

ISSN 0385-566x

LEXICON

No. 44

2014

IWASAKI LINGUISTIC CIRCLE

Tokyo

The NICT JLE Corpus in Well-Formed XML Format: From Analyses of Surface Forms to Functions in Longer Stretches of Discourse	AIKA MIURA HIROSHI SANO	1
An Analysis of Three Collocations Dictionaries for Learners of English	YASUTAKE ISHII SATORU UCHIDA KYOKO HAYASHI YUICHIRO KOBAYASHI	24
The Motivation for Using Detached Participial Clauses in English	HITOSHI KAWABATA	54
The Grammatical Meanings of the Adjectival Participle in English	TAKERU ITOKAWA	73
投稿規定		86
編集後記		86

Iwasaki Linguistic Circle

c/o Eigo-senko kyoiku junbishitsu
Tokyo University of Foreign Studies
3-11-1 Asahi-cho, Fuchu-shi
Tokyo 183-8534
Japan

Shoji Miyai, Chairman

2014年6月20日印刷 2014年6月29日発行

編集兼発行者 岩崎研究会 代表者 宮井捷二
振替口座 00140-2-8737

〒183-8534 東京都府中市朝日町 3-11-1 東京外国語大学 英語専攻教育準備室
email: iwaken.info@gmail.com

印刷所 研究社印刷株式会社

The NICT JLE Corpus in Well-Formed XML Format: From Analyses of Surface Forms to Functions in Longer Stretches of Discourse

AIKA MIURA
HIROSHI SANO

1. Introduction

This paper explores the possibilities of expanding the scope of spoken learner corpora from investigations of forms (e.g. lexico-grammatical features) to those of pragmatic functions (e.g. speech act expressions). It focuses on the National Institute of Information and Communications Technology Japanese Learner English Corpus (NICT JLE), which contains the data of approximately 1,200 Japanese English as a foreign language (EFL) learners taking an oral interview test. The corpus comprises annotated transcripts that contain data on the speakers (i.e. interviewers and interviewees), including their level of English proficiency, and extra-linguistic features of the utterances, for example, fillers, pauses, self-corrections, and overlaps between speakers.

Recently, the advent of multi-modal corpora of spoken data has made it possible to analyse not only prosodic features and turn-taking in interactions, but also speakers' gestures, facial expressions, and posture, provided by audio or audio-visual files (Adolphs; Knight and Adolphs; O'Keeffe, Clancy, and Adolphs). Although the NICT JLE Corpus was created before the advent of multi-modal corpora and is composed of only the transcribed data of spoken interactions, marked-up elements annotated in the corpus are highly detailed and informative in the investigation of interactive spoken data of EFL learners. However, the markup used to encode extra-linguistic features in the

NICT JLE Corpus does not correspond to well-formed XML format. For example, there are missing start- and end-tags, especially when the data include intersecting utterances by different speakers. Once the corpus is formatted in valid XML, it will be relatively convenient to investigate the pragmatic functions of the utterances in longer stretches of discourse and to examine the contexts of interactions between the interviewers and interviewees.

The main aim of this paper is to explain the process of converting the NICT JLE Corpus text files into well-formed XML documents and to describe how the XML-converted corpus can be applied to research on pragmatic functions. The paper describes how valid XML documents allow researchers to expand the scope of analysis of the NICT JLE Corpus data. Specifically, the operation of handling the data becomes less constrained, and thus, it is easy to extract specific corpus data according to the researchers' needs (e.g. by segmenting the data by learner groups or interview stages) and to apply additional annotations to the corpus, such as annotations representing pragmatic features.

2. Corpus-based Studies in the Domain of Pragmatics and Interlanguage Pragmatics

With the recent compilation of language corpora, researchers have gained access to a large collection of naturally occurring data (O'Keeffe et al.). However, corpus-based research on the nature of the relationship between linguistic form and function has remained relatively scarce (Adolphs; Knight and Adolphs; O'Keeffe et al.). The study, which aims to explain the disparity between linguistic form and meaning in context, is an investigation in pragmatics. Further, it explores how a speaker's intended meaning to the hearer can be realised in particular linguistic forms. The functions, for example, of speech act expressions and conversational implicatures, are not necessarily equivalent to their surface meaning.

The background to the difficulties of applying a corpus-based approach to pragmatics is as follows. First, not many spoken corpora

have been available for pragmatic analysis, as it is difficult to compile spoken corpora (Adolphs). According to O’Keeffe et al., it takes ten hours to transcribe one hour of talk, which typically consists of approximately 10,000 to 15,000 words. They note that “spoken corpora are few, compared to written corpora, and those that are available may not be designed in such a way that suits the study of pragmatic features” (33). For example, it is necessary to search manually for instances of pragmatic features such as speech acts. Manual tagging for pragmatic features is, of course, time-consuming, yet the ability to interpret a particular function of a speech act expression based on the context is vital (Adolphs). Thus, because it is necessary but not always easy to infer the various contexts of utterances in the corpus (Romero-Trillo), researchers often make their classifications “partly based on intuition” (Adolphs 9). Traditional pragmatics has been discussed in terms of “invented examples of utterances based on native speakers’ intuitions” to support a division between form and function (Adolphs 18, 21). These drawbacks lead to a general scepticism of corpora exploration focused on extended discourse stretches (Adolphs).

Nevertheless, Adolphs and O’Keeffe et al. emphasise the importance of spoken corpora for pragmatic investigations. As mentioned above, “much of the work in pragmatics has been based on invented examples of utterances based on native speaker intuition” (Adolphs 21). However, this intuitive aspect can be overcome by the nature of corpora, as they allow us to re-examine the researchers’ initial analysis of speech act expressions and to “re-evaluate more traditional frameworks for assigning functions to utterances” (Adolphs 90).

Learner corpora, which are the main focus of this study, contain collections of texts produced by second or foreign language learners. Researchers aim to track the developmental aspects of learners’ language and, particularly, to highlight areas which are difficult for learners to learn or acquire (O’Keeffe et al.). Studies of learner corpora predominantly employ frequency-based lexico-grammatical analysis. There is no doubt that interlanguage pragmatics has been heavily based on data collected from elicitation tasks such as the Discourse

Completion Task (DCT) (Kasper and Blum-Kulka; Kasper and Rose; Takahashi; Schauer), while studies based on spoken learner corpora remain scarce.

3. The Aim of the Study: Overcoming Difficulties with Pragmatic Analyses

This paper explains the conversion of TXT files of the NICT JLE Corpus into XML format. As mentioned in Section 2, it is our desire to establish frameworks for interpreting patterns of use that go beyond the lexico-grammar in order to conduct a pragmatic analysis of the NICT JLE Corpus data.

The NICT JLE Corpus may be downloaded for free on the website. It is one of the largest spoken corpora, and is comprised of written transcripts of audio-recorded speech samples of learners at different levels of proficiency. The corpus contains not only information on interactive features such as overlap between speakers, but also data on extra-linguistic features such as fillers, pauses, and repetitions. The marked-up information in the corpus is very useful for investigating pragmatic features. For example, interactive contexts in a role-play session about shopping can be very informative in the identification of requestive speech acts, as they allow the researchers to examine the intended meaning of the speaker.

However, difficulties arise in pragmatic analyses of the NICT JLE Corpus data. First, the corpus analysis tool, called Analyzer, that is provided by the compiler does not easily allow the researchers to amend the data, for example, to separate the data into smaller parts according to the segments annotated or to add more pragmatic annotations. Secondly, only limited frequency counting based on lexical items and tags are available with Analyzer. Finally, even though the TXT files of the corpus data have recently been made freely available online, thereby enabling researchers' direct access, it is still difficult to amend the data. Since the format of annotation tags in the corpus is XML-like but not completely well-formed, it is difficult to use Perl, a programme that allows researchers to amend and segment data in

well-formed XML files.

