

Phraseology in a Learner Corpus Compared with the Phraseology of UK and US Students

Yoshihito Kamakura¹

1. Introduction

Describing learner language is a primary objective and a most important approach to the study of second language acquisition (Ellis 1994). Quite a few attempts to describe learner language have been made so far by researchers. For instance, error analysis, which was in its heyday in the 1970s, is employed to assess whether an output by a learner is correct or erroneous according to the intuition of native speakers. In comparison to the productions of native speakers, the infrequent written and spoken outputs by learners are likely to be identified as ‘foreign’ in their use of the language; such atypical production, although conforming to the norms of prescriptive grammar, are liable to be considered erroneous. It is, therefore, implied that the judgment of part of a learner’s sentence, such as the use of prepositions, cannot describe the productions of learners in detail: an assessment which focuses merely on prepositions is unlikely to be adequate and it is better to adopt collective observations of phrases and sentences which learners produce. Corpora are eligible for collective comparison in terms of the frequency of given words and phrase, the internal and external structures of phrases and the composition of sentences containing key words.

Corpora are capable of supplying a comprehensive description of language (Sinclair 1991, Hunston 2002, Partington 1998, Biber *et al.* 1999, Kennedy 1998)². The large amount of storage of text gives enough resources to shed light on remarkable aspects of language. Corpora make it easier to compare learners’ and native speakers’ types of text, in terms of the frequency of words and phrases (Ringbom 1998, Lorenz 1998, De Cock *et al.* 1998, Altenberg and Tapper 1998, Virtanen 1998, Petch-Tyson 1998). Concordance lines extracted from a corpus are suitable for comparing sentences which include the same key word and phrase. The combination of words, the environment of a given word and phrase and the position

¹ Department of English, University of Birmingham
e-mail: yxk343@bham.ac.uk

² Hunston (2002) points out that corpora by themselves are not able to show novel features of language, but the software which helps to show frequency, phraseology and collocation enables us to identify novel features.

of a particular phrase can then be observed by comparing the two corpora. Hence, the purpose of this present study is to make a comparative analysis of two corpora, one compiled from essays written by Japanese learners of English and one from a corpus of British and American university students' essays³.

This study focuses mainly on a comparison of two-, three- and four-word clusters containing the prepositions *IN*, *ON*, *AT* from the learners' and native speakers' corpora. It comprises five chapters. This first chapter introduces the questions to be answered in the course of the study. The second chapter gives an account of the data of the present study. In the third chapter, work related to this study on learner corpora is presented. In the main chapter, Chapter Four, appears the accounts of multi-word units containing the given prepositions according to the hypotheses formulated by the analysis. To close the present study, its conclusions are discussed in Chapter Five.

2. Data Description

2.1 Japanese Learner Corpus: Showa Women's University Corpus

This Japanese learner corpus has been provided by Professor Kaneko, of Showa Women's University. The collection of the data virtually emerged as a project undertaken for the International Corpus of Learner English (henceforth ICLE). Meanwhile, another set of Japanese learner corpora, which is the principal data of this research, was being compiled by Professor Kaneko for her own interest. With regard to the shared features, year three and four university students, whose speciality of study is English language, mainly contribute their writings.

In terms of the variable features, task settings seem to vary from learner to learner: some of the writings took place in a timed setting, but not all. The bulk of the learners preferred to consult a dictionary when writing, but some chose not to. Writing topics were selected from the ones, which were provided to the learners, while some wrote on their original topic. Given the diversity of these attributes, the topics are recorded in the corpus profile.

³ Comparable corpora are defined as two or more than two corpora composed of different varieties of a language (e.g. Indian English and Canadian English) in pursuit of the comparison of varieties on the same language (Hunston 2002: 15) In the present study, the varieties of English – learners' English and native speakers' English – are being compared in order to identify the differences between the two varieties.

2.2 Native English Speakers' Essay Corpus: the Louvain Corpus of Native English Essays (LOCNESS)

The LOCNESS corpus has been compiled for the purpose of contrasting it with the ICLE learner corpora; the design criteria are established to comply appropriately with the ICLE ones. The corpus comprises essays written by British and American students, namely British pupils' A-level essays, British university students' essays, and American university students' essays.

Essay topics retain a wide variety: from social issues, which are overwhelmingly chosen by British A-level essays, to literary or sociological subjects, which the American university students wrote for assessment. Despite the fact that English is their first language, it is stated that reference tools, such as dictionaries, were sometimes used when writing. Like the Japanese learner corpus, it is likely that task settings vary with each learner.

Essay writer	Corpus name	National subcorpus	Number of essays	Number of words
Japanese learners of English (JP: Japanese)	Showa Women's University corpus	Japanese	175	85,098
Native English speakers (NS:native speaker)	Louvain Corpus of Native English Essays (LOCNESS)	English American	436	324,304

Table 1: Sizes and distribution of analysed corpora

3. Previous Research on Phraseology in Learner Corpora

The research into learner corpora conducted in the 1990s overwhelmingly concentrated on quantitative differences between native speakers' texts and those of learners (Ringbom, 1998), although the studies of learner corpora can also be approached qualitatively. Biber and Reppen (1998), comparing the frequency of complement clauses following a verb in the learner corpora of four different L1 groups from that in native speakers' English, have found with high frequency verbs that they co-occur with *that*-and *to*-clauses and that these occur more frequently in learner texts than in native ones. Like the studies of Ringbom (1998) and Biber and Reppen (1998), it is word frequencies in the main which draw the particular attention of learner corpus researchers (de Haan 1997, De Cock 1998, De Cock *et al.* 1998, Virtanen 1998, Patch-Tyson 1998, Granger and Rayson 1998, Belz 2004, Pulcini and

Furiassi 2004), since they are to some extent easily observable and can be compared by analysing computerised data with the help of software tools or a computer program (Nesselhauf 2004, Myles 2005). Such a ‘frequency based approach’ (Nesselhauf 2005) seems to exploit one advantage of learner corpora. More importantly, the ‘phraseological approach’ is likely to make good use of learner corpora and shed light on the features of learner language which are not revealed by conventional research approaches. Technically, by means of observing co-occurring words and phrases along with a key word (e.g., a word which may frequently precede or follow another word or phrase), one may identify the lexico-grammatical patterning (Flowerdew 1998) of learner language.

