered as they occur in the corpora.

al. 1998: 349).⁹ However, this was not the case. The preposition most frequently preceding *the fact that* in LOCNESS is *due to*; it occurs 21 times, reflecting the extended pattern *due to the fact that*. This pattern is less frequent in ICLE-NO, although with eight occurrences, it is the most common pattern with PREP + *the fact that*. (There were also eight other occurrences of *to* + *the fact that* in ICLE-NO.)

Interestingly, the second most common preposition to precede *the fact that* is *by*, with 10 occurrences in LOCNESS and 7 in ICLE-NO. With one exception in ICLE-NO and two in LOCNESS, *by the fact that...* functions as an agent adjunct in a passive construction, as exemplified by (4), thus mirroring the relatively frequent use of this word combination as subject.

- (4) This is explained *by the fact that* everyone is free and can make choices for his or herself... (LOCNESS)

However, some of the uses of PREP + *the fact that* in ICLE-NO are dissonant,¹⁰ because of a wrong choice of preposition (5).

- (5) This is a contradiction to *the fact that* we support the human rights. (ICLE-NO)
 (6) ... they ignore *the fact that* it is not right that this discrepancy exists. (LOCNESS)

The verbs occurring to the immediate left of *the fact that* are a mixed lot; only *be* occurs above two or three times. However, the verbs can be grouped according to meaning. A striking group in LOCNESS is made up by *ignore/overlook/mask/reject/resent*; i.e. what people do with objectionable facts (6). A second group shows a more positive attitude: *amplify, express, give, mention, point out, present, respect, state, support*; see example (2). The smallest group is made up by *address* and

⁹ Norwegian allows prepositions in front of clauses corresponding to English *that*-clauses, as in *De profitterte på at politimennene gjorde en dårlig jobb.* (ENPC: KA1). Literal translation: "You profited *on that* the policemen did a bad job." The published translation uses *fact*: *You benefited from the fact that the police did a poor job.*

¹⁰ The term 'dissonant' comes from Hasselgren (1994) and covers everything from ungrammatical to stylistically inappropriate (1994: 242 f.).

challenge. The same verb meanings were found in ICLE-NO, with *face* as an addition to the *address/challenge* group. Some verbs preceding *the fact that* in ICLE-NO, however, appear to be infelicitous collocates, e.g. *agree on* and *underestimate* in (7) and (8).

- (7) Most of us agree on *the fact that* we all are born equal and deserve and have the right to the same things. (ICLE-NO)
- (8) ... you can not underestimate *the fact that* many college degrees also need a practical side. (ICLE-NO)

In both cases the verb would suggest that the following proposition is *not* a fact. On the other hand, it is also questionable whether the proposition in the *that*-clause is really a fact. Thus (7) could be improved by omitting *the fact* together with the preposition, or *fact* might be replaced by *idea*. In (8) the label could be avoided by rephrasing the proposition, e.g. by using nominalization: ... *underestimate the need for a practical component*. Both examples give an impression of verbosity; for the latter point, see Granger (1998: 155). Note, however, that the type of dissonance shown in (8) can also occur in native English, particularly in informal registers.

When *the fact* is the head of a subject NP, as in example (3), it typically functions as clause theme and thus the entity that the proposition is about. As shown in (3), these subject NPs may be preceded by a conjunction or an adverbial. The conjunctions before *the fact that* are almost always co-ordinating. The tendency to verbosity also shows up when *the fact that* is in subject position, as in (9), where *the fact that* is superfluous (and a construction with extraposition would have been more natural).

- (9) *The fact that* the child needs to be taken care of after birth is obvious. (ICLE-NO)

The dissonant use in (9) may be a case of hypercorrection, i.e. the learner avoids a 'bare' *that*-clause even in contexts where it might be acceptable, or more likely, she uses *the fact that* as an equivalent of the Norwegian *det at* ('that_{dem} that_{conj}'), which is typically used in sentence-initial subject position. This correspondence is also found in the ENPC:

- (10) *Det at han så så "ung" ut vekket plutselig en uro i meg ... (KF2)*
The fact that he looked so young suddenly aroused a certain unease in me... (KF2T)

It seems that *the fact* is sometimes used in front of a *that*-clause to fit it more smoothly into a nominal position, as is evidenced by (11), in which *the fact that* is co-ordinated with a noun phrase. This use is found both in ICLE-NO and in LOCNESS.

- (11) That has a lot to do with equality of status, and *the fact that* women's sexuality no longer is something shameful and embarrassing. (ICLE-NO)

There is evidence in both ICLE-NO and LOCNESS that the shell noun *fact* does not always refer to a factual situation, as in (12) and (13), where what is labelled as 'fact' is rather an opinion and possibility, respectively (see also (7) and (8) above).

- (12) With this essay I have tried to share my feelings about abortion, and *the fact that* it can be right in some situations and wrong in other. (ICLE-NO)
- (13) One of the most important benefits of drug legalization is *the fact that* the prices of drugs would decrease and there would not be as much drug trade. (LOCNESS)

A likely explanation for this type of dissonance could be that the high frequency of *the fact that* leads to overgeneralization and semantic bleaching. Schmid (2000: 99) observes on the basis of native speaker data that "the construction *the fact that* seems to have lost a considerable part of its 'original' meaning and has come to be used as *the* general-purpose shelling device", thus it does not necessarily refer to a factual state of affairs. "What counts is simply that the construction *the fact that* is a very handy means of shelling events and abstract relations together" (ibid: 100).

4.1.2 In fact

In fact is the second most frequent expression with *fact* across the corpora. As the expression can be said to be a lexicalized adverbial expression, where *fact* does not have the potential of functioning as a shell noun, it will be dealt with only briefly here. Compared to LOCNESS, Norwegian learners underuse *in fact*, even though Norwegian has the cognate expression *faktisk*. However, contrastive studies have shown that the uses and meanings of the cognates overlap only partially: *faktisk* is less frequent than *in fact*, and more importantly, *in fact* is used predominantly as a connector and *faktisk* as an evidentiality marker ('in truth/reality'); cf. Hasselgård (2009: 257 ff) and Johansson (2007: 85 ff). The meanings of *in fact* correlate systematically with placement: the connector occurs predominantly in initial position, as in (14) and the evidentiality marker in medial position, as in (15), where the meaning of 'in reality' is predominant.