The remainder of this paper is organised as follows. Section 4 is an introduction to the NICT JLE Corpus, including a description of how and in what format it has been provided by the compiler. After describing the rules of XML files in Section 5, next, in Section 6 we point out examples of ill-formed tags within the corpus and explain the process of converting the ill-formed tags into well-formed XML. Finally, Section 7 discusses how the converted XML may be used for pragmatic analyses, and frequency information on types and tokens is generated.

4. The NICT JLE Corpus

4.1. What is the NICT JLE Corpus?

The NICT JLE Corpus was created by the National Institute of Information and Communication Technology (NICT) in Japan. It contains 1.2 million words in transcripts of 1,281 Japanese EFL learners taking a speaking proficiency test called the Standard Speaking Test (SST) (Izumi, Uchimoto, and Isahara). The SST is a 15-minute oral interview which was developed based on the Oral Proficiency Interview (OPI) of the American Council on the Teaching of Foreign Languages (ACTFL). A total of 300 hours of interviews were transcribed. The SST has five stages: (1) answering warm-up questions (3–4 minutes), (2) describing a single picture (2–3 minutes), (3) doing a role-play with the interviewer (1–4 minutes), (4) narrating picture sequences (2–3 minutes), and (5) answering wind-down questions (1–2 minutes). The participants who took the test were assessed holistically and classified into one of nine proficiency levels called SST Levels: Levels 1 to 3 (Novice), Levels 4 and 5 (Intermediate Low), Levels 6 and 7 (Intermediate Mid), Level 8 (Intermediate High), and Level 9 (Advanced). Stages 2 to 4 consist of the “task” and “follow-up” sessions.

In addition, 167 files of the 1,281 files in the corpus are error-tagged in terms of 47 grammatical and lexical features. Also, for the purpose of comparison with the learner data, a subcorpus containing the data

of 20 native English speakers is available. However, it should be noted that the error-tagged corpus and native corpus are not dealt with in this study, and they are not examined in the conversion to well-formed XML files.

4.2. Tags Annotated in the Corpus

There are 30 basic annotation tags used in the corpus. They can be classified broadly into four types based on the type of information they denote: interview structure, the interviewee's profile, speaker turns, or utterance phenomena such as fillers, repetitions, self-corrections, and overlapping (National Institute of Information and Communication

```

<interview>
<filename></filename>
<head version="1.3">
<date></date>
<sex></sex>
<age></age>
<country></country>
<overseas></overseas>
<category></category>
<step></step>
<TOEIC></TOEIC>
<TOEFL></TOEFL>
<other_tests></other_tests>
<SST_task2></SST_task2>
<SST_task3></SST_task3>
<SST_task4></SST_task4>
<SST_level></SST_level>
</head>
<body basictag_version="2.1.3">
<stage1></stage1>
<stage2>
<task></task>
<followup></followup>
</stage2>
<stage3>
<task></task>
<followup></followup>
</stage3>
<stage4>
<task></task>
<followup></followup>
</stage4>
<stage5></stage5>
</body>
</interview>

```

Fig. 1. Tags for representing the interview structure and the interviewee's profile.

```

<A>How are you?</A>
<B>Fine. Thanks. How are you?</B>
<A>I'm fine, too. Thank you.</A>

```

Fig. 2. Tags for representing speaker turns.

Table 1 Tags for representing utterance phenomena

Tag	Meaning
<F></F>	Filler / Filled Pause
<R></R>	Repetition
<R?></R?>	Repetition (which the transcriber is not confident transcribing)
<SC></SC>	Self-correction
<SC?></SC?>	Self-correction (which the transcriber is not confident transcribing)
<CO></CO>	Utterances which are cut off
<?></?>	Utterances which the transcriber is not confident transcribing
<??></??>	Utterances which are impossible to transcribe
<H pn="X"></H>	Hidden personal information or discriminatory term
<JP></JP>	Japanese
<.></.>	Pause which lasts 2 to 3 seconds
<..></..>	Pause which lasts more than 3 seconds
	Overlapping utterances of Speaker A and Speaker B
<nvs></nvs>	Non-verbal sounds such as a sniff, laughter, cough, or sigh
<laughter></laughter>	The speaker produces the utterance while laughing.
<ctxt></ctxt>	Non-linguistic events or information to be described

Technology). It should be noted that some of these tags are not considered well-formed XML, as described further in Section 6. Figures 1 and 2 show the first three tag types as listed in NICT's online manual, and Table 1 lists tags of the fourth type.

4.3. How has been the corpus provided?

4.3.1. The NICT JLE Corpus Analysis Tool: Analyzer

Originally, access to the NICT JLE Corpus was only available on a CD-ROM that accompanied the book *Nihonjin 1200 Nin No Eigo Speaking Corpus [L2 Spoken Corpus of 1200 Japanese Learners of English]* (Izumi et al.). In 2012, the TXT files were made available for free online (National Institute of Information and Communication Technology). The CD-ROM contains "The NICT JLE Corpus Analysis Tool" (i.e. Analyzer) with corpus data such as "LearnerOriginal" (i.e. learner data including written interview transcripts of 1,281 interview-

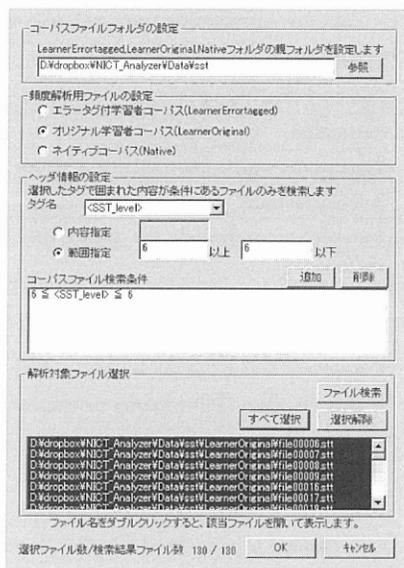


Fig. 3. Analyzer specifying the division of the corpus (Level 6 of LearnerOriginal).

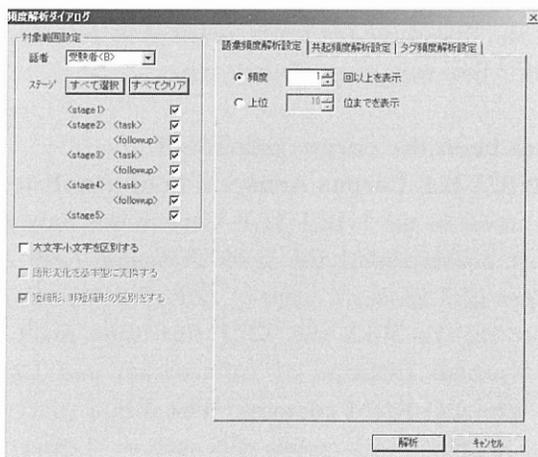


Fig. 4. Analyzer specifying the divisions of the file.

ers and interviewees), “LearnerErrortagged” (i.e. error-tagged learner data), and “Native” (i.e. the data of 20 native speakers taking the SST). These data may only be analysed with Analyzer; they cannot be downloaded as TXT files. Analyzer allows researchers to make a word list of the particular part of the corpus which can be segmented by annotation tags. Figure 3 shows part of the user interface. In this sample window shot, the researcher has specified the corpus version “LearnerOriginal” and a particular part of the corpus (130 files annotated as Level 6), which is segmented by a tag showing the SST Level in the header of each file. Figure 4 shows the interface for generating a word list by specifying the learner (e.g. Interviewee B) and stage (i.e. Stages 1 to 5, including “task” and “follow-up”).

Analyzer can easily segment the corpus according to the annotation tags. Then, it generates frequency results and concordance lines depending on the stage or learner proficiency levels, which can also be downloaded as CSV files. Figure 5 shows the concordance lines of “I mean” from the learner data tagged as Level 6. Table 2 shows the distribution of learners (i.e. interviewees), types, and tokens for each level of proficiency based on Analyzer. It should be noted that the types provided are not lemmatised.