Drawing on the classification of collocational occurrence observed in native speaker productions by Howarth (1998), Nesselhauf (2003, 2005) examines the combinations of verb and complement nouns in English essays by German-speaking learners and finds that restrictedly combined ‘idioms’ and unrestricted ‘free combinations’ (Howarth 1998: 167, Nesselhauf 2003: 226) turn out to be less deviant and less problematic for learners. Consequently, given the interrelation of inter- and intralingual difficulties, Nesselhauf concludes that the congruency⁴ between learners’ L1 and the target L2 and the degree of restrictedness of the L2 word combinations (not too restricted and too free) are probably ascribable to deviant word combinations by learners (Nesselhauf 2005: 238).

Flowerdew (1998, 2003) argues that most corpus research in the 1980s focused on lexico-grammatical patterning and provided pedagogical insights on the basis of the findings of this patterning. She then suggests that learner corpora can be applied to the idiosyncratic discourse accounts of learner outputs (Flowerdew 1998). Analysing the problem-solution rhetorical patterns in academic writing by undergraduate students in Hong Kong, Flowerdew reveals that there are some patterns in native-speakers’ academic writing: e.g., *problem* followed by an implicit causative verb has a positive semantic prosody (Sinclair 2004), e.g. ...*will minimise much of the problem*, whilst explicit causative verbs retain a negative semantic prosody, e.g. ...*will create a noise problem* (Flowerdew 2003: 500). In addition, some distinctive phraseological patterns are suggested as patterns: notably, ‘BE a problem’ and ‘there BE a problem’ occur frequently in the native corpus, while the co-occurrence of *solution(s)* and *problem* in the same phrase, as in ‘BE a solution to this problem’, appears to be frequent in the learner corpus (*ibid*: 502). Given that learner corpus research lends itself to classifying and correcting an erroneous word (for instance, ‘error tags’ in Dangneaux *et al.* 1998, Granger 2002, 2003), learners’

⁴ Congruency is defined as ‘word-for-word equivalence of a collocation in the learner’s L1 and the L2’ (Nesselhauf 2005: 238).

insufficiency in lexico-grammar is best solved by providing pedagogical indications on the basis of findings of learner corpus research (Flowerdew 2003: 508).

4. Hypotheses and Frequent Sequences with Prepositions *IN*, *ON* and *AT*

This chapter will focus mainly on the frequent phrases which co-occur with preposition *in*. The phrasal sequence will be examined in extracts from N-grams, which consist of two to four words. Having examined the frequent clusters, the outline of the analysis has emerged: the following statements are key arguments of the study:

- i) *On the whole, the same N-grams are frequent in both corpora, but they tend to occur with different co-texts in each corpus.*
- ii) *There are a number of clusters occurring persistently in particular positions in a clause. In particular, the Japanese learners tend to restrict certain clusters to certain position.*
- iii) *Frequent clusters can be classified into flexible vs. consistent sequences, in terms of their internal structure.*
- iv) *There are some frequent longer clusters composed of frequent shorter clusters. The short N-grams gradually extend the string boundary from two- to three- and then three- to four-word clusters.*
- v) *Frequent long clusters can be divided into two types – topic-specific and topic-unspecific – as regards their relationship to the context of the essay. The emergence of the topic-specific ones highly correlates to the topic of the essay.*

In the frequent cluster table of JPC and NSC, identical clusters are found; however, it is interesting that the surroundings of the cluster appear to differ between the two corpora, indicating that the usage of an identical cluster may show a distinctive interpretation of the phrase by JP learners, and imply a peculiar ‘avoidance’ of a grammar attribute. Some clusters tend to occur in a peculiar position, such as clause-initial or clause-final. This tendency is shown eminently in the JPC. Frequent clusters fall into two types in internal sequences: flexible and consistent. Flexible clusters show a minor variation in phrase, for instance, by inserting an adjective before a noun in a prepositional phrase or by elaborating a complex grammar structure. It turns out that inflexible clusters are likely to occur frequently in the JPC. The development of two- to four-word clusters illustrates the finding that

JP's shorter clusters, bi-grams and trigrams, consistently occur as part of four-word sequences, which implies that a given four-word sequence consistently occurs in uniform phrases. This corresponds to the finding that JP are inclined to use consistent sequences. When the four-word clusters are decoded, they appear to be classifiable into either 'topic-specific' or else 'topic-unspecific': the former clusters are exclusively used in essays on a particular topic, while the latter occur frequently in those on any general topic. Taking these key points in the argument into account, we move on to examine individual multi-word clusters.

4.1 Accounts of hypothesis I

'IN the world' and 'IN the future' frequently emerge in both JPC and NSC; however, the environments preceding or following the trigrams differ slightly in lexical and grammatical terms. In the JP trigrams '*language*' recurrently precedes *IN*, as in '*common language in the world*', '*important language in the world*', '*second language in the world*'. This may be because the topics for JP essays in studying a second language may involve '*language*'. Regarding the connotations of '*IN the world*', the JP invariably expresses 'commonality' in all the countries and people, as in '*... it [English] is used in common in the world by other countries*' and '*So most people in the world learn English*' (my underlining throughout). However, the NS also uses the trigram as an intensifier of the preceding noun or a part of a superlative clause and also the same connotation as the JP trigram: for example, '*... including our travel, almost everywhere in the world*', '*... a tradition in UK and anywhere in the world*'. Despite not being entirely clear about the line between commonality and intensifier, the NS seem to use the trigram in order to emphasise the preceding noun in completeness, as in '*Everything in the world appears to be given ...*' and '*Nothing in the world is necessarily good or bad*'. The trigram also often recurs with superlative adjectives or phrases composed of a superlative adjective and noun, as in '*It is this desire to be the best in the world*', and '*... and the Queen, one of the richest women in the world*' .

lish is thought as it is used in common in the world by other countries. The
me popular. We can access to everywhere in the world by using a personal comput
age? Today English is a common language in the world. Many people speak it as
ubject. So English is a common language in the world. Moreover, computer have
y. English is used as a common language in the world if we can speak English
lish? Because it is important language. In the world almost all country seems
ill become most important language in the world. And also I heard m
meday English will be a second language in the world, by that time we should
Now, English is spoken by many people in the world as a measure of commun
fact that English is spoken many people in the world. English is the most ba
ommon language of world. So most people in the world learn English. Secondly
s, and communicate with a lot of people in the world. But the most important

Figure 1a: IN the world in the JPC

, including our travel, almost anywhere in the world. With the advent of the
chy is a a tradition in UK and anywhere in the world, UK is known by its royal
far would you go to be the best athlete in the world? Some people would go t
11 . It is this desire to be the best in the world that drives some athlet
18 well loved sport amongst most countries in the world however it is also punis
19 which is recognised by other countries in the world. This government makes a
25 could everything be false? Everything in the world appears to be given and
39 he Procurante would be the happiest man in the world but not until Martin tells
42re also rejects this attitude. Nothing in the world is necessarily good or
49 all the way to Beijing -- every person in the world can watch this event as
60 and for her T.V. is the greatest thing in the world. She sits in front of it
64 and the Queen, one of the richest women in the world, did not pay tax. The r