- (14) He repeats this like a child all the way through. *In fact* he is very much the child. (LOCNESS)
- (15) My final comment about Marx is that I *in fact* agree with him. It may sound like a paradox ... (ICLE-NO)

Faktisk does not show a similar correlation (Hasselgård 2009: 262); the evidentiality marker and the more bleached connective both typically occur medially (Hasselgård 2009: 260). Considering the differences between *in fact* and *faktisk*, Norwegian learners were expected to overuse *in fact* as an evidentiality marker, to overuse medial position for *in fact*, and to be unaware of the correlation between the meaning and position of *in fact*. It was indeed found that the Norwegian learners overuse the evidentiality marker. However, when *in fact* is used as a connector, it is placed in initial position. An apparent overuse of medial position for *in fact* in ICLE-NO is thus due to a slight overuse of the evidentiality meaning rather than to the wrong placement of the connector.

The French overuse of *in fact* along with (*as a*) *matter of fact* has often been commented on (see e.g. Granger & Tyson 1996: 22) and related to the more frequent French *en effet*. In the present material, the French overuse of *in fact* is not significant in relation to the number of times *fact* occurs (cf. Table 2), but it is highly significant relative to the number of words in ICLE-FR vs. LOCNESS ($\chi^2=19.9$, $p=0.000$). The

expression is used both as an evidentiality marker and a connector. In the latter function it can be semantically bleached, carrying practically no overtones of 'contrary to expectation' that was suggested by Oh (2000) as the core meaning of *in fact*; see (16).

- (16) As far as the military aspect is concerned we can see that the unification of the twelve nations will also be problematic. *In fact* there are different reasons accounting for this: (ICLE-FR)

4.1.3 It is a fact

The sequence *it is a fact* is frequent in ICLE-NO, but not in LOCNESS, cf. Table 2. The sequence is invariably followed by *that*, as shown in (17). Thus, like *the fact that*, this expression contains *fact* as an advance label with its lexicalization in a *that*-clause.

- (17) *It is a fact* that those who shout out loud get more attention. For centuries, women had been taught to keep quiet and to mind their own business, and those who first started to shout to get attention were first looked upon as a disgrace to their gender. (ICLE-NO)

A striking number of the *it is a fact that*-constructions occur paragraph-initially and are accompanied by some kind of contrast or comparison, as evidenced by (17). Incidentally, this contrastive feature is also present in the only example of the word combination in LOCNESS; cf. (18), which, however, is not paragraph-initial.

- (18) However, *it is a fact* that most of the recipients of welfare are white. (LOCNESS)

4.1.3 Phrase variability and learner problems

Both *the fact that* and *in fact* allow modification of *fact*. The BNC offers *the very/mere/simple fact that* and *in actual fact* as the most frequent variations. ICLE-NO and LOCNESS have three examples each of *the ADJ. fact that*, but there are no recurrent patterns (ICLE-NO has *cruel*,

scientific, simple and LOCNESS has *mere, only, very*).¹¹ *In fact* does not occur with modification in either ICLE-NO or LOCNESS. *It is a fact* occurs with an adverb after the verb; twice in ICLE-NO and once in LOCNESS (*obviously, also, still*). It also occurs five times in ICLE-NO and twice in LOCNESS with an adjective modifying *fact* (e.g. *hard, known, common, unfortunate, undeniable*); cf. (19).

- (19) *It is a known fact* that for most people, the biggest fear in life is the fear of death. (LOCNESS)

Norwegian learners have few problems with *in fact*. As regards *the fact that*, dissonant uses are mainly of the following types: (i) the shell noun does not label a ‘fact’, as in (12); (ii) *the fact* is superfluous, as in (6); (iii) *the fact that* is preceded by the wrong preposition, as in (2). Types (i) and (ii) occur in LOCNESS too, as shown by (13). *It is a fact* is overused by Norwegian learners, but there were no examples of dissonant use of *fact* as a shell noun in this construction.

4.2 Idea

Table 3 shows the distribution of recurrent combinations with *idea* across the corpora, selected according to the same criteria as those outlined for *fact* (see 4.1). It occurs in recurrent combinations most often in LOCNESS (76%) and least in ICLE-NO (57%). The patterns *the idea of* and *the idea that* are most frequent among native speakers, closely followed by the French learners, whose use of *idea* in general seems to be fairly close to the native speakers. The German and Norwegian learners underuse *idea* on the whole (see Figure 1), though ICLE-GE has more occurrences of *idea* as well as a higher proportion of recurrent combinations than ICLE-NO; in particular *the idea of* is more frequent. However, the Norwegian learners overuse *good idea* (relative to the total occurrences of *idea*), a combination shown in the BNC to be more frequent in speech than in writing.

¹¹ The variations on the recurrent combinations discussed here and in other sections on phrase variability were identified in separate searches using wildcards, e.g. <the * fact that>.

Table 3. Recurrent word combinations containing *idea* across corpora: raw frequencies and frequencies per 100,000 words.

IDEA	ICLE-NO (54)		ICLE-FR (122)		ICLE-GE (91)		LOCNESS (205)	
	raw	relative	raw	relative	raw	relative	raw	relative
<i>the idea of</i>	5	2.3	43	20.9	29	12.0	77	23.6
<i>this idea</i>	2	0.9	14	6.8	9	3.7	17	5.2
<i>the idea that</i>	5	2.3	9	4.4	4	1.7	22	6.7
<i>idea is</i>	7	3.3	3	1.5	5	2.1	15	4.6
<i>good idea</i>	8	3.7	6	2.9	6	2.5	9	2.8
<i>idea to</i> *	3	1.4	6	2.9	6	2.5	4	1.2
<i>no idea</i>	0	0	1	0.5	5	2.1	6	1.8
total	30	14.0	82	39.8	64	26.6	150	46.0
% in recurrent combinations	57.4		67.2		71.4		75.6	

* *Idea to* often overlaps with *good idea*.