However, there are several drawbacks to using Analyzer. First, it is not possible to mark up elements other than tags already annotated in the corpus because the data can only be accessed using this tool. Hence, there is no way to amend the corpus data by annotating additional markup, split into the data into several segments, or even transform the corpus to be analysed by other tools. Next, it is possible to create a word list but not to generate a lemmatised token. Finally, although anyone can download the corpus data from the accompanying CD-ROM and install the software onto his or her own computer, the book is now out of print.

4.3.2. The NICT JLE Corpus Data Provided as Text Files on the Website

In October 2012, NICT began providing the NICT JLE Corpus on

its website (National Institute of Information and Communication Technology). The corpus data are the same as published by Izumi et al. Now, instead of files that can only be processed by Analyzer, the corpus offers each interview transcript in a TXT file. Therefore, researchers can access the data directly and add more tags for further analyses. However, as each file in the corpus contains both an interviewer and interviewee and all the stages of the interview script, it is not possible to retrieve a word list or concordance lines by specifying the data according to speaker tags or stage tags using general concordancers such as AntConc. If researchers wish to focus on specific parts or stages of the interviews in the corpus, the data have to be automatically segmented by other tools. However, there is still a problem with automatic segmenting, as the XML files are not well-formed. Section 5 describes the rules of valid XML files, and Section 6 shows how the corpus data were converted into well-formed XML.

5. The Rules of XML files

XML is defined as “an extensible markup language used for the description of marked-up electronic text” (Sperberg-McQueen and Burnard 13). Markup language means a set of markup conventions used for encoding texts. The purpose of XML is to allow different kinds of processing to “be carried out with the same part of a file” (Sperberg-McQueen and Burnard 14). XML “ensures the documents encoded according to its provisions” (14) and can move “information from place to place, even between different software products and platforms” (Goldfarb and Prescod 6) without loss of information.

Therefore, an XML document can be transported into and processed by any programme without any transformations or translations if it is well-formed. There are three simple rules for writing an XML document. First, there should a single “element” encoded with a start- and end-tag, which is known as the “root element” (Shibano 56). Second, “the tags marking the start and end of each element must always be present” (Goldfarb and Prescod 17). Then, elements should not partially overlap with one another. For example, Shibano writes that

the street name “Rue Slater Street” written in French and English in Canada should be marked up as “<fr>Rue </fr><en><fr>Slater </fr>Street</en>” (59). The markup “<fr>Rue <en>Slater </fr>Street</en>” is not allowed because the tags overlap; that is, the start-tag <en> is opened within <fr></fr> but closed outside it. The uppermost elements (i.e. parent node) should always contain the lower elements (i.e. child nodes).

Once an XML document is confirmed to be well-formed, it is called a “valid document”, and a document which states the criteria for successful validation is known as document type declaration (DTD) or an XML schema (Sperberg-McQueen and Burnard 17–18; see also Shibano 66–67; Goldfarb and Prescod 15–16, 40, 42). An element can have attribute values. “Attribute-value pairs” can be found inside the start-tag “<poem id='P1' status='draft'>”, where “the value part must always be given inside matching quotation marks, either single or double” (Sperberg-McQueen and Burnard 22). On the other hand, an end-tag does not contain an attribute value specification, as illustrated by “</poem>” (22).

According to various websites that list XML rules, such as W3C (Bray, Hollander, and Layman), there are also several rules regarding elements:

- i. The first character should be “_” (underscore), “:” (colon), or one-byte English letter or Japanese letter (except for one-byte kana characters and two-byte English letters or numbers).
- ii. Characters such as a one-byte number, “.” (full stop), “-” (hyphen), and letters with accent symbols should be used from the second letter.
- iii. One-byte kana characters, two-byte English letters or numbers, and two-byte spaces cannot be retrieved.
- iv. Reserved keywords such as “xml” cannot be used.

6. Data Cleansing with the NICT JLE Corpus

To convert the NICT JLE Corpus into well-formed XML format, we performed automatic modification in Perl and manual modification

Table 3 Tags automatically converted into well-formed XML

Original Tag	Modified Tag	Notes
<.></.>	<pause duration="long"></pause>	The tags showing pauses, <.> and <..>, are combined. To distinguish the length of pause, attribute values are added.
<..></..>	<pause duration="short"></pause>	
<?></?>	<scripting unclarity="partly"></scripting>	The tags showing how confident the transcriber is in the transcription, <?> and <??>, are combined. To distinguish the degree of confidence, attribute values are added.
<??></??>	<scripting unclarity="all"></scripting>	
<SC></SC>	<SC unclarity="none"></SC>	The tags showing self-correction, <SC> and <SC?>, are combined. To distinguish the degree of the transcriber's confidence in the transcription, attribute values are added.
<SC?></SC?>	<SC unclarity="partly"></SC>	
<R></R>	<R unclarity="none"></R>	The tags showing repetition, <R> and <R?>, are combined. To distinguish the degree of the transcriber's confidence in the transcription, attribute values are added.
<R?></R?>	<R unclarity="partly"></R>	

Table 4 Overlapping tags (between speakers) manually converted into well-formed XML

Original Tag	Modified Tag	Number of Modified Tags
<CO></CO>	<CO segment="inter">	48
<R unclarity="none">	<R unclarity="none" segment="inter">	8
<SC unclarity="none">	<SC unclarity="none" segment="inter">	50

matically modified to include attributes, the tags below were manually modified and checked for validity in Chrome. For example, the original version of “file00074” had an overlapping tag “repetition”. Chrome indicated that there was an “error on line 189 at column 110: Opening and ending tag mismatch: R line 0 and B”. Figure 6.1 shows the original data, in which the start-tag <R> in line 189 does not have the end-tag </R> before the end-tag of speaker . Instead, the end-tag

```

189 <B><F>Mhm</F> I watched "Life Is Beautiful". <F>Mhm</F>. And on Sunday, <R><OL>I went
to</OL></B>
190 <A><OL><F>Uh-huh</F></OL>.</A>
191 <B><R>Yokohama</R> I went to Yokohama.</B>

```

Fig. 6.1. The original “file00074” showing the overlapping tags between speakers.

```

189 <B><F>Mhm</F> I watched "Life Is Beautiful". <F>Mhm</F>. And on Sunday, <R unclarity="none"
segment="inter"><OL>I went to</OL></R></B>
190 <A><OL><F>Uh-huh</F></OL>.</A>
191 <B><R unclarity="none" segment="inter">Yokohama</R> I went to Yokohama.</B>

```

Fig. 6.2. The modified version of “file00074”.

Table 5 Overlapping tags (single speaker’s utterance) manually converted into well-formed XML

Original Tag	Modified Tag	Number of Modified Tags
<CO>	<CO segment="intra">	1
	<OL segment="intra">	3
<SC>	<SC unclarity="none" segment="intra">	8
<R>	<R unclarity="none" segment="intra">	9

appears inside the next speaker tags but without the start-tag in line 191. Therefore, the end-tag in the first utterance of Speaker B in line 189 and start-tag in the second utterance in line 191 were added as in figure 6.2.

There are also tags which overlap other tags within the same speaker’s utterance, as shown in table 5. For example, as shown in figure 7.1, the original version of “file00287” is not valid, as Chrome showed an “error on line 40 at column 33: Opening and ending tag mismatch: SC line 0 and OL”. The tags “overlap with other speakers” and “self-correction” overlap in the utterance of Speaker B. Figure 7.2 shows that not only was the start-tag “self-correction” modified with additional attributes in line 40, but also the start- and end-tags were inserted after .

```
40 <B><OL><SC>it's</OL> kind of</SC> <F>well</F> it's nice and safe. <nvs>laughter</nvs></B>
41 <A><F>Uhhh</F>. <OL><CO>I</CO></OL>.</A>
```

Fig. 7.1. The original version of “file00287” showing overlapping tags within Speaker B’s utterance.

```
40 <B><OL><SC unclearness="none" segment="int ra">it's</SC></OL> <SC unclearness="none"
segment="intra">kind of</SC> <F>well</F> it's nice and safe. <nvs>laughter</nvs></B>
41 <A><F>Uhhh</F>. <OL><CO>I</CO></OL>.</A>
```

Fig. 7.2. The modified version of “file00287”.

```
And <></> <R>two</R> two wòman are talking. <F>Mm</F>. And <></> <F>uhm</F>
<laughter><R>bo</R> <SC>boys</laughter> <F>uhmm</F> playing</SC> boys are playing volleyball.
```

Fig. 8.1. The original line 97 from “file00141” showing overlapping tags within the same utterance.

```
And <pause duration="short"></pause> <R unclearness="none">two</R> two woman are talking. <F>Mm</F>.
And <pause duration="short"></pause> <F>uhm</F> <laughter><R unclearness="none">bo</R></laughter> <SC
unclearness="none"><laughter>boys</laughter> <F>uhmm</F> playing</SC> boys are playing volleyball.
```

Fig. 8.2. The modified version of line 97 from “file00141”.