Figure 1b: IN the world in the NSC

2 people but we must grow ourselves. And in the future we can work all over
3 I would like to be a flight attendant in the future. So now I learning Eng
5 that many students have no dream to be in the future. Thus they don't know
8 I understand English without caption. In the future, I want to be tour Gui
14 So everyone should to learn English. In the future children will learn E
17 University. But I need to speak English in the future. So I want to be able
29 everyone say "English is very Important. In the future, you can't work without
33 to master English as a second language. In the future if we don't know or c
38 to person's contact with each mind. In the future, we need to use Engl
44 e to be more good international person. In the future I want to go abroad b

Figure 2a: IN the future in the JPC

1 lk, which means there will be less cows in the future possibly. If cows are
2 stance to draught, pests and disease. In the future it will be possible t
3 ances of being crippled by this disease in the future. If a lot of people d
8 vorce will never be gone, but hopefully in the future there will stronger m
11 to see how it might be able to lose it in the future. Sovereignty, origina
16 t mean that they will make that mistake in the future. The past is past. Le 17
prevent injuries such as these occurring in the future and they feel the onl
20 onviction that military conflict should in the future be avoided; after all
25 ey knew that profit would come to them in the future after a good relation
27 re new problems will cause great trauma in the future. I agree, that genera

Figure 2b: IN the future in the NSC

The second most frequent trigram in the JPC, '*IN the future*', to some extent tends to locate at the sentence-initial position, with 16 instances out of 52 (30.8%), in contrast to 4 instances out of 30 in the NSC (13.3%). The surroundings of the trigram in the JPC represent a wish or need, such as '*want to*', '*need to*' and '*must*'; alternatively, they indicate the possibility of being able to do something (Figure 2a). The same sequence in the NSC shows a distinctive feature in co-occurring words: negative and troublesome words predominately emerge in the clause, for instance, '*pests*', '*disease*', '*mistake*', '*conflict*', '*trauma*', with some exceptions such as '*hopefully*' (Figure 2b). Lastly, a frequent trigram in the JPC '*IN the first*' is frequently found in the sequence of '*in the first place*' and '*in the first year*'. It could, then, be inferred that JPs are likely to account for their prospects by expressing their demands and duties with '*in the future*'. On the other hand, NSc seem to mainly describe their anticipation by mentioning concerns with the three-word phrase.

4.2 Accounts of hypothesis II

AT that time' and '*AT the time*' occur frequently in the two essay corpora, but the frequent clusters do not support hypothesis ii). As far as the averaged frequency is concerned, '*AT that time*' in the JPC, occurring 1.76 times per 10,000 words, greatly exceeds the NSC equivalent cluster, accounting for only 0.28. The nearly eight times greater discrepancy in frequency may to some extent relate to the use of the cluster by JP learners: the cluster position appears to locate at either clause-initial, followed by a phrase delimiter comma, or clause-final followed by a sentence delimiter period (Figure 5), which confirm hypothesis ii). However, the NS do not show such consistency in terms of position. In addition to the occurrence of the three-word cluster at clause-initial and final positions, it appears after the subject position as in '*..., but the population of children at that time would generally level...*'; and despite the clause-final position the cluster emerges before the subordinate clause, as in '*I told them at that time that they had lost...*' The trigram, hence, turns out to retain a different environment in terms of its position, occurring frequently in both the JPC and NSC. '*AT the time*' does not show such a distinctive feature in its positioning.

As regards the JPC, the cluster of '*AT the same time*' in the NSC frequently follows '*and*', '*but*', '*while*' (Figures 6a and b) and occurs somewhat infrequently with a preceding conjunction (hypothesis □)). It appears to a large extent in particular positions in a clause — clause-initial and final — as shown in Figure 6a (hypothesis □)). This cluster demonstrates a number of features in the two corpora: the JPC

1 h school. I liked a English lesson at that time. But I could not learn
2 ever forget the painful experience at that time. But it will always re

3 ears behind Europe in this respect at that time. But now, Japan became
4 e met him. He had already grown up at that time. For a long time since
6 need a proper Japanese education. At that time, I didn't understand w
8 , I couldn't only say "I'm sorry." At that time, I thought that I wish
9 I could get the news immediately. At that time, I thought, e-mail is
10 ld think about, this is in Japan." At that time, I was shocked. Have I
11 nese will go to a foreign country. At that time, if we are able to spe
15 the far past, there are not money. At that time, they were subjected t
5 or me. Then I entered high school. At that time I began thinking about
7 I thought that I did not work yet. At that time I felt anxious about m
12 ing in it only women and children. At that time men thought much of Lad
13 Then they gradually spoke English. At that time racial discrimination
14 te of at least five thousands yen. At that time the victim was hurt th

Figure 5: *AT that time* in the JPC

4 vvinistence to the partner country. At the same time, we'll can give im
7 vvthere are many foreigner in Japan. At the same time, Japanese who go t
10 oney is the most necessary for me. At the same time, money is the most
11 etely. I feel regrettable so much. At the same time, I thought if I co
15 unemployed increase year by year. At the same time crime is on the in
5 vvBecause we can know their culture at the same time. Third, English wi
12 ence, and we can enjoy to talk at the same time. However, the
8 e. Because teacher ask many things at the same time. I can't do so man
1 vvcalled international society, and at the same time, information-oriен
2 vvhowever hard they try to be. And at the same time, they are not Japa
3 vvre junior high school student. But at the same time, I worry about the
6 vlearn Japan's culture and history at the same time learning English.
9 vvow to do we enjoy foreign language at the same time because to learn E

Figure 6a: *AT the same time* in the JPC

1 ecause of their aggressive action. At the same time other feminists su
6 ay the role of mother and father. At the same time, there are also me
8 cepts - has if anything increased. At the same time, workers have suff
10 oponents and continue to have sex. At the same time, teens will not b
7 at more than one can be fertilised at the same time to increase the ch
9 ehicles want to use the same roads at the same time. The 'A' roads int
2 ing against the white culture and, at the same time, are fighting agai
3 any women followed this advice and at the same time did their best to
4 taxes which they are paying. But, at the same time, with an ever decr
5 . He understands Caligula, but at the same time is unable to recon
11 uch of the crime in America, while at the same time regulate the consu
12 ngeles to go back to Oakland while at the same time the Los Angeles Ra

Figure 6b: *AT the same time* in the NSC

cluster tends to occur in a peculiar position in the clause, while the NSC cluster extends the cluster in a relatively consistent sequence by synthesizing with particular conjunctions (hypothesis (□)). Nevertheless, both corpora appear to retain the frequent fixed cluster '*AT the same time*', which seems to endorse hypothesis □).