The Norwegian underuse of *idea* is surprising in view of the existence of a Norwegian cognate (*idé*). However, searches in the ENPC show that *idea* is almost twice as frequent as *idé*, and moreover, that the cognates do not totally overlap in meaning. The fact that the lemma *idea* is translated into *idé* only 40% of the time, while *idé* is translated into *idea* 72% of the time, indicates that *idea* covers some meanings not shared by *idé*. The typical meaning of Norwegian *idé* is 'thought that you have about how to do something or how to deal with something' (Macmillan), which shows up in the most frequent cluster with *idea* in ICLE-NO, *good idea*. Other meanings of *idea* are 'information/knowledge', 'purpose/intention' and 'principle' (ibid.), which are present in Norwegian *idé* too, but typically belong to a relatively formal register. However, Norwegian learners do use them in the top four clusters in Table 3.

The patterns of *idea* in ICLE-GE are not significantly different from LOCNESS in spite of the general underuse of the noun. German has a cognate noun *Idee*, though searches in the English-German part of the Oslo Multilingual Corpus show that the two words do not have the same frequencies and distribution. In contrast to the Norwegian learners, the Germans have acquired *the idea of*, but they use *the idea that* as infrequently as the Norwegians.

4.2.1 The idea of

The idea of is the most frequent expression with *idea* in LOCNESS. *The idea of* functions with fairly equal frequencies as complement of preposition (20), subject (21) and verbal complement (object or predicative). *The idea of* something can for instance be addressed, attacked, believed in, discussed, endorsed, evoked, liked, preferred, rejected and supported. The prepositions preceding the cluster may be part of a prepositional verb or introduce a prepositional phrase, as in (20). Whether or not *idea* is a shell noun in this expression depends on its complement; a noun phrase complement, as in (20), cannot be said to lexicalize the content of *idea*, in contrast to a clausal complement, as in (22).

- (20) There seems also to be some ambiguity in *the idea of* innocence too. (LOCNESS)
- (21) *The idea of* a nuclear war is practically non-existent today. (LOCNESS)
- (22) ... some feminists focus on *the idea of* changing society into a more “womanly” one, ... (ICLE-NO)

While *the idea of* is underused by Norwegian learners, it is usually used correctly, as in (22). The only example of dissonance is found in (23), where the problem lies with the collocation of *fear* and *the idea of* rather than with *idea* itself.

- (23) Why doesn't criminals fear *the idea of* going to prison for several years. (ICLE-NO)

4.2.2 The idea that

Like *the idea of*, *the idea that* is most frequent in LOCNESS, but is also used by Norwegian learners. Syntactically, *the idea that* is also similar to *the idea of*, with a close to equal distribution between subject, complement of preposition and verbal complement in LOCNESS, while it takes subject function only once (out of 5) in ICLE-NO. As object, it most commonly follows verbs such as *develop*, *establish*, *come up with* or *point to*, *focus on*, see (24) Another, less frequent, group is made up by the verb phrases *stem from* and *be based on*.

- (24) Over the years society has established *the idea that* violence influences other modes of violence. (LOCNESS)

The Norwegian learners underuse *the idea that*, but they do use it correctly. The underuse may be partly related to the overuse of *the fact that*. Example (25) is one where *idea* might be a more fortunate choice of shell noun than *fact*.

- (25) In Norway we find some resistance against immigration. This is a contradiction to *the fact that* we support the human rights. (ICLE-NO)

4.2.3 This idea

This idea may function as a double marker of cohesion through the demonstrative reference of the determiner (Halliday & Hasan 1976: 57 ff) plus the retrospective labelling function provided by the (shell) noun (Francis 1994). This is demonstrated in (26), which is text-initial, and where *this* provides a referential link to the title of the essay ('Money is the root of all evil'); *idea* shows the writer's conceptualization of that proposition along with his/her explicit evaluation of it.

- (26) *This idea* is completely erroneous. (LOCNESS)

However, the cohesive link provided by *idea* may also consist in lexical repetition (Halliday & Hasan 1976: 278) rather than labelling, as exemplified by (27).

- (27) Most nations support *the idea that* everyone is born equal, and that there should not be ill treatment of people on any grounds; whether religious, racial, sexist or ethnic. *This idea* is also backed up by the nations legislation which prohibit discrimination, racism etc. (ICLE-NO)

4.2.4 Phrase variability and learner problems

The idea that, *this idea* and *the idea of* all allow modification of the noun. The only expression that was found to recur (twice in LOCNESS

and once in ICLE-NO and ICLE-GE) was *the whole idea of*, which is also the most frequent realisation of the pattern *the + ADJ + idea of* in the BNC. We may note the pattern ‘X’s/POSS DET *idea of*’, which is clearly related to *the idea of*. It occurred 4 times in ICLE-NO and 6 in LOCNESS and was thus too infrequent to be included in Table 3. Meanings of *idea* in these clusters are ‘principle’ and ‘understanding’.

- (28) Is keeping scared-to-death prisoners in coffin sized boxes *their idea of* humane convict treatment? (ICLE-NO)

There are few cases of dissonant labelling with *idea* in either ICLE-NO or LOCNESS. As mentioned above, the Norwegian learners’ underuse of *idea* may be partly due to the differences in frequency and semantic coverage of the cognates *idea* and *idé*. In the ENPC, *idea* was found to have a range of Norwegian correspondences. The most frequent nouns were *tanke* (‘thought’) and *anelse* (‘feeling’/‘hunch’), but interestingly correspondences with mental verbs such as *tenke* (‘think’) and *ane* (‘feel’/‘sense’) are also quite common. There are indeed some instances of *thought* in ICLE-NO where *idea* could have been used instead, e.g. (29). Furthermore, wildcard searches in ICLE-NO for patterns in which *idea* is used in LOCNESS (e.g. <support the * that/of>) suggested that Norwegian learners may be using *fact* and *statement* in contexts where *idea* would be a better choice; see (25) above and (30).