Next, the original corpus contains 16 “laughter” tags that overlap either between different speakers’ utterances or within the same utterance. The modification of “file00141” illustrates how this problem was solved. Chrome displayed the message “error on line 97 at column 509: Opening and ending tag mismatch: SC line 0 and laughter” in a part “<SC>boys</laughter> <F>uhmm</F> playing</SC>”, as shown in figure 8.1. The end-tag </laughter> after “bo</R>” and start-tag <laughter> before “boys” were added, as shown in figure 8.2. While “self-correction” is supposed to mark certain lexical items (in this case, “boys uhmmm playing”), “laughter” is an additional non-linguistic annotation that occurs simultaneously with the utterances of lexical items. In the process of modification, only the first end-tag and the second start-tag of “laughter” were inserted to resolve the problem of overlapping tags.

6.2.2. Provision of Missing Tags

There were four tags missing in the original corpus that were added in during the modification process. They are shown in table 6.

6.2.3. Correction of Erroneous Symbols

Erroneous symbols were found in the original corpus, as table 7 shows. First, two-byte characters or spaces were corrected to one-byte ones. Second, irrelevant symbols were deleted.

6.2.4. Modification of the Wrong Order of End-Tags

Eighty-five parts in the original corpus which had the wrong order of end-tags, as in the example of line 60 from “file00008” in figure 9.1.

Table 6 Modification of missing tags

Missing start-tag	<A>	file00062	line 37
		file00003	line 47
		file00794	line 112
		file00963	line 183

Table 7 Modification of erroneous tags

Two-byte characters	I	file00081	line 69
	space	file01165	line 157
	space	file01165	line 157
Irrelevant symbols	.	file00743	line 14
	.	file01270	line 14
]	file00848	line 36

```
<SC>very</SC> totally different from Japan, so I really like that place even though it's <SC>a
<laughter>very</SC></laughter>
```

Fig. 9.1. The original line 60 from “file00008” showing the wrong order of end-tags.

```
<SC>very</SC> totally different from Japan, so I really like that place even though it's <SC unclearness="none">a
<laughter>very</laughter></SC>
```

Fig. 9.2. The modified version of line 60 from “file00008”.

Figure 9.2 shows the corrected order of the end-tags “laughter” and “self-correction”.

7. Application of Well-Formed XML Files in the Analysis of Longer Stretches of Discourse in the NICT JLE Corpus

7.1. Research on Request Strategies in Speech Acts

As described in Section 3, researchers now have direct access to the data and markup annotations of the NICT JLE Corpus, thanks to NICT providing TXT files. Pragmatic annotations were manually added to the segmented parts of the valid XML files in the corpus, following the established rules of XML format (Miura, *Criterial Features; The NICT JLE*). Her studies attempt to investigate request strategies in speech acts in the extracted learner data of role-play sessions that involve transactions to obtain goods, negotiations for refunds, or item exchanges. In this role-play (segmented by the tag <stage3> and </stage3>), the interviewer plays a shop assistant or train staff member, while the interviewee is given the role of a customer or passenger. Figure 10 shows an example from “file00001” with added annotations of request strategies (shown in bold). The annotation was done according to the coding scheme developed in the area of cross-linguistic pragmatics (Blum-Kulka; Blum-Kulka, House, and Kasper; Trosborg; Salgado).

As figure 10 shows, the *head act* (i.e. the core of the request sequence) of request strategies was first identified and tagged as <HA></HA>.

```

121 <A>Hello. May I help you, miss?</A>
122 <B><F>Er</F> yes. <F>Mmm</F> <HA><RQ dmc="s"><DR str="desire"><R unclarity="none">I
    want to</R> <SC>I want to</SC> <SC>I want</SC> <F>mm</F> sorry, <F>mm</F> <pause
    duration="short"></pause> I want to <F>err</F> watch.</DR></RQ></HA></B>
123 <A><F>Uhm</F>.</A>
124 <B><F>Um</F>.</B>
125 <A>Yes.</A>
126 <B><F>Um</F> and <F>mmm</F> <HA><RQ dmc="s"><ID>I prefer <F>mm</F> leather
    watch.</ID></RQ></HA></B>
127 <A><F>Uhm</F>.</A>
128 <B><F>Uhm</F> and <F>mmm</F> <HA><RQ dmc="h"><ID><SD mkr="intrg"><R
    unclarity="none">do you have</R> <F>mm</F> do you have something special
    one?</SD></ID></RQ></HA></B>

```

Fig. 10. An excerpt from “file00001” with annotations of request strategies.

Then, whether the dominance of the requestive perspective is on the speaker or hearer was identified (<RQ dmc="s"> or <RQ dmc="h">). There are three types of request strategies: *direct*, *conventionally indirect* and *non-conventionally indirect* strategies. In this example, the direct request strategy is expressed with a pattern of desires ("I want/need X") and annotated as <DR str="desire"> in line 122. Then, head acts of request in lines 126 and 128 were identified as an indirect strategy, annotated as <ID>. The syntactic downgrader as interrogative as internal modification was also identified as <SD mkr="intrg">.

After manual tagging of these pragmatic features, annotated features were retrieved using Perl. The Perl script was written to retrieve not only the search tags, but also the whole utterance of a particular line where the search tags appeared, as well as the line number and file identification number. This made easier to examine the neighbouring contexts of the target pragmatic features, as we could retrieve longer stretches instead of a KWIC (Key Word In Context) concordance with a limited number of words provided by general concordancers. It was also possible to check the preceding and following utterances of the learners and their interactions with the interviewers, by detecting the place where the target features were produced according to the given line numbers.

In summary, the valid XML documents of the NICT JLE Corpus allow us to expand the scope of analysis of pragmatic features in longer stretches of discourse, rather than restricting the scope to word frequency or simple concordance lines; this means lexical and grammatical analyses are no longer limited to surface forms, as when Analyser or general concordancers are used. The details of coding scheme and results of analyses drawing on different proficiency levels and tasks given role play sessions can be found in Miura's studies (*Criterial Features; The NICT JLE*). It is also reported that the development and availability of spoken learner corpora not only give new insights into the development of tools or coding schemes for analysing the relationship between lexico-grammatical features and discourse functions, but also allow us to re-examine the results of the previous stud-

ies on interlanguage pragmatics.

7.2. Word List Based on Well-Formed XML Files in Comparison to Analyzer

Now the valid XML documents are available for the NICT JLE Corpus, the originally designed programme can retrieve its tokens and lemmatised types of learner data at different proficiency levels. The following programme was made to generate a word list and frequency information.

- i. Elements marked up by speaker tags “A” and “B” were extracted respectively.
- ii. Elements annotated by tags such as “F”, “H”, “ctxt”, and “nvs” were excluded and those annotated by “laughter”, “R”, “CO”, and “OL” were counted.
- iii. Elements segmented by a space, comma (,) and question mark (?), and exclamation mark (!), double quotation (“ ”), and semicolon (;) were identified as lexical words. In this case, upper case and lower case letters were not distinguished.
- iv. Tokens and types for proficiency level and speaker were counted. *E_lemma.txt* (Ver1.1), which was compiled by Yasumasa Someya (Izumi et al.), was used to lemmatize the retrieved types. Inflected verbs and pluralised nouns were converted into dictionary forms in the list, but the part-of-speech information were not considered.
- v. Genitive forms, contracted forms, or hyphenated words were not deconstructed, and counted as single lexical words.