4.3 Accounts of hypothesis III

The most frequent trigram in the NSC, '*IN order to*', contains the trigram of '*IN order*' which is listed in the top 20 bi-grams both in the JPC and the NSC. However, the trigrams containing the two-word sequence in the corpora lead to different three-word sequences: '*IN order*' in the JPC accounts for 22 instances with 22 instances of '*IN order to*', as shown in Figure 7a. The JPC does not vary the trigram '*IN order to*', while the NS use two types of trigram containing '*IN order*': '*IN order to*' and '*IN order for*'. As seen in Figure 7b, the latter trigram is a modification of '*IN order to*' into '*IN order for + noun/pronoun + to*'. However, the JP appear not to have learnt '*in order for + noun/pronoun + to + infinitive*': it is not the case that they never see the sequence. English grammar textbooks employed for teaching English to JP students apparently introduce the phrase of '*for + noun/pronoun*' preceding '*to + infinitive*' function as the agent of base-form verb, as in '*It is necessary for you to see a doctor*' (My underlining, Ishiguro 2003: 168). But JP are unlikely to be encouraged to generalise from the account to other contexts: '*in order to*' always appears in their text as a fixed phrase without modification. The account of '*IN order to*' in the textbooks is inevitably presented as if it were an invariable phrase '*in order to*', which may prevent JP students from producing the modified structure of '*IN order for noun/pronoun to*'.

1 e to have a large vocabulary of English in order to be useful in the feuture. 3It
is important for us to speak English in order to become a person who can
4 h is considered as an official language in order to communicate on keep contact
6 ry for students to wear school uniforms in order to control public morals, but
8 dren are usually use their own language in order to express their feeling and 9edy.
Gradualy, people would do anything in order to get money. At this time,
10 d "you must enter the famous university in order to get a position in a excellen
14 n a continent must speak many languages in order to live. Some people are taught
15 ry to force them to wear the uniforms in order to make them good students,
16 ve chance to speak English that's why in order to master English , we need
17 their opinions with parental affection in order to solve this problem. Now Eng
18 ce and farms needs English capacity. in order to talk foreigner in their
19 re in different ways. We need to master in order to understand foreigners. The

Figure 7a: IN order to in the JPC

1 of the computer. It is clear that in order for a computer to work it
3 literary education must be present in order for any good work to exist
7 ell if he would want a new stadium in order for him to become a part o
10 eturn of prayer into the classroom in order for students to pray toge
11 ablished. Jefferson believed that in order for the government to oper
15 er 1962 to reform the constitution in order for the president to be el
18 c Hero must display good qualities in order for the readers to have sy
19 rent, it is very scary for people. In order for these people to be su
25 g health throughout the community, in order to achieve the ultimately
26 on. He spent 18 months negotiation in order to achieve a plan for a un
60 search of the meaning of freedom, in order to discover himself. Initi
79 nfer the instabilities people have in order to get people to buy their
80 i lies about the nightmare she had in order to get away from a disaste
100 Division Series to have to survive in order to make it to the League C
103 rtain sacrifices that must be made in order to make peace between the
112 her, Sophia moved out of the house in order to prevent further violenc
115 lie persists, she decides to leave in order to prevent any future viol
150 love of oneself to protect oneself in order to survive. It was self-co

Figure 7b: IN order to/for in the NSC

The phrase ‘*in a way*’ shows flexibility in its position and grammar relating to the qualifying clause: the trigram is likely to appear at the sentence-initial or ending position, and precede relative clauses beginning with ‘*in which*’ and ‘*that*’, as seen in Figure 8. ‘*In a way*’ emerges frequently showing minor variation in the words prior to ‘*way*’. Table 4, which tabulates a sequence of ‘*IN a ...*’ in the NSC, illustrates that the phrase of ‘*way*’ varies in three ways: ‘*in a way*’, ‘*in a + adjective + way*’, ‘*in a + intensifier + adjective + way*’ – ‘*in a number/variety of ways*’ is also found in the five-word sequence containing ‘*in a ... way*’ (Figure 9a). The adjectives employed in the phrase are *certain, cynical, different, negative, positive, reasonable, similar, small, substantial* and *useless*, as in Figure 9a. The table also indicates that ‘*manner*’, as a synonym of ‘*way*’, frequently emerges in the sequence of ‘*in a + adjective + manner*’. The adjectives of the phrase identify with those of ‘*in a + adjective + way*’, for instance *different, positive, similar*, as in Figure 9b.

L3	L2	L1	Centre	R1	R2	R3
THE	THE	WAS	IN A	WAY	OF	AND
TO	IS	LIVE		WORLD	THAT	THE
A	BE	THEM		MORE	WAY	OF
OF	TO	BEING		VERY	IN	WHICH
AND	OF	IS		SENSE	AND	IS
CAN	A	PRESENTED		SINGLE	MANNER	THIS
HE	WAS			NUMBER	HE	IT
THAT	AND			TEST	WHERE	NOT
	HE			DIFFERENT	EUROPE	WAY
	ARE			COUNTRY	LIGHT	WAS
				PUBLIC		
				POSITIVE		

Table 4: Patterns of ‘*IN a + noun/adjective*’ in NSC

1 is worth living. He feels despair in a way about man's situation. He
2 other realization that . So in a way, Caligula can be seen as u
3 in a more logical and ordered way. In a way Da Vinci is a bad example,
4 still free to think and was happy in a way. He chose to treat the tas
5 e committing intellectual suicide. In a way he is driving for people t
6 and is ashamed of his background, in a way his entry into the party i
7 t is the money which protects them in a way. How could we ever ban a
8 doption institution has hurt them in a way in which they've never bee
9 eader on his own particular level, in a way in which it does not reach
16 depiction, he chose to communicate in a way that would last through th
17 cause people to conduct themselves in a way that they might not otherw
18 ble to see life in other countries in a way that was never before poss
19 ways to make a point, others dress in a way that has no particular mea
20 sh murderers or serious criminals in a way that will set an example f
21 ation of couch potatos. It is true in a way, the development of portab
22 e players that sit the bench. But in a way this is wrong, giving more