- (29) My guess is that it has to do with *the thought that* the more efficient the society is, the more time we will gain to do whatever it is that we are dreaming of doing. (ICLE-NO)
- (30) A totalitarian system of government could be said to support *the statement that* some are more equal than others. (ICLE-NO)

Interestingly, *statement* is greatly overused in ICLE-NO, with 57 occurrences per 100,000 words as against 17 in LOCNESS and similar frequencies in the other corpora. Norwegian learners use *statement* almost exclusively to refer to the essay prompt, i.e. the issue they are asked to discuss.

4.3 Question

Question can be a shell noun, but it can also refer to a concrete question being asked; sometimes to the essay question itself. Table 4 surveys the recurrent word combinations with *question* in the corpora. A first observation is that *question* occurs in recurrent phrases much less frequently than both *fact* and *idea* in all the corpora. LOCNESS has the highest proportion of *question* in recurrent combinations (50%), while the learner corpora have similar proportions of 43-44%. The most frequent combination overall is *the question of*. Note, however, that LOCNESS accounts for about half of its uses;¹² it is significantly underused ($p \leq 0.01$) in all the learner corpora, most clearly so in ICLE-NO and ICLE-GE. By contrast, *the question is* is overused in ICLE-NO and ICLE-GE. *This question* is overused in ICLE-FR, while frequencies in the other corpora are similar and well below that of ICLE-FR.

Table 4. Recurrent word combinations containing *question* across corpora: raw frequencies and frequencies per 100,000 words.

QUESTION	ICLE-NO (131)		ICLE-FR (149)		ICLE-GE (135)		LOCNESS (147)	
	raw	relative	raw	relative	raw	relative	raw	relative
<i>the question of</i>	14	6.5	17	8.2	9	3.7	41	12.6
<i>the question is</i>	17	7.9	10	4.8	17	7.1	5	1.5
<i>this question</i>	8	3.7	21	10.2	7	2.9	10	3.1
<i>a question of</i>	5	2.3	6	2.9	9	3.7	5	1.5
<i>question whether</i>	4	1.9	4	1.9	8	3.3	5	1.5
<i>in question</i>	5	2.3	2	1.0	2	0.8	6	1.8
<i>question that arises</i>	0	0	6	2.9	1	0.4	2	0.6
<i>question if</i>	4	1.9	0	0	5	2.1	0	0
total	57	26.6	66	32.0	58	24.1	74	22.7
% in recurrent combinations	43.5		44.3		43.0		50.3	

4.3.1 The question of and question if/whether

The question of can be followed by a noun phrase or a nominal clause, as in (31) and (32), respectively. When *the question of* is followed by a noun phrase, *question* is not a shell noun; i.e. the question is not

¹²LOCNESS has 12 instances of *the question of philosophical optimism*, probably reflecting an essay prompt.

lexicalized, but the labelling function may still be present, construing something as for example more debatable than an idea or less problematic than a problem.

- (31) Voltaire has tackled *the question of* philosophical optimism in a very successful way, in *Candide*. (LOCNESS)
- (32) For supporters of a single Europe *the question of* whether it will entail a loss of British sovereignty is not a primary issue. (LOCNESS)

The clauses lexicalizing the question are typically introduced by *whether*, which occurs ten times in LOCNESS; see (32), or by *what* and *where* (three occurrences in LOCNESS).

In ICLE-NO, *the question of* occurs before a *wh*-clause seven times (introduced by *how*, *what*, *whether* and *which*) and once erroneously before an indirect question introduced by *if*; see (33). The writer may have transferred the interchangeability of *if/whether* from the related expression *the question if (whether)*, shown in (34).

- (33) In my opinion, *the question of if* there is place enough for both science technology and imagination, I would say that the question is quite irrelevant. (ICLE-NO)
- (34) In *the question if* abortion can be both right and wrong, I would say that it depends. (ICLE-NO)

Question if occurs 4 times in ICLE-NO and 5 in ICLE-GE but is not used in ICLE-FR and LOCNESS, which seem to prefer *question whether*. Searches in the BNC show that the expression *question whether* has a distinct peak in academic prose, while *question if* is most frequent in spoken English; thus its use in the ICLE corpora shows the familiar influence of speech on learner writing (see e.g. Gilquin & Paquot 2008). Another difference, apparent from the concordances, is that *question* is a verb in all five cases of *question whether* in LOCNESS, but a noun in all four instances in ICLE-NO. The same applies to all instances of *question whether* in ICLE-GE and three out of the four occurrences in ICLE-FR. In LOCNESS, the noun *question* is not followed directly by *whether*, but instead has an intervening preposition in *the question of whether* (see above).

4.3.2 The question is and this question

The question is is far more frequent in the learner corpora than in LOCNESS, and more frequent in ICLE-NO than in the other learner corpora. The combination may refer to the essay prompt, as in (35).¹³ This is a metatextual function (i.e. the writer's comment on his/her text; cf. Ädel 2006). This function of *the question is* was found only in ICLE-NO. *The question is* may also be used rhetorically to preface a question posed by the writer, a function that is found both in ICLE-NO and LOCNESS. In (36) it contributes to text structure by marking a stage in a line of reasoning and also signalling the start of a problem-solution pattern (cf. Hoey 1983). *The question* functions as an advance (cataphoric) label (cf. Francis (1994) with the lexicalization of the shell noun in the predicative clause.