Table 8 shows a comparison between the distribution of learners, types, and tokens for each level of proficiency retrieved from Analyzer and XML files using the Perl programme. It should be noted that Analyzer cannot generate lemmatised types. The differences in the tokens and types of the present study from those of Analyzer may be due to the way of counting of contracted forms, genitive forms, hyphenated words, and words annotated as hidden information, and the process of lemmatisation. We aim to improve the precision of the

Table 8 The Distributions of Learners, Types, and Typees for Each Level in the NICT JLE Corpus

SST Level	Analyzer's Tokens	Analyzer's Types Types (not lemmatised)	Tokens in the Present Study	Lemmatised Types in the Present Study
1	413	208	411	238
2	7,654	1,259	7,498	1,408
3	95,494	4,670	94,290	4,946
4	308,477	7,410	306,243	7,540
5	204,617	5,893	203,146	5,669
6	132,885	5,034	130,287	4,611
7	87,574	3,953	85,018	3,582
8	70,404	3,607	68,349	3,220
9	56,118	3,264	54,251	2,840

programme for generating a word list and frequency in the future.

8. Conclusion

The automatically and manually converted 1,281 files into well-formed XML format have made it possible to amend the corpus data with any tools in a way that researchers wish. Especially, pragmatic analyses can be more easily done with valid XML versions. In order to investigate instances of requestive speech acts in the role-play stage of learner data, it was highly useful to retrieve a word list or frequency information from the segmented corpus data according to the annotation tags, for example, stages of the interview, proficiency levels of learners, turns of interviewers and interviewees (i.e. learners). Interactive spoken data with abundantly annotated extra-linguistic elements of the XML-formatted version of the NICT JLE Corpus allow researchers to expand their analysis from surface forms to pragmatic functions in longer stretches of discourse.

WORKS CITED

- Adolphs, Svenja. *Corpus and Context: Investigating Pragmatic Functions in Spoken Discourse*. Amsterdam: John Benjamins, 2008. Print.

- Blum-Kulka, Shoshana. "Indirectness and Politeness in Requests: Same or Different?" *Journal of Pragmatics* 11 (1987): 131–46. Print.
- Blum-Kulka, Shoshana, Juliane House, and Gabriele Kasper. *Cross-cultural Pragmatics: Requests and Apologies*. Norwood, NJ: Ablex, 1989. Print.
- Bray, Tim, Dave Hollander, and Andrew Layman, eds. "Namespaces in XML." *XML Schema*. W3C, 14 January 1999. Web. 8 July 2014.
<<http://www.w3.org/TR/1999/REC-xml-names-19990114>>.
- Goldfarb, Charles F., and Paul Prescod. *The XML Handbook*. Upper Saddle River, NJ: Prentice Hall, 2011. Print.
- Izumi, Emiko, Kiyotaka Uchimoto, and Hitoshi Isahara, eds. *Nihonjin 1200 Nin No Eigo Speaking Corpus*. [L2 Spoken Corpus of 1200 Japanese Learners of English]. Tokyo: ALC, 2004. Print.
- Kasper, Gabriele, and Shoshana Blum-Kulka. "Interlanguage Pragmatics: An Introduction." *Interlanguage Pragmatics*. Ed. Gabriele Kasper and Shoshana Blum-Kulka. Oxford: Oxford University Press, 1993, 3–17. Print.
- Kasper, Gabriele, and Kenneth R. Rose. *Pragmatic Development in a Second Language*. Oxford: Blackwell, 2002. Print.
- Knight, Dawn, and Svenja Adolphs. "Multi-modal Corpus Pragmatics: The Case of Active Listenership." *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Ed. Jesús Romero-Trillo. Berlin: Mouton de Gruyter, 2008. 175–90. Print.
- Miura, Aika. "Criterial Features of Pragmatic Competence in a Spoken Corpus of Japanese Learners of English to Profile Different Levels of Proficiency." Learner Corpus Research Conference (LCR 2013). Solstrand Hotel & Bad, Bergen. 27 Sep. 2013. Address.
- Miura, Aika. "The NICT JLE (Japanese Learner English) Corpus No XML Seikeishika To Sore Wo Tsukatta Shutokubetsu No Goyoronteki Gengo Tokuchou No Bunseki [Constructing Well-Formed XML Files of the NICT JLE Corpus and Its Application to Research into Pragmatic Competence Across Different Levels of Proficiency]." The 39th Conference of the Japanese Association for English Corpus Studies. Tohoku University, Sendai. 5 Oct. 2013. Address.
- National Institute of Information and Communication Technology. *The NICT JLE (Japanese Learner English) Corpus*. Information Analysis Laboratory. Web. 15 Feb. 2014.
<http://alaginrc.nict.go.jp/nict_jle/index_E.html>.
- O'Keeffe, Anne, Brian Clancy, and Svenja Adolphs. *Introducing Pragmatics in Use*. Abingdon: Routledge, 2011. Print.
- Romero-Trillo, Jesús. "Introduction: Pragmatics and Corpus Linguistics—a Mutualistic Entente." *Pragmatics and Corpus Linguistics: A Mutualistic Entente*. Ed. Jesús Romero-Trillo. Berlin: Mouton de Gruyter, 2008. 1–10. Print.
- Salgado, Elizabeth Flores. *The Pragmatics of Requests and Apologies: Developmental Patterns of Mexican Students*. Amsterdam: John Benjamin Publishing Company, 2011. Print.
- Schauer, Gila A. "Pragmatic Awareness in ESL and EFL Contexts: Contrast and Development." *Language Learning* 56 (2006): 269–318. Print.
- Shibano, Koji. *SGML/XML Ga Wakaru Hon [SGML/XML]*. Tokyo: Ohmsha, 2000.

Print.

Sperberg-McQueen, C. M., and Lou Burnard, eds. *Guidelines for Electronic Text Encoding and Interchange: TEI P4*. Oxford: Published for the TEI Consortium by the Humanities Computing Unit, University of Oxford, 2002. Print.

Takahashi, Satomi. "Pragmatic Transferability." *Studies in Second Language Acquisition* 18 (1996): 189–223. Print.

Trosborg, Anna. *Interlanguage Pragmatics: Requests, Complaints and Apologies*. Berlin, NY: Mouton de Gruyter, 1995.

An Analysis of Three Collocations Dictionaries for Learners of English

YASUTAKE ISHII SATORU UCHIDA
KYOKO HAYASHI YUICHIRO KOBAYASHI

1. Introduction

This paper is a brief critical review of three collocations dictionaries for learners of English with special emphases on the differences in their headwords, collocates and microstructures, as well as how they are perceived by their users.

A “collocation” of a given word refers to its most typical word combinations, and those typical words the given word combines with are called “collocates”¹⁾. The mastery of collocations is known to be essential for sounding natural in any language.

The following three dictionaries, given in the order of publication, are reviewed in this paper.

- *Oxford Collocations Dictionary*, Second Edition (2009) (“OCD2” henceforth): The number of pages for the A-Z body part is 958. A CD-ROM accompanies each copy of the print dictionary, and some entries are offered only on the CD-ROM version. On the CD-ROM, in addition to regular searching for headwords, we can also search for collocates given under other headwords, which can be of benefit to users. The print dictionary states about the CD-ROM that “30,000-word index makes searching easy” (back cover), although the number of index keys is 22,864²⁾ according to our counting. The first edition of this dictionary came out in 2002 without a CD-ROM, and its electronic version was released as part of a standalone CD-ROM, *Oxford Phrasebuilder Genie* in

2003.