Figure 8: *IN a way* in the NSC

1 lieves that others are pushed forward in a certain way but he is. At the begin
2 hapiness and freedom, is portrayed in a cynical way by Camus. To summariz
3 m sure we can deal with the situation in a different way. I am sure nobody wil
4 uis it is clear that he is portrayed in a negative way and cannot put trust
5 ive light, which reflects, therefore, in a positive way on the PP. Similar
6 der ways to use these characteristics in a positive way? Our society and the
7 good values and it can still be lived in a reasonable way: . We can now
8 the fall of man, Adam's original sin. In a similar way to the situation in Arg
9 ht betray the party to his background, in a similar way to his father, thus epi
10 of reusing what we have. (Recycling in a small way also helps those w/o home
11 ism in the novel must be considered in a substantial way. Voltaire, like
12 nd, like all inventions can be used in a useless way. Not even truly deali

Figure 9a: *IN a + adjective + way* in the NSC

1 ion as a whole could work together in a democratic manner. To a cer
2 s also an accessible character but in a different manner. He has an un
3 mary episode in the book describes in a jovial manner the castle of Ba
4 d the media portray certain people in a poor manner and the majority o
5 ment as a question of consequence in a positive manner can not possib
6 ey gather and celebrate happiness. In a similar manner, Celie writes G
7 l arts link communities together. In a simplistic manner, it is often
8 r the legitimate cases to be heard in a timely manner. It now takes f

Figure 9b: *IN a + adjective + manner* in the NSC

4.4 Accounts of hypothesis IV

The hierarchy of four-word clusters which have been decomposed into a shorter corresponding cluster becomes increasingly explicit if we include frequent bi-grams. ‘*Language IN the world*’ can be decomposed into the frequent trigrams ‘*language IN the*’ and ‘*IN the world*’. ‘*Language IN the*’ is composed of the bi-grams ‘*language IN*’ and ‘*IN the*’, while ‘*IN the world*’ could be divided into ‘*IN the*’ and ‘*the world*’. It should again be noted that the bi-grams and trigrams are found as the 20 most frequent word clusters. The clusters in Tables 5a and 5b are extracted exclusively from those containing preposition *IN*. The compartment which ‘*the world*’ is supposed to fill is crossed out, since the two-word cluster is not found as a frequent bi-gram containing *IN*. Table 5a illustrates the completeness of the two- to four-word clusters in the JPC, indicating that frequent word clusters are consistently employed when its boundary is extended. At the same time, a frequent four-word cluster of NSC ‘*prayer IN public schools*’ shows a close composition of corresponding clusters: the four-word cluster can be decomposed into ‘*prayer IN public*’ and ‘*IN public schools*’, and the latter trigram into ‘*IN public*’ and ‘*public schools*’. But, as stated in the account of the JPC four-word cluster, the compartment of ‘*public schools*’ is deleted because of not having *IN* in the cluster. Bearing in mind that ‘*prayer IN public*’ is divided into ‘*prayer IN*’ and ‘*IN public*’, however, the compartment for ‘*prayer IN*’ appears to be left blank. This is because the bi-gram is not found in the 20 most frequent uses in the NSC. In order to distinguish it from clusters without *IN* being crossed out, a compartment of infrequent clusters is assigned to the category of ‘unfilled’. Tables 5a and 5b suggest that the four-word cluster of JPC establishes a rigid hierarchy of frequent word clusters, while that of NSC creates, to some extent, a flexible hierarchy.

2word cluster	3word cluster	4word cluster
<i>language IN</i>		
<i>IN the</i>	<i>language IN the</i>	
<i>IN the</i>	<i>IN the world</i>	<i>language IN the world</i>

Table 5a: Cluster progression of two-word to four-word frequent sequences of *language in the world* in the JPC

2word cluster	3word cluster	4word cluster
<i>IN public</i>	<i>prayer IN public</i>	
<i>IN public</i>	<i>IN public schools</i>	<i>prayer IN public schools</i>

Table 5b: Cluster progression of two- to four-word fsequences of *prayer in public schools* in the NSC

This is not only the case in the second *IN* clusters, but also in the third and final-position *IN* clusters, which clearly indicate the inflexibility of the JPC four-word clusters and the flexibility of the NSC ones. Tables 6a and 7a of JPC clusters show a complete structure of frequent word clusters, while Tables 6b and 7b retain some blank chambers, supposedly being unfilled by infrequent clusters. The frequency is likely to arise from the variable and flexible clusters which consist in an NSC four-word cluster.

2word cluster	3word cluster	4word cluster
<i>language IN</i>	<i>second language IN</i>	
<i>language IN</i>		<i>second language IN the</i>
<i>IN the</i>	<i>language IN the</i>	

Table 6a: Cluster progression of two-word to four-word frequent sequences of *second language in the* in the JPC

2word cluster	3word cluster	4word cluster
		<i>the way IN which</i>
<i>IN which</i>		

Table 6b: Cluster progression of two-word to four-word frequent sequences of *the way in which* in the NSC

2word cluster	3word cluster	4word cluster
		<i>to study English IN</i>
<i>English IN</i>	<i>study English IN</i>	

Table 7a: Cluster progression of two-word to four-word frequent sequences of *to study English in* in the JPC

2word cluster	3word cluster	4word cluster
		<i>to an increase IN</i>

Table 7b: Cluster progression of two-word to four-word frequent sequences of *to an increase in* in the NSC

4.5 Accounts of hypothesis V

Four-word clusters can be seen as classifiable into two sub-categories, one topic-specific and the other topic-unspecific. '*Homosexuals IN the military*' and '*prayer IN public schools*' are apparently identified as topic-specific, since they literally imply the topic of the essay containing the sequences. Whereas, '*argument IN favor of*' and '*increase IN the number*' may fall into the sub-category of topic-unspecific, where the sequences do not indicate a topic. Interestingly, the components of the clusters reflect a lexical difference before and after preposition *in*: in terms of topic-specific clusters, '*homosexual*' and '*military*' are not frequent as topics and they are not expected to co-occur often in everyday speech. In turn, '*homosexuals in the military*' can be supposed to be an essay topic with no internal flexibility in its components, making the frequency largely depend on the topic. Likewise, '*prayer in public schools*' may seem unusual; it thus can be a topic-specific cluster. By contrast, the components of topic-unspecific clusters are expected to co-occur: '*argument*' is predictably followed by '*favor*', because a primary aim of arguments is to determine '*favor*' or opposition. Similarly, '*number*' closely relates to '*increase*', not specifically to a topic. As a result, topic-unspecific clusters seem to be available irrespective of essay topic, and the environment of such clusters does not appear as consistently as those of topic-specific clusters.