- (35) I also think *the question is* too extensive to simply answer yes or no. (ICLE-NO)
- (36) Mostly, we agree on the fact that people should be protected against criminal actions, *the question is*, however, how we can do that in a satisfactory way. (ICLE-NO)

The shell function of the noun can also be apparent in *this question*. In contrast to *the question is* (as well as *the question (of) whether*), *this question* functions as a retrospective (anaphoric) label; it typically follows a question that has been lexicalized in the text, as in (37). However, *this question* is also found to refer to the essay prompt in many cases in ICLE-NO, as shown in (38). Similar cases were found across the corpora, typically at the opening or end of the essay.

- (37) So who was the true number 1 and true national champion in the 1993-94 college football season, Florida State or Notre Dame. Again, the only way to answer *this question* fairly is to have a playoff system. (LOCNESS)
- (38) The subject of "Abortion - right and wrong" is a delicate and difficult matter that must be handled accordingly. You can get

¹³ The prompt was 'Most university degrees are theoretical and do not prepare students for the real world. They are therefore of very little value.' (<http://www.uclouvain.be/en-317607.html>)

professional help before and after your decision is made. But it can never completely heal the pain and scars left in your soul. Therefore, no one can ever answer *this question*. (ICLE-NO)

4.3.3 *Phrase variability and problems of use*

The BNC contains numerous examples of premodified *question* in the top four phrases in Table 4. However, the phrases do not show much variability in ICLE-NO or LOCNESS. *The question of* occurs with a premodifier twice in each corpus (*philosophical/whole* in ICLE-NO; *ethical/growing* in LOCNESS), while *this* PREMODIFIER *question* occurred twice in LOCNESS only (*this ethical/whole question*). ICLE-NO contained no variations on *the question is*, *this question* or *a question of*. LOCNESS gives one or two examples of each: *the real question is*; *this ethical/whole question*; and *a major question of*.

As a shell noun, *question* seems to be easier to handle for the learners than *fact*. The only example in ICLE-NO where the use of *question* was dissonant was (39), where *description* would be a better collocate of *fit*. However, the underused pattern *the* NOUN *of* seems to be a stylistic problem for the Norwegian learners; the corpus contains some stylistically awkward examples such as (40).

- (39) What kind of food is it so that results in a good and healthy breakfast? There is of course several provisions that fit *this question*. (ICLE-NO)
- (40) *The question of* equality has drawn more to *the question of* races the last decades. (ICLE-NO)

4.4 *Problem*

Problem was found to be significantly underused in ICLE-NO and ICLE-GE (cf. Figure 1), which may be surprising in view of the fact that a cognate word exists in both Norwegian and German. However, relative to the total frequency of *problem* in each corpus, most differences between learners and native speakers in the distribution of recurrent combinations are not significant, the exception being the overuse of *the problem is* in ICLE-FR. Table 5 shows that *problem* occurs in recurrent combinations between 39% and 47% of the time. Like *question* it is used

more frequently in recurrent combinations by native speakers than by learners.¹⁴ The pattern *the NOUN of* is frequent in LOCNESS, and equally so in ICLE-FR, no doubt inspired by the equivalent *le problème de*. Norwegian learners use this pattern least frequently, and the underuse is highly significant when calculated relative to corpus size ($p < 0.001$).

Table 5. Recurrent word combinations containing *problem* across corpora: raw frequencies and frequencies per 100,000 words.

PROBLEM	ICLE-NO (121)		ICLE-FR (179)		ICLE-GE (144)		LOCNESS (271)	
	raw	relative	raw	relative	raw	relative	raw	relative
<i>the problem is</i>	9	4.2	25	12.1	8	3.3	19	5.8
<i>the problem of</i>	6	2.8	15	7.3	11	4.6	22	6.7
<i>this problem</i>	9	4.2	6	2.9	16	6.6	21	6.4
<i>problem with</i>	5	2.3	5	2.4	4	1.7	17	5.2
<i>another problem</i>	4	1.9	9	4.4	3	1.2	12	3.7
<i>problem that</i>	3	1.4	4	1.9	4	1.7	14	4.3
<i>of the problem</i>	6	2.8	7	3.4	6	2.5	4	1.2
<i>main problem</i>	3	1.4	5	2.4	0	0	5	1.5
<i>major problem</i>	0	0	4	1.9	0	0	9	2.8
<i>big problem</i>	2	0.9	3	1.5	7	2.9	0	0
<i>solution to the problem</i>	0	0	0	0	5	2.1	5	1.5
total	47	22.0	83	40.3	64	26.6	128	39.3
% in recurrent combination	38.8		46.4		44.4		47.2	

Most of the recurrent combinations with *problem* are not frequent enough to show clear patterns. We may, however, note *problem that*, which is more frequent in LOCNESS than in the learner corpora. In most cases this word combination is part of the pattern *the NOUN that*, which is generally disfavoured by learners. The slightly dissonant *big problem* is recurrent chiefly in ICLE-GE. It does not occur in LOCNESS (which instead has *major problem*), and would not normally be considered an elegant collocation in academic writing. In the BNC it occurs predominantly in speech and very rarely in the written registers.

¹⁴ LOCNESS had 14 instances of *social problem*, but this was not included in the study as it seemed to be related to a specific essay topic and was also scarce in the learner corpora (1 in ICLE-NO and 2 in ICLE-GE).

4.4.1 The problem is

The problem is can be a shell noun signalling a problem-solution pattern and preceding its lexicalization. In LOCNESS *the problem* typically functions as a subject, as in (41), but also as the complement of a preposition in an extended noun phrase, such as *the solution to the problem* or *the extent of the problem*. With one single exception, *is* functions as the main verb in this sequence in LOCNESS. The predicatives are realized by clauses in 11 cases (6 *that*-clauses, 2 infinitive clauses, 2 *wh*-clauses, and one *ing* participle), noun phrases and adjective phrases three times each. In one case the predicative is a deleted quotation.