- *Macmillan Collocations Dictionary* (2010) (“MCD” hereafter): It has 911 pages in the body. There is no electronic version of this dictionary.
- *Longman Collocations Dictionary and Thesaurus* (2013) (“LCD” below): The main body has 1,447 pages. There is no CD/DVD-ROM version of this dictionary, but a one-year subscription to the online version is provided to those who have bought a copy. Some additional entries, collocates and examples are available on the online version.

According to the publishers’ catalogues (Oxford University Press 2013; Macmillan Education 2013; Pearson ELT 2013), the target level of users of all three dictionaries is B2-C2 on the CEFR, Common European Framework of Reference for Languages: Learning, Teaching, Assessment, which is a framework that describes the proficiency level of a learner’s foreign language ability, wherein the levels are divided into six stages from A1, beginner, to C2, the most advanced.

This paper consists of four sections excluding this introduction and the conclusion in Section 6. Section 2 compares the headwords in the three dictionaries. Section 3 focuses on the microstructures of each dictionary. Section 4 highlights the differences in collocates across the three dictionaries based on a sampling study. Section 5 deals with the users’ perspective by showing the results of our user survey.

(Section 1 by Ishii)

2. Headwords¹⁾

This section looks at the headwords in each dictionary, and compares the items across the three dictionaries. For this purpose, we made a complete list of headwords for each dictionary. We made the lists for both print and electronic versions except for MCD, and they include those items that are present either in the print edition or in the electronic version or in both. The following numbers and discussions are based on the lists we have created.

2.1. Number of headwords in each dictionary

The number of headwords in OCD2 is 8,418. Empty headwords for cross reference are excluded from this number, and words which have two or more parts of speech are counted separately; for example, **abuse**²⁾ (n)³⁾ and **abuse** (v) are counted as two items, which is how all three dictionaries give their entries. The print edition of OCD2 states that it includes “[c]ollolocations for 9,000 nouns, verbs and adjectives” (back cover), which could be judged to be dishonest according to our count⁴⁾. 356 out of 8,418 items are given exclusively in the electronic version.

MCD has 4,305 headwords. There are no empty headwords in this dictionary, so this number is the actual total size. The dictionary states that it has “collocations for over 4,500 carefully-selected key words” (back cover), so this dictionary can also be criticized for exaggerating the count assuming that our count is correct.

LCD presents 3,819 headwords which include 36 items available only in the electronic version that do not give collocates but thesauri and/or other notes. 741 empty headwords for cross reference are excluded from this number. 550 (besides the above 36) out of 3,819 items are given only in the electronic version. We were not able to find any mention of the number of headwords neither in the print dictionary nor in the electronic version.

The breakdown of the headwords in terms of parts of speech is given in Table 2.1. All three dictionaries share a similar tendency in terms of the proportion of parts of speech; about 60% of the headwords are nouns while adjectives and verbs share roughly 20% each.

In terms of the numbers of headwords, OCD2 can be praised for its

Table 2.1 Breakdown of Headwords

	OCD2 ⁵⁾	MCD	LCD ⁷⁾
Adjectives	1,486(17.7%)	997(23.2%)	680(17.8%)
Nouns	5,443(64.7%)	2,406(55.9%)	2,483(65.0%)
Verbs	1,489(17.7%)	902 ⁶⁾ (21.0%)	573(15.0%)
Adverbs	—	—	49(1.3%)
(Total)	8,418	4,305	3,819 ⁸⁾

relatively wide coverage; users are more likely to find the headwords they want to look up in this dictionary.

2.2. Comparison of the coverage

The number of all different types of headwords covered in at least one dictionary is 9,168. 26.3% of the items, namely 2,409 headwords, are covered in all three dictionaries, while 27.8% or 2,549 words are included in two and 45.9%, i.e., 4,210 items, are present only in one dictionary. The more detailed breakdown of these numbers in terms of parts of speech is laid out in Table 2.2. About 70% (1,676 out of 2,409) of the headwords covered in all three dictionaries are nouns, which conforms to the expectation that we look up nouns more often than other parts of speech in collocations dictionaries.

It is natural to expect that the widely covered headwords are frequently used words. In order to confirm whether this is the case, we extracted the first ten headwords in alphabetical order from each group of words that are covered in three/two/one dictionaries in our list, and checked their raw frequency in the British National Corpus (BNC)¹¹. The result is shown in Table 2.3. In general, those headwords that are covered in all three dictionaries are observed more frequently than those items covered in two or one dictionary, with a few exceptions such as **able** (adj) and **about** (prep).

Now let us see how these headwords are covered in each dictionary. Table 2.4 summarizes the relationship between the coverage and the number of headwords in each dictionary. Although MCD and LCD fall behind OCD2 in terms of the numbers of headwords, they have

Table 2.2 Breakdown of All Different Types of Headwords Covered in Three/Two/One Dictionaries⁹⁾

	3	2	1
Adjectives	391	579	833
Nouns	1,676	1,445	2,405
Verbs ¹⁰⁾	341	525	891
(Total)	2,409	2,549	4,210

Table 2.3 First Ten Items Covered in Three/Two/One Dictionaries and Their Raw Frequency in the BNC

Covered in 3 Dictionaries		Covered in 2 Dictionaries		Covered in 1 Dictionary	
abandon (v)	4,332	abbreviation (n)	229	abashed (adj)	44
ability (n)	10,378	abnormal (adj)	801	abhorrent (adj)	66
abortion (n)	1,488	abolish (v)	1,864	ablaze (adj)	162
absence (n)	5,859	absent (adj)	1,451	able (adj)	29,657
abstract (adj)	1,860	absorb (v)	2,619	abode (n)	250
abuse (n)	3,664	abstract (n)	370	about (prep)	146,934
academic (adj)	4,612	absurd (adj)	926	abscess (n)	99
accent (n)	1,764	abuse (v)	1,206	absorbed (adj)	92
accept (v)	19,811	abusive (adj)	277	abundance (n)	665
acceptable (adj)	3,608	accelerate (v)	1,064	academic (n)	965

Table 2.4 Relationship between the Coverage and the Number of Headwords in Each Dictionary

	OCD2	MCD	LCD
3	2,409 (28.6%)	2,409 (56.0%)	2,409 (63.1%)
2	2,456 (29.2%)	1,501 (34.9%)	1,141 (29.9%)
1	3,547 (42.1%)	395 (9.2%)	268 (7.0%)
(Total)	8,418	4,305	3,819

relatively few headwords that are unique to only one dictionary. Considering the fact that MCD and LCD give at least one example for each collocation or each set of synonymous or related collocates while OCD2 does not (see 3.3.2 for more details), we can say that MCD and LCD focus on frequent headwords and try to give as many examples as possible. Users are likely to find the headword whose collocates they want to know as long as the headword is frequent enough, while in other cases for more infrequent headwords it is plausible that only OCD2 can offer them a solution.

While there seems to be no special characteristic found in those headwords covered only in OCD2, the items included only in MCD or in LCD have some features that are worth mentioning. The first 10 items of those headwords covered only in MCD in alphabetical order

are as follows: **academic** (n), **accepted** (adj), **account for** (phr vb), **add to** (phr vb), **added** (adj), **adhere to** (phr vb), **adverse** (adj), **advisory** (adj), **allege** (v) and **alleged** (adj). We will soon notice the three phrasal verbs here. The inclusion of phrasal verbs as headwords is a contrastive feature of this dictionary. The items extracted in the same procedure from the LCD are as follows: **about** (prep), **actress** (n), **adult** (adj), **affectionate** (adj), **after** (prep, conj, adv), **again** (adv), **alarm clock** (n), **allergic** (adj), **allow** (v) and **almost** (adv). Prepositions and adverbs are included as headwords only in this dictionary¹²⁾, which is a noticeable feature of this dictionary.