	TOPIC-SPECIFIC	TOPIC-UNSPECIFIC
JPC	<i>language IN the world</i>	
	<i>English IN the future</i>	
	<i>interested IN English and</i>	
	<i>people IN the world</i>	
NSC	<i>prayer IN public schools</i>	<i>argument IN favor of</i>
	<i>homosexuals IN the military</i>	<i>increase IN the number</i>
		<i>people IN the UK</i>

Table 8a: Topic-specific and unspecific four-word clusters with second-position *IN*

Table 8a shows the classification in specifying clusters in relation to essay topic. However, ‘*people IN the world*’ among JPC does not fall into either sub category: it should be somewhere between topic-specific and topic-unspecific. According to the earlier finding, the components of topic-unspecific clusters, due to their related meaning, should be supposed to occur simultaneously, for example, ‘*argument*’ and ‘*favor*’; however, this may not be the case regarding ‘*people IN the world*’. The cluster retains the words —*people* and *world* — in the same category as the NSC topic-unspecific ‘*people IN the UK*’, which differs only in the indicated places. But, by reading the JPC concordance lines, the JPC four-word cluster appears rather to incline toward being topic-specific, inferring ‘*people outside Japan who speak or learn English*’. The cluster has been classified into an extra sub-category containing clusters which seem to be topic-unspecific, but in fact are somewhat topic-specific if judged from their context. This is summarised as the finding that apparently topic-unspecific clusters seem to shift into being assessed as topic-specific when read in context—but the converse scarcely ever happens.

	TOPIC-SPECIFIC	TOPIC-UNSPECIFIC
JPC	<i>second language IN the</i>	
	<i>common language IN the</i>	
	<i>are written IN English</i>	
	<i>to speak IN English</i>	
	<i>many people IN the</i>	
NSC	<i>I was IN the</i>	
	<i>of prayer IN public</i>	<i>the way IN which</i>
	<i>not believe IN God</i>	<i>the people IN the</i>
	<i>to power IN 1981</i>	<i>the best IN the</i>
		<i>it is IN the</i>
	<i>an increase IN the</i>	
	<i>are already IN existence</i>	

Table 8b: Topic-specific and unspecific four-word clusters with third-position *IN*

	TOPIC-SPECIFIC	TOPIC-UNSPECIFIC
JPC	<i>to study English IN</i>	
	<i>to speak English IN</i>	
	<i>to learn English IN</i>	
	<i>to use English IN</i>	
	<i>English is spoken IN</i>	
	<i>a second language IN</i>	
<i>over the world IN</i>		
NSC		<i>for the best IN</i>
		<i>came to power IN</i>
		<i>to an increase IN</i>

Table 8c: Topic-specific and unspecific four-word clusters with final-position *IN*

The assessment of topic-specific and topic-unspecific is likely to depend to some extent on the segment of the cluster: for example, '*increase IN the number*', '*an increase IN the*', '*to an increase IN*' seemingly belong to a larger sequence of '*to an increase IN the number*'. However, the concordance lines of '*increase IN the number*' and '*an increase IN the*' do not show any inclination to be topic-specific. When they include '*to*', which results in '*to an increase IN*', they may attract '*lead to*' and then the environment and context will turn out to be relatively negative. Similarly, '*came to power IN*' means, in general, to establish government authority, while '*to power IN 1981*' explicitly specifies in this corpus the regime of President Mitterrand.

Table 8 a, b and c illustrates that JP employ no topic-unspecific four-word clusters. It may well be deduced, however, that they are prone to use topic-specific ones fairly frequently, although topic-unspecific clusters occur infrequently in their writings. It is speculated that pre-fabricated clusters may help learners to produce with less effort. Certainly, they are likely to persistently resort to topic-specific clusters, whereas they do not make the most of topic-unspecific clusters in their writing. Consequently, learning to use topic-unspecific clusters may encourage them to develop expressiveness in producing English verbalisation.

5. Discussion

5.1 Implications of the key hypotheses

We have observed so far what distinguishes the features of the multi-word sequences containing *IN*, *ON* and *AT*, as used by Japanese-speaking learners, compared with those used by native speakers. In the light of the comprehensive accounts of the

co-occurring words with the prepositions, I want to propose some key hypotheses concerning Japanese learners' phraseology for the three prepositions, following the discussions in the previous chapters:

- i) *On the whole, the same N-grams are frequent in both corpora, but they tend to occur with different co-texts in each corpus.*
- ii) *There are a number of clusters occurring persistently in particular positions in a clause. In particular, the Japanese learners tend to restrict certain clusters to certain positions.*
- iii) *Frequent clusters can be classified into flexible vs. consistent sequences, in terms of their internal structure.*
- iv) *There are some frequent longer clusters composed of frequent shorter clusters. The short N-grams gradually extend the string boundary from two- to three- and then three- to four-word clusters.*
- v) *Frequent long clusters can be divided into two types – topic-specific and topic-unspecific – as regards their relationship to the context of the essay. The emergence of the topic-specific ones highly correlates to the topic of the essay.*

The first hypothesis states that identical multi-word clusters in the Japanese-speaking learner corpus (JPC) and the native speakers' corpus (NSC) co-occur with different collocational words. '*IN the world*' in the JPC frequently follows '*language*' or '*people*' along with such adjectives as *centre*, *common*, *important* and *language* which represent uniformity and commonality. Drawing on Sinclair's (2004) definitions of language description, '*IN the world*' when used by Japanese-speaking learners collocates with *many*, *most*, *centre*, *common* the 'colligation' of which are identified adjectives. The 'semantic preferences' of the collocates, in turn, represent abundance, plentifullness, uniformity and commonality and the phrase seems to indicate the 'semantic prosody' of worldwide pervasiveness (Table 9). Given that '*English*' as a language frequently co-occurs in the JPC outputs with '*IN the world*', the semantic prosody of the phrase could possibly imply that Japanese-speaking learners – and indeed, other learners – consider *English* as common and a language which is learnt by many people. The psychological representation of the JP learners results in the realization of their distinctive collocates, which, under the name of 'semantic association', is accounted for by Hoey (2005). However, it should be noted that there is a possibility that specific essay topics may encourage these learners to establish such collocational behaviour. In fact, '*learning English*' is the most frequently selected topic in the JPC.