- (41) As stated, *the problem is* how these two desires are to be reconciled ... (LOCNESS)
(42) *The problem is* that the word “feminism” has a number of negative connotations. (ICLE-NO)

In ICLE-NO *the problem is* constitutes subject and (main) verb in all nine cases. It is followed by a clause in seven cases (six *that*-clauses, as in (42), and one infinitive clause), and an adjective phrase in two. Clausal predicatives thus dominate in both corpora, but the native speakers use a greater variety of clause types. There were no examples of *the* NOUN PREP *the problem is* in ICLE-NO or ICLE-FR, and only one in ICLE-GE).

4.4.2 The problem of

The problem of differs markedly in frequency between LOCNESS and ICLE-NO. Interestingly, it also differs markedly in the lexical and syntactic patterns it enters into. In LOCNESS, *the problem of X* is clause subject in seven cases, notional subject in existential clauses in three, object of transitive verbs in 11 (*address, ease, examine, face, make, solve, tackle, understand*), and prepositional complement in one. In ICLE-NO it functions as notional subject in an existential clause once, object twice (*face, avoid*), and prepositional complement three times; see (43). Two transitive verbs take *the problem of ...* as object more than once in LOCNESS, namely *solve* and *tackle*, exemplified by (44).

- (43) In addition to *the problem of* overcrowding, there is a lot of abuse. (ICLE-NO)
- (44) ... Voltaire tackles *the problem of* thoughtless optimism. (LOCNESS)

In contrast to *the idea of*, *the problem of* is invariably followed by noun phrase complements. The noun phrases chiefly denote phenomena that would normally be regarded as negative anyway, as in (43). However, the expression may also signal the writer's negative evaluation of something, as in (45).

- (45) It is obvious that Mr Gingrich does not understand *the problem of* Welfare Reform at all. (LOCNESS)

4.4.3 This problem

LOCNESS and ICLE-NO alike use *this problem* predominantly as object or prepositional complement. *This problem* thus typically functions as a retrospective label and also contributes to cohesion through demonstrative reference (Halliday & Hasan 1976: 57 ff). In LOCNESS *this problem* is most commonly the object of a verb referring to a way of dealing with it: *combat, curb, eliminate, solve, get round, put an end to, and deal with*, or a way of relating to it: *examine, look at*. *Solve* and *deal with* are also found in ICLE-NO, together with *do something with*; see (46).

- (46) ...but today psychiatrists and other experts seem to deal with *this problem* in another way. (ICLE-NO)
- (47) *This problem* could easily be curtailed by lowering the drinking age from twenty-one to eighteen. (LOCNESS)
- (48) *This problem* I have met in several subjects at the different colleges I have attended. (ICLE-NO)

This problem also functions as subject (in four out of 21 occurrences in LOCNESS and one out of nine in ICLE-NO). Three of the sentences in LOCNESS with *this problem* as subject are passive constructions; see (47). Such passive constructions are not found in ICLE-NO, but there is a case of *this problem* as a fronted object (48). While *this problem*

functions adequately as a retrospective label in (48), the word order signals a contrast which was probably not intended by the writer, and which may be due to transfer of the somewhat freer word order of Norwegian.

4.4.4 Phrase variability and learner problems

All the combinations discussed in this section allow premodification of *problem*. *The problem is* occurs with a premodifier three times in LOCNESS (*only* and *other*) and four in ICLE-NO (*biggest*, *major*, *only*, *other*). *The problem of* has an intervening adjective only in LOCNESS (seven times); the adjectives are *common*, *major*, *mounting*, *perpetual*, and *social*. ICLE-NO has two examples of *this* + ADJ. + *problem* (*complex* and *particular*), while LOCNESS only has one (*this same problem*).

The use of *problem* as a label or a shell noun does not seem difficult for learners; no cases of dissonant labelling were found. Any ‘foreign accent’ in the phraseology of *problem* in ICLE-NO is rather caused by the differences in overall frequencies of some constructions and in the lexical and syntactic environments of the combinations, as outlined above.

4.5 Issue

As was shown in Figure 1 above, *issue* is underused by all learner groups, and recurrent patterns are therefore scarce. The frequencies are too low for significance testing to be meaningful: Table 6 shows that recurrent combinations with *issue* are frequent only in LOCNESS, and notably quite absent from ICLE-GE.

Table 6. Recurrent word combinations containing *issue* across corpora: raw frequencies and frequencies per 100,000 words.

ISSUE	ICLE-NO (58)		ICLE-FR (22)		ICLE-GE (10)		LOCNESS (157)	
	raw	relative	raw	relative	raw	relative	raw	relative
<i>this issue</i>	7	3.3	5	2.4	0	0	25	7.7
<i>the issue of</i>	6	2.8	2	1.0	0	0	26	8.0
<i>of the issue</i>	2	0.9	0	0	0	0	12	3.7
<i>important issue</i>	7	3.3	1	0.5	0	0	4	1.2
<i>issue that</i>	1	0.5	0	0	0	0	7	2.1
total	23	10.8	8	3.9	0	0	74	22.7
% in recurrent combinations	39.7		36.4		0		47.1	

The most frequent phrases in LOCNESS are *this issue*, *the issue of* and *of the issue*. The latter two overlap in (49):

- (49) The Ethnic American Authors' addressing *of the issue of* self understanding. (LOCNESS)

It may be noted that the combination *issue that* in LOCNESS does not reflect the pattern *the NOUN that*; *that* is a relative pronoun in this combination and thus does not preface a lexicalization of the noun. (*Issue* followed by a nominal *that*-clause providing a lexicalization was, however, found in the BNC.) In addition to the two patterns discussed below, ICLE-NO has seven instances of *important issue*. Four of them are preceded by *an* or *one*, and thus resemble the only pattern that can be identified in ICLE-GE, namely *a(n) ADJ. issue*.

4.5.1 The issue of

In LOCNESS, *the issue of* is often part of a subject noun phrase, either clause-initially or as notional subject in an existential clause. Alternatively it is the object of the same type of verb that tends to precede *problem*: *address*, *attack*, *bring up*, *confront*, *discuss*, *tackle*, and *relate*. Occurrences in ICLE-NO do not reveal any patterns, but it may be noted some of the examples reveal usage problems; see (50) and (51).