(Section 2 by Ishii)

3. Microstructures

3.1. Introduction

This section examines the microstructures of OCD2, MCD and LCD. The structure of entries in these dictionaries is specially devised to list collocates. The basic framework can be described as follows:

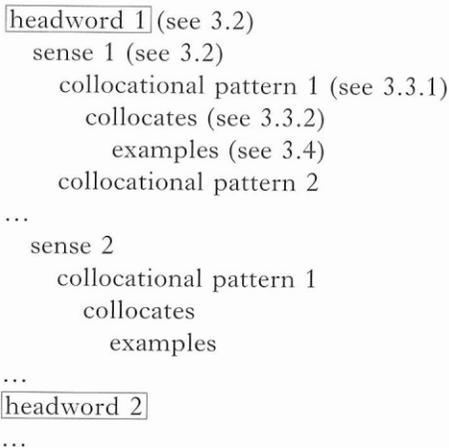


Fig. 3.1, part of an entry taken from the CD-ROM version of OCD2, is a typical example of how each entry looks.

We will begin by looking at the framework of entry structures (3.2), and then discuss how collocational patterns and collocates are pre-

rhyme *noun*

1 using words that have the same sound as each other

ADJECTIVE

- **internal**

RHYME + NOUN

- **scheme**

PREPOSITION

- **in ~**
a story in ~

2 word that has the same sound as another

ADJECTIVE

- **half**

PREPOSITION

- **~ for**
Can you think of a ~ for 'tragic'?

Fig. 3.1. Entry of **rhyme** (n) on the CD-ROM version of OCD2.

sented in the entry (3.3) together with examples (3.4) and labels (3.5).

3.2. The basic structure of entries

Let us start by looking at the main framework of entries, that is, headwords and meaning groups. OCD2, MCD and LCD employ the same policy for splitting headwords; they separate entries based on word classes. For example, **review** (n) and **review** (v) are listed as separate entries. Unlike ordinary learner dictionaries, pronunciations of headwords are not presented in the target dictionaries¹⁾.

Within each entry, headwords with more than one sense are divided into different parts. For example, as can be seen in Fig. 3.1, each sense of **rhyme** (n) (“1 using words that have the same sound as each other” and “2 word that has the same sound as another”) is treated separately. In such cases, short definitions of each sense are given so that users can understand which sense of the word goes with the collocations listed under the sense.

There are some differences as to the policy for the definitions.

Unlike MCD and LCD, OCD2 does not provide a definition when entries have only one sense, which can be problematic. Take **compatible** (adj) for example, which has the following three senses according to OALD8:

1. (of machines, especially computers) able to be used together
2. (of ideas, methods or things) able to exist or be used together without causing problems
3. if two people are compatible, they can have a good relationship because they have similar ideas, interests, etc.

At the entry of **compatible**, OCD2 lists collocates such as *highly*, *very*, *entirely*, *fully*, *perfectly* and *mutually*. However, users may wonder which of the above senses can be used with these collocates without the specification of the sense of the headword.

LCD clarifies its policy for ordering each sense of polysemous words as follows: "Meanings of the word are listed in frequency order" (iv). However, OCD2 and MCD do not clearly state their policies in this respect, which could cause user-unfriendliness.

3.3. Collocates

3.3.1. Collocational patterns

Collocates are categorized into several groups according to their syntactic patterns. The basic patterns in each dictionary are as follows²⁾:

OCD2

verbs' collocates: adv, v + V, prep, phrases

nouns' collocates: adj, n of N, v + N, N + v, N + n, prep, phrases

adjectives' collocates: v, adv, prep, phrases

MCD³⁾

verbs' collocates: adv + V, V + n, and/or

nouns' collocates: adj + N, n + N, v + N, N + v, N + n, and/or

adjectives' collocates: adv + ADJ, ADJ + n, v + ADJ, and/or

LCD⁴⁾

verbs' collocates: adv, prep, phrases, n

nouns' collocates: adj, adj/n + N, n + N, N + n, v + N, N + v, v, prep, prep/adv, phrases

adjectives' collocates: n, ADJ + n, n + ADJ, adv, v, prep, phrases

While OCD2 and MCD list collocational patterns in the same order in every entry, LCD seems to show collocational patterns in the order of frequency. This results in the following differences in the order of collocational patterns within LCD (print edition):

charter (n): v, phrases, prep

chat (n): adj, v, prep

chocolate (n): phrases, adj/n + **chocolate**, **chocolate** + n

This shows, for example, that “adj + **chat**” is more frequent than “v + **chat**” and “**chat** + prep.” Also, we can tell that the “phrases” category is the most frequent for **chocolate**, which lists *a bar of chocolate*, *a piece of chocolate* and *a box of chocolates*.

Let us now look closely at each pattern. First, a note should be made about pre-modifying categories. OCD2 includes many nouns with adjectives because such nouns “function like adjectives” (x). MCD distinguishes nouns and adjectives that come before nouns. LCD employs a moderate policy: in some cases nouns and adjectives are separate while in other cases they are treated together⁵. The differences can be seen in the entries of **review** (n) as the following excerpts indicate:

review (n) (sense 2) in OCD2

adj: *good, bad, book, film*

review (n) (sense 2) in MCD

adj + N: *good, bad*

n + N: *book, film*

review (n) (sense 1) in LCD

adj/n + **review**: *a thorough/comprehensive/full review, peer review*

review (n) (sense 2) in LCD

adj: *a good/bad review*

n + **review**: *a film/movie review*

We can say that the policy employed in OCD2 is practical, since there are some cases where such distinction is difficult, which might be a motivation for LCD's ambiguous treatment. It should be noted that

there seems to be confusion in MCD's treatment as *management committee* and *cotton curtain* are treated under the label of "adj + N," though they are more likely to be categorized as the "n + N" pattern.

As Coffey (2010: 334–335) points out, it is one of MCD's characteristic features that it includes noun collocates in verb and adjective entries. LCD also gives noun collocates in its verb and adjective entries. Unlike these two dictionaries, OCD2 excludes noun collocates from verb and adjective entries based on the assumption that a user would not look up noun phrases there⁶. This may hold true for many users, but this can also lead to a serious disadvantage as a collocations dictionary. To take **injure** (v) 'to harm physically' for example, OCD2 (sense 1) lists only adverbs such as *badly*, *seriously* and *severely*. MCD, in contrast, includes not only adverbs such as *badly* and *critically* (adv + V) but also nouns that indicate "people involved in a traffic accident" such as *cyclist*, *driver* and *motorcyclist* (V + n), and also nouns for "body parts" such as *ankle*, *knee* and *wrist* (V + n). Including such nouns undoubtedly contribute to clarifying the verb's usage. Also, it makes it possible to include less frequent nouns that are not treated as headwords such as "cyclist" and "motorcyclist." OCD2 does not include them as headwords, so the collocations of "**injure** + cyclist" and "**injure** + motorcyclist" have no chance to appear in this dictionary.

Another advantage of MCD is that it consistently includes "and/or" category⁷. Let us look at some examples from MCD:

age (n) (sense 1)

and/or: *ability, background, ethnicity, gender, nationality, occupation, ...*

angry (adj)

and/or: *aggressive, annoyed, irritable, depressed, sad, ...*

From the collocates for the entry of **age**, for instance, we can assume that the coordination pattern of **age** is often used in the context of job requirements in such phrases as *age and ability* and *age and background*. Also, information for the entry of **angry** enables us to describe somebody's anger in a more precise way. Thus, the pattern of "and/or" can be of great use for encoding purposes.

The final point to be considered as to the collocational patterns is about “phrases.” OCD2 and LCD have this category and put miscellaneous phrases here. For example, OCD2 includes *stand accused of* in the phrases category of **accuse** (v). This category makes it possible to list expressions that do not fall into any standard collocational patterns but are worth including. MCD does not use this category and tries to make its patterns as clear as possible. However, MCD’s policy may also cause a lack of information. For example, phrases such as *sb’s nerves are on edge* (=they feel slightly nervous or worried) (LCD) and *in your mind’s eye* (=your imagination) (OCD2) are certainly useful for learners, but they are not included in MCD due mainly to the lack of this category.