	Semantics			Grammar	Core
PROSODY	pervasiveness				
PREFERENCE			abundance commonality		
COLLIGATION			<u>adjective</u>	<u>noun</u>	<u>preposition</u>
COLLOCATION	<i>English</i>	...	<i>many most a lot of centre common important official</i>	<i>people language</i>	<i>IN</i>
					<i>the world</i>

(Adapted from Sinclair 2003: 151)

Table 9: Lexical description of *IN the world* in the JPC

By contrast, native speakers exhibit a different collocational behaviour from the JP. '*IN the world*' functioning as intensifier follows '*every + noun*', '*every-/nothing*' and superlative clauses as in ...*to be the best in the world*. Some research claims that the overuse of a given word or phrase by learners can be attributed to translation from their mother language (Granger 1998, Granger and Tyson 1996, Altenberg and Granger 2001). However, the claim runs counter to the finding that the intensifiers and superlative clauses alongside '*IN the world*' occur infrequently in the JPC, although translations of these phrases are accessible in the Japanese language. This may imply that the transferability from learner's language to the target language may be restricted by the learner's perceptions about the target language. Hoey (2005) accounts for the process of psychological association found in collocations: the associations primed by learning a second language are held to overlap those of the earlier language, resulting in creating a new association for the target language. In addition to this, he suggests that establishing semantic associations is a personal experience and it is subject to different experiences of language among language users (Hoey 2005). Hence, it can be concluded that the extent of transferability to the target language is unlikely to be accounted for by contrastive analysis with the learners' first language; rather, detailed accounts of the learner's language, such as learner corpora, could illustrate the way in which the language learner elaborates his/her own second language.

The second hypothesis indicates that some multi-word clusters in the JPC in particular are prone to occur at the sentence-initial and final position. '*AT that time*'

accounts for the disposition of its positioning in sentences by Japanese-speaking learners: the phrase is often followed by a comma or sentence delimiter. Examined in the concordance lines of the JPC, '*AT the time*' occurs at the sentence-initial position followed by comma; otherwise the phrase is located at the sentence-final position. Linking adverbials '*AT the same time*' (Biber *et al.* 1999: 886) account for the discrepancy in position between the JPC and NSC. A number of instances where the linking adverbial is located at sentence-medial position, as well as -initial and -final, are found in the NSC. Unlike the NSC, Japanese-speaking learners employ them only at the sentence-initial and -final position.

Hoey (2005) points out that a given word occurs at a particular position in a clause, while its synonyms may occur at a different position: preference is given to a particular word. As has been noted, some of the linking adverbials occasionally occur after subject noun phrases. Japanese-speaking learners, however, persist in locating the adverbial phrases exclusively at the sentence-initial or -final position. As a result, it can be inferred that learners may resort to locating them in these positions although they may have learnt that medial-position is permissible, as indicated in an English grammar reference book used in classrooms (Ishiguro 2003) Sinclair's 'slot-and-filler' concept underlying the 'open choice principle' (1991: 109) accounts for this inclination to position the adverbials in a particular place: the learners slot adverbial phrases in the slot at sentence-initial or -final position. Note that this tendency for learners to locate adverbial phrases and connectors is not specific to learners with any particular L1 (Granger and Tyson 1996). In order to raise awareness of phrase location, the comparison between learners' and natives' texts can highlight the difference and encourage learners to improve the sequence of their sentences.

As for hypothesis iii), '*IN order to*' in the JPC shows no variability in its internal word combination. The phrase seems to be considered a pseudo-fixed phrase by Japanese learners. In contrast, the phrase denoting a purpose or intention of doing something in the NSC shows syntactical flexibility by embedding '*for + pronoun*' before *to-infinitive*. Similarly, the phrases extracted from the NSC are observed to allow for internal flexibility, while those from the JPC maintain consistency in the internal variation of phrases. In addition to this phrase, in the JPC, '*AT + determiner + time*' is identified as a quasi-consistent phrase, e.g., '*AT the time*' and '*AT that time*', while it varies in such determiners as *the, any, this, that* in the NSC. Native speakers also make use of internal lexical syntactic variation in terms of phrases containing *IN* and *way*: '*IN a way*', '*the way IN which*', '*IN a + adjective + way*', '*IN very/more + adjective + way*'.

On the whole, phrases in the JPC are prone to have less flexibility and maintain consistency in internal variation. A consistent phrase is found to be introduced as if it

were a stable unit of phrase in a textbook (Kasashima 2006). Institutional inputs at school and classroom are highly likely to affect the outputs of language learners, since teachers and teaching materials may encourage learners to establish shortcuts in the form of semantic associations (Hoey 2005). As regards its being a strategy of language learners, they tend to output common words and phrases which are learnt in the early stages of language learning – these are dubbed ‘lexical teddy bears’ (Hasselgren 1994: 237). ‘Lexical teddy bears’ are seemingly helpful for learners to produce outputs, but such words and phrases, which denote a wide range of meanings, are likely to deviate disadvantageously from the typical phraseological behaviours of the target language. The comparison of collocational behaviours in a native corpus and learner corpus may be expected to encourage learners to raise their awareness of collocation.

Given that hypotheses iv) and v) interrelate closely, the implications of these will be discussed in this section. The former hypothesis suggests that a frequent four-word cluster appears to incorporate a frequent two- or three-word cluster. As regards the JPC, ‘*language IN the world*’ occurs more frequently than the other four-word clusters; not only ‘*language IN*’ and ‘*IN the*’ among the two-word clusters but also ‘*language IN the*’ and ‘*IN the world*’ among the three-word ones occur frequently, occupying the principal twenty clusters. This finding is likely to reveal the consistency of ‘*language IN the world*’ – this could be deemed relevant to hypothesis iii) in terms of consistency. It can be speculated that such a phrase learnt in the early stage of learning functions as one of the ‘lexical teddy bears’ in the JP’s outputs (Hasselgren 1994: 237).

Some research of learner corpora argue that learners resort to an idiosyncratic multi-word unit (Hasselgren 1994, De Cock 1998, Altenberg and Granger 2001). As has been observed, ‘*language IN the world*’ is employed as if it were a ‘formula’ (Wray 2002) in the JPC. However, it turns out that native speakers output a number of the clusters which remain consistent in the extension from two- to four-word clusters: ‘*prayer IN the public schools*’, ‘*come to power IN*’, ‘*not believe IN God*’, ‘*effect ON the agriculture*’, ‘*number of cars ON*’ in the NSC. It should be noted that these clusters apparently specify the context of the essays in which they occur in the text (topic-specific in hypothesis v)). It seems unusual, however, that native speakers persist in outputting an inflexible phrase, although they show flexibility in the accounts of hypothesis iii). It could, in turn, be speculated that replicating an essay title or part of the essay question may result in outputting a consistent multi-word cluster. Hence, essay titles and written information for essay writing should be registered in the course of compiling essay corpora.

5.2 Ways forward

The implications of the five hypotheses which arise from the comparison of a Japanese-speaking learner corpus and a reference native speaker one appear to indicate the contrast of collocational behaviour between learners and native speakers. There are many studies which rely only on the intuition of native speakers in terms of the assessment of learners' outputs, such as error analysis. However, the comparison of the concordance lines of the two corpora clearly illustrate how and in what learners' outputs differ from what native speakers typically output, rather than what is correct or wrong. As a result, the comparison of learners' and natives' outputs may encourage learners to develop an awareness of collocation. Johns (1991) suggests that a 'data-driven learning (DDL)', which employs concordance lines from a reference corpus for second language learning, elicits autonomous findings on collocation from learners. In order to increase the awareness of collocation, it might be helpful for learners or teachers to notice the differences of outputs in collocation and learn the collocational behaviour which native speakers typically show.