- (50) One could question *the issue of* whether nations really need infantry, or foot soldiers at all. (ICLE-NO)

- (51) Another aspect which may seem more and more important is *the issue of prevention*. (ICLE-NO)

The collocation of the verb *question* with the object *issue* in (50) is unfortunate; the sentence might be improved by replacing *question* with *discuss* or simply omitting *the issue of*. In example (51) the word *issue* is used correctly; however, the sentence is clumsy because the writer has used *aspect* and *issue* synonymously. The example shows the verbosity described by Granger (1998) as typical of learner style and would benefit from some pruning, e.g. *The issue of prevention may seem more and more important*.¹⁵

4.5.2 This issue

This issue can function as a retrospective label. In LOCNESS it is typically an object following verbs such as *address*, *discuss* and *surround*, as in (52), as well as prepositions in phrases like *part/side of this issue*. It functions as subject only once. Again ICLE-NO has too few examples to reveal patterns, but there are dissonant uses, as in (53).

- (52) There are numerous debates surrounding *this issue* leaving people to wonder, what is the right thing to do? (LOCNESS)
(53) Lastly *this issue* leads us to yet another argument against the prison system... (ICLE-NO)

Example (53) contains a clear attempt at creating text structure, but it is unclear what *this issue* refers to. The sentence is paragraph-initial, and what is described in the preceding context seems to be a *situation* rather than an *issue*.

4.5.3 Phrase variability and learner problems

The BNC contains examples of noun modification in *this issue*, e.g. *this important/ particular/whole issue*, but the phrase does not show any

¹⁵ Incidentally, the writer probably means *contraception* rather than *prevention* here, and has stumbled over a false friend (Norwegian *prevensjon* = 'contraception').

variability in either LOCNESS or ICLE-NO. *The issue of* occurs in the BNC with premodifiers of *issue* denoting importance, complexity, difficulty or specificity (e.g. *central, complex, difficult, thorny, particular, whole*). The two examples of an extended phrase in LOCNESS reflect this tendency: *the whole/thorny issue of*. ICLE-NO does not have any variation of the phrase.

The underuse of *issue* in ICLE-NO, along with a relatively large proportion of dissonant examples, shows that *issue* is not well-established in the vocabulary of most Norwegian learners. The learners seem to have trouble with the semantics as well as the pragmatics of *issue*. The learning problem seems to be widespread, as *issue* is one of the words discussed in the 'Improve your writing skills' section in the *Macmillan English Dictionary*: "If you want to present the topic as an important subject that people discuss and have opposing views about, use the nouns *issue* or *question*." (*Macmillan* 2007: IW21) Learners are also advised on how to avoid confusing *problem* and *issue*. While Norwegian learners seem to have little trouble using *question* and *problem*, there is at least one example where *issue* has been used in lieu of *question*; see (54).

(54) ... *the issue whether* abortion is right or wrong has turned into a great discussion. (ICLE-NO)

Moreover, a search for contexts typical of *issue* showed that Norwegian learners sometimes use *aspect* instead, as shown in (55). The sentence is paragraph-initial and brings up revenge as a topic for discussion; precisely the type of context where native speakers use *issue*.

(55) Then there is *the aspect of* revenge. (ICLE-NO)

As mentioned above, Norwegian does not have a direct equivalent of *issue*, which will make it difficult for Norwegian learners to conceptualize the term. In the ENPC *issue* is translated by *spørsmål* ('question'), *problem*, and *tema* ('topic'). Thus, some of the instances of *question* and *problem* could probably be replaced by *issue*, for instance in (56).

50 Hilde Hasselgård

- (56) In addition to the short sighted and politically motivated slant in favor of “irrelevant” studies, there is *the problem of* the actual content of higher education. (ICLE-NO)
- (57) *The issue of* the open market therefore continues to be problematical ... (LOCNESS)
- (58) The [*issue of*] ‘everyday-racism’ is very much in the spotlight in Norway these days. (ICLE-NO)

Another reason why learners underuse *the issue of* may be that it is often syntactically omissible, as in (57), which would be grammatical without it. However, what is lost by such omission is the rhetorical function of flagging a topic as up for discussion. Example (58) is one that might be improved by such a rhetorical use of *issue*, as indicated in brackets.

5 Concluding remarks

This paper set out to explore the use of the nouns *fact*, *idea*, *question*, *problem* and *issue* and the ways they habitually combine with other words in native English and three varieties of learner English. The recurrent word combinations in ICLE-NO and LOCNESS received special attention. The nouns differ markedly in frequency across the corpora as does their tendency to occur in recurrent word combinations. As shown in section 3, most of these nouns tend to be underused by most of the learner groups; the exceptions are *question* and the frequent use of *fact* and *problem* in ICLE-FR. The noun that is most markedly underused by all learner groups is *issue*. Clearly, in a study of learner language, quantitative observations need to be supplemented with qualitative analysis. Closer scrutiny thus revealed that Norwegian learners sometimes misuse this word. A possible reason for the underuse, besides the lack of an equivalent Norwegian word, might be that the function of *issue* is mainly rhetorical; i.e. signalling a topic for discussion.

All the learner corpora contained examples of these nouns used as shell nouns. Norwegian learners were shown to have problems with *issue* in this function, but also with *idea*, due to semantic differences from the Norwegian cognate. The expression *the fact that* deserves special mention. All the corpora, including LOCNESS, had examples of *fact* labelling propositions that would not normally be considered facts.

Similar uses were noted by Schmid (2000). This indicates that *the fact that* may be on its way to becoming an extended conjunction that helps accommodate a *that*-clause in nominal positions. Even so, Norwegian learners seem to exaggerate the need for *the fact* as a preface to *that*-clauses, and moreover, they may be unaware of more appropriate alternatives to *fact* to label non-facts.

Tables 2–6 show the way and extent to which the nouns occur in recurrent combinations across the corpora. An interesting observation is that the percentage of the time each noun occurs in recurrent combinations is almost consistently higher in LOCNESS than in the learner corpora. This seems to indicate a higher degree of routinization of the phrases among native speakers. Of the learner corpora, ICLE-FR has the highest proportion of recurrent word combinations. The percentage is generally lowest in ICLE-NO, but ICLE-GE has lower proportions of recurrent word combinations with *question* and *issue*.

The pattern where native speakers differ most markedly from learners is *the NOUN of/that*. This pattern belongs to syntactically complex phrases, which may be a reason why learners underuse it (disregarding *the fact that*). French learners, however, use the pattern more than German and Norwegian learners, possibly due to the frequent use of similar constructions in French (e.g. *l'idée de/que*). Simpler combinations are more popular with the learners, such as *the NOUN is* and *this NOUN*. The question of phrase complexity in learner language must, however, await further study. Another question worthy of further investigation concerns the extent to which the use of shell nouns depends on writing experience as well as language proficiency. Since both LOCNESS and ICLE represent novice writing, it would be interesting to compare the results of the present study to more skilled writing, such as press editorials or published academic papers.

This paper has shown that Norwegian learners use most of the nouns investigated in a different manner from native speakers. The learners do not seem fully aware of the semantics and pragmatics of *idea* and *issue*, which leads to underuse as well as misuse. However, even with words that are more firmly established in their vocabulary, they tend to prefer simple patterns, in particular avoiding *the NOUN of/that*. Learners could usefully be made aware of the rhetorical and text-structuring potential of phrases involving shell nouns. Moreover, some focus on syntactically

complex phrases would bring the learners further from the stylistic ideals of Norwegian and closer to a style that is valued in academic English.

References

- Ädel, A. 2006. *Metadiscourse in L1 and L2 English*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Aktas, R.N. & V. Cortes. 2008. "Shell nouns as cohesive devices in published and ESL student writing." *Journal of English for Academic Purposes* Vol 7, 3-14.
- Altenberg, B. 1998. "On the phraseology of spoken English" in Cowie (ed.), 101-122.
- Cowie, A.P. (ed.) 1998. *Phraseology. Theory, Analysis, and Applications*. Oxford: Oxford University Press.
- Ellis, N. 2008. "Phraseology: The periphery and the heart of language" in Meunier, F. & S. Granger (eds), *Phraseology in Foreign Language Learning and Teaching*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1-14.
- Firth, J.R. 1957. *Papers in Linguistics 1934-1951*. Oxford: Oxford University Press.
- Flowerdew, J. 2006. "Use of signalling nouns in a learner corpus." *International Journal of Corpus Linguistics* 11 (3), 345-362.
- Francis, G. 1994. "Labelling discourse: an aspect of nominal-group lexical cohesion" in Coulthard, M. (ed.) *Advances in Written Text Analysis*. London: Routledge, 83-101.
- Gilquin, G. & Paquot, M. 2008. "Too chatty: Learner academic writing and register variation." *English Text Construction* 1(1): 41-61.
- Gläser, R. 1998. "The stylistic potential of phraseological units in the light of genre analysis" in Cowie (ed.).
- Granger, S. 1998. "Prefabricated patterns in EFL writing" in Cowie (ed.), 145-160.
- Granger, S. and S. Tyson. 1996. "Connector usage in the English essay writing of native and non-native EFL speakers of English." *World Englishes* 15, 19-29.
- Halliday, M.A.K. and C.M.I.M. Matthiessen. 2004. *An Introduction to Functional Grammar*. 3rd edition. London: Arnold.
- Hasselgård, H. 2009. "A textual perspective on the pragmatic markers *in fact* and *faktisk*" in Slembrouck, S., M. Taverniers, and M. Van

- Herreweghe (eds.) *From Will to Well: Studies in Linguistics offered to Anne-Marie Simon-Vandenberghe*. Ghent: Academia Press, 257-266.
- Hasselgård, H., S. Johansson and P. Lysvåg. 1998. *English Grammar: Theory and Use*. Oslo: Universitetsforlaget.
- 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: 237-259.
- Hoey, M. 1983. *On the Surface of Discourse*. London: George Allen & Unwin.
- Hunston, S. & G. Francis. 1999. *Pattern Grammar. A Corpus-driven Approach to the Lexical Grammar of English*. Amsterdam: Benjamins.
- Johansson, S. 2007. *Seeing through Multilingual Corpora*. Amsterdam/New York: John Benjamins.
- Macmillan English Dictionary for Advanced Learners*, 2nd edition. 2007.
- Nesselhauf, N. 2005. *Collocations in a Learner Corpus*. Amsterdam: Benjamins.
- Oh, S.-Y. 2000. "Actually and in fact in American English: a data-based analysis." *English Language and Linguistics* 4(2): 243-268.
- Paquot, M. 2010. *Academic Vocabulary in Learner Writing. From Extraction to Analysis*. London: Continuum.
- Pawley, A. and F. H. Syder. 1983. "Two puzzles for linguistic theory: nativelike selection and nativelike fluency" in Richards, S.C. and R.W. Schmidt (eds). *Language and Communication*. London & New York: Longman, 191-226.
- Schmid, H.-J. 2000. *English Abstract Nouns as Conceptual Shells: From Corpus to Cognition*. Berlin: Mouton de Gruyter.

Corpora used

Oslo Multilingual Corpus (OMC):

www.hf.uio.no/ilos/english/services/omc/

The International Corpus of Learner English (ICLE):

www.uclouvain.be/en-cecl-icle.html

The Louvain Corpus of Native English Essays (LOCNESS):

www.uclouvain.be/en-cecl-locness.html

The British National Corpus (BNC): www.natcorp.ox.ac.uk/ and

<http://corpus.byu.edu/bnc/>

The English-Norwegian Parallel Corpus (ENPC):

www.hf.uio.no/ilos/english/services/omc/enpc/