Learner corpora still allow us the possibility of investigating learner outputs. Exploiting the technology of computers enables us to account for the process of language learning and spot the determinants which may obstruct language learning. The development of retrieval and analysis in corpus software and the connection of text with sound and image recording in the process of language learning will be able to account for the processes of language acquisition in detail (Reder *et al.* 2003, Myles 2005). This study has sought to shed light on the phraseology in learners' outputs by means of retrieving the multi-word clusters containing *IN*, *ON* and *AT* and comparing them in the two corpora. I hope that the implications of this study will contribute to further analysis of learner corpora and descriptions of learner language.

References

- Altenberg, B. and S. Granger (2001). 'The Grammatical and Lexical Patterning of MAKE in Native and Non-native Student Writing'. *Applied Linguistics*, 22(2), 173–95.
- Altenberg, B. and 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: Addison Wesley Longman Limited
- Biber, D., S. Conrad and V. Cortes (2004). 'If you look at...: Lexical Bundles in University Teaching and Textbooks'. *Applied Linguistics*, 25(3), 371–405.
- Biber, D., S. Johansson, G. Leech, S. Conrad and Finegan, E. (1999). Lexical Bundles.

- In D. Biber, S. Johansson, G. Leech, S. Conrad and E. Finegan (eds.) *Longman Grammar of Spoken and Written English*, pp. 990–1024. Harlow: Pearson Education Limited.
- Biber, D. and R. Reppen (1998). Comparing native and learner perspectives on English grammar: a study of complement clauses. In S. Granger (Ed.) *Learner English on Computer*, pp.145–58. New York: Addison Wesley Longman Limited.
- Dagneaux, E., S. Deness and S. Granger (1998). ‘Computer-aided error analysis’. *System*, 26, 163–74.
- De Cock, S. (1998). ‘A Recurrent Word Combination Approach to the Study of Formulae in the Speech of Native and Non-native Speakers of English’. *International Journal of Corpus Linguistics*, 3(1), 59–80.
- De Cock, S., S. Granger, G. Leech and T. McEnery (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (ed.) *Learner English on Computer*, pp. 67–79. London: Addison Wesley Longman Limited.
- De Haan, P. (1997). An experiment in English learner data analysis. In I. d. M. H. Wekker (Ed.) *Studies in English Language and Teaching: In Honour of Flor Aarts*, pp.215–29. Amsterdam: Rodopi.
- Ellis, R. (1994). *The Study of Second Language Acquisition*. Oxford: Oxford University Press.
- Flowerdew, L. (1998). ‘Corpus linguistic techniques applied to textlinguistics’. *System*, 26, 541–52.
- Flowerdew, L. (2003). ‘A Combined Corpus and Systemic-Functional Pattern in a Student and Professional Corpus of Technical Writing’. *TESOL Quarterly*, 37(3), 489–511.
- Granger, S. (1998). ‘Prefabricated Patterns in Advanced EFL Writing: Collocations and Formulae’. In A. P. Cowie. (Ed.) *Phraseology : theory, analysis, and applications*, pp. 145–60. Oxford Clarendon Press.
- Granger, S. (2002). A Bird’s-eye view of learner corpus research. In J. H. a. S. P.-T. Sylviane Granger (Ed.) *Computer learner corpora, second language acquisition and foreign language teaching*, pp. 3–33. Amsterdam and Philadelphia: John Benjamins.
- Granger, S. (2003). ‘The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second Language Acquisition Research’. *TESOL Quarterly*, 37(3), 538–46.
- Granger, S. and S. Tyson (1996). ‘Connector usage in the English essay writing of native and non-native EFL speakers of English’. *World English*, 15(1), 17–27.
- 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), 237–59.
- Hoey, M. (2005). Lexical Priming: A new theory of words and language. London/New York: Routledge.
- Howarth, P. (1998). The Phraseology of Learners' Academic Writing. In A. P. Cowie (ed.) *Phraseology : theory, analysis, and applications*, pp. 161–86. Oxford: Clarendon Press.
- Hunston, S. (2002). Corpora in Applied Linguistics. Cambridge: Cambridge University Press.
- Ishiguro, A. (ed.). (2003). *Sogo Eigo Forest*. Tokyo: Kirihsara Shoten.
- Johns, T. (1991). 'Should you be persuaded - two examples of data-driven learning materials'. *English Language Research Journal*, 4, 1–16.
- Kasashima, J., H. Asano and Y. Shimomura (2006). New Horizon English Course. Tokyo: Tokyo Shoseki.
- Lorenz, G. (1998). Overstatement in advanced learners' writing: stylistic aspects of adjective intensification. In S. Granger (ed.), *Learner English on Computer*. pp. 53–66. London: Addison Wesley Longman Limited.
- Myles, F. (2005). 'Interlanguage corpora and second langauge acquisition research'. *Second Language Research*, 21(4), 373–91.
- Nesselhauf, N. (2003). 'The Use of Collocations by Advanced Learners of English and Some Implications for Teaching'. *Applied Linguistics*, 24(2), 223–42.
- Nesselhauf, N. (2004). Learner corpora and their potential for language teaching. In J. M. Sinclair. (ed.), *How to use corpora in language teaching*, pp. 125–52. Philadelphia: John Benjamins.
- Nesselhauf, N. (2005). Collocations in a Learner Corpus. Amsterdam / Philadelphia: John Benjamins Publishing Company.
- Partington, A. (1998). Patterns and Meanings. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Patch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on Computer*, pp. 107–118. London: Addison Wesley Longman Limited.
- Reder, S., K. Harris and K. Setzler (2003). 'The Multimedia Adult ESL Learner Corpus'. *TESOL Quarterly*, 37(3), 546–58.
- 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: Addison Wesley Longman Limited.
- Sinclair, J. (1991). *Corpus Concordance Collocation*. Oxford: Oxford University Press.

- Sinclair, J. (2004). Trust the Text: language, corpus and discourse. New York: Routledge.
- Virtanen, T. (1998). Direct questions in argumentative student writing. In S. Granger (ed.), Learner English on Computer. pp. 94–106. London: Addison Wesley Longman Limited.
- Wray, A. (2002). Formulaic Language and the Lexicon. Cambridge: Cambridge University Press.