3.3.2. Groups of collocates

Collocates are grouped according to their meanings in each dictionary. In OCD2 and MCD, words are listed in alphabetical order within each group. OCD2 states that “[t]he groups are arranged in an order that tries to be as intuitive as possible” (viii). For example, adjective collocates for the entry of **behaviour** (n) are grouped as “*exemplary, good | acceptable | normal | bizarre, strange, suspicious | ...*” (“|” indicates group boundaries) and the order of groups seems to be intuitive as the list starts from positive collocates and then followed by neutral and negative words; notice also that words are listed alphabetically within each group. However, each semantic group is not given any explanations in OCD2⁸). By contrast, MCD gives a brief meaning for each group, though the policy for the order of the groups is left unexplained. An example below is taken from the entry of **child** (n), where, again, collocates appear in alphabetical order within each set.

adj + N

of a particular age: *adolescent, little, newborn, preschool, school-aged, small, teenage, young*

with problems: *abused, at-risk, disabled, disadvantaged, neglected, special-needs, underprivileged, vulnerable*

badly behaved: *badly behaved, delinquent, difficult, naughty,*

problem, spoiled

This strategy seems to be more user-friendly than OCD2's policy. However, as Coffey (2010: 339) points out, meaning descriptions do not always successfully function as sign posts. The above example includes such a case; the definition "badly behaved" is just a repetition of the first collocate.

LCD states in its preface that "[c]ollocations are listed in order of frequency, so that you can see the most common collocations first" (iv). Compare the following examples taken from the "v + N" sections in the entries of **cigarette** in each dictionary:

OCD2

smoke | draw on, pull on, suck on | light | extinguish, put out, stub out | roll | flick, flick away | bum | advertise

MCD

light or smoke a cigarette: *draw on, light, puff on, smoke*
finish a cigarette: *extinguish, finish, put out, stub out*

LCD

smoke a cigarette
light a cigarette
put out a cigarette also *extinguish a cigarette* formal (=stop it burning)
pull on/drag on/draw on a cigarette (=smoke a cigarette with deep breaths)
roll a cigarette (=make your own cigarette using special paper)

LCD's entry tells us that the most frequent collocation is "smoke + **cigarette**," and also "pull on + **cigarette**" is the most common collocation in the meaning of 'smoking with deep breaths.' Rather than simply giving words in alphabetical order, this is far more informative for users. One possible problem of this policy is that similar words may appear separately. For example, at the entry of **citizen** (n), *a good citizen, a respectable/decent citizen* and *a model citizen* are listed separately⁹), making it difficult to recognize the semantic closeness of these expressions.

Another advantage of LCD is that "[t]he meanings of difficult collocations are explained in brackets" (iv). This means that semantic

annotation is not given to the group of collocates but to the specific phrases as can be seen above (the paraphrase starts with “=”). This policy also makes it possible to annotate the meanings of idiomatic or semantically opaque phrases, or of genre-specific phrases such as *peer review*, which is annotated as “(=in scientific and other studies, the examination of someone’s work by other scientists, researchers etc).” Since the meaning of the phrase is not always the sum of each word, this surely helps users understand the meaning and the usage of the phrase.

3.4. Examples

Just like other learner dictionaries, the collocations dictionaries have examples in each entry. MCD and LCD give sentence examples whereas OCD2 has both sentence and phrase examples. Some random examples are given below:

clue (n) [prep] ~ (*as*) *to*: a ~ as to her whereabouts (OCD2)

departure (n) [v + N] *delay*: She’d heard that I’d had to delay my departure. (MCD)

citizen (n) [adj] *a good citizen*: The education system is designed to produce good citizens. (LCD)

LCD has the richest examples; almost every collocation is accompanied by an example. By contrast, MCD provides only one example for most semantic groups and OCD2 offers examples to only a small part of semantic groups; both dictionaries leave many of their collocations unexplained and are less attractive in this respect. Obviously, having more examples is an advantage especially for encoding purposes (see 5.2.2 and 5.2.3).

3.5. Labels

OCD2 and LCD employ various labels. The following are the labels listed on each dictionary’s inside front cover and some random examples:

OCD2

[related to attitudes/situations] disapproving, figurative, formal,

historical, humorous, informal, ironic, literary, offensive, old-fashioned, saying, slang, technical
 [related to subject areas] biology, business, computing, economics, finance, law, etc.
 [related to regions and institutions in particular countries] (esp.) AmE, (esp.) BrE, in Australia, in England and Wales, etc.

call (v) (sense 3) [adv] *free* (BrE), *toll-free* (AmE)

self (n) [adj] *good* (humorous, esp. BrE)

LCD

Ac (used to show that a word is on the Academic Word List), AmE, BrE, formal, informal, spoken, written, literary, technical, old-fashioned, disapproving, humorous

accident (n) [v] *be involved in an accident* formal

eye (n) [v + N] *avert your eyes* literary

OCD2 notes that “[l]abels used in the dictionary generally apply to the *collocation*—that is to the two words in combination—and not to the individual words” (inside front cover). To take “do drugs” as an example, “neither *do* nor *drugs* is informal in itself” (ix) but as a phrase it is assigned an “informal” label. The same rule is true in LCD, which states, “Labels tell you if the collocation is only used in formal or informal English” for the use of “formal” label (iv).

MCD does not mention its policy on its use of labels in the preface or elsewhere, but it uses “informal” labels as can be seen in the following example:

recipe (n) (sense 1) [adj + N] *yummy* informal

However, compared to OCD2 and LCD, MCD is insufficient in this respect. Specifically, no attention is paid to regional varieties, which results in a serious defect of the dictionary. For example, according to OCD2 and LCD, “have a **guess**” is a Britishism whereas “take a **guess**” is an Americanism, but this kind of information is not found in the entry of **guess** in MCD which simply gives *have, hazard, make* and *take* as “v + N” collocates.

3.6. Summary

The microstructure of each dictionary is, in general, similar to each other, but there are some differences as Table 3.1 summarizes. It can be said that LCD is advantageous in many respects and, as far as the microstructure is concerned, this is one of the most powerful tools for users to produce natural English. (Section 5 will show that most of our survey participants actually preferred LCD.)

Since collocations dictionaries are based on corpora as stated in the preface in each of our target dictionaries (see also 4.2), it is not difficult to add frequency information (as in LCD), and register and genre information (as in OCD2 and LCD). Those dictionaries that do not take advantage of these kinds of information can be improved by using their corpora more extensively in future editions. Also, it might be

Table 3.1 Summary of Differences in the Microstructure

		OCD2	MCD	LCD
Definitions of Headwords		only for polysemous words	✓	✓
Collocational Patterns	Noun Collocates at Verb and Adjective Entries		✓	✓
	“phrases”	✓		✓
	“and/or”		✓	
	Frequency Order			✓
Order of Collocates		alphabetical	alphabetical	frequency
Meaning Descriptions of Collocations		occasional	for each semantic group	for semantically opaque phrases
Examples		for part of semantic groups	for each semantic group	for each phrase
Labels	Register	✓	“informal” only	✓
	Geographical	✓		✓

possible to add more notes for learners using learner corpora¹⁰). The recent trends toward creating huge corpora will surely contribute to improving collocations dictionaries in the not-so-distant future.

(Section 3 by Uchida)

4. Comparison of collocates given in three collocations dictionaries

This section compares the collocates given under several same headwords in each dictionary especially in terms of their frequency. We will show the quantitative summaries of the collocates of five sampled headwords (4.1), and give an overview of those collocates' frequency in large corpora to verify whether highly frequent collocates are thoroughly included in our target dictionaries (4.2).

The five sample headwords considered in this section were selected taking the following steps:

1. We picked the first headwords, excluding empty ones only for cross reference, starting with each alphabet on the print edition of LCD whose coverage is the smallest among the three dictionaries: **abandon** (v), **baby** (n), **cab** (n), **dam** (n), **ear** (n), and so on.
2. Only those headwords covered in all three dictionaries were kept. This step excluded seven items, and the 17 remaining words include **abandon** (v), **cab** (n), **dam** (n), **face** (n), **gadget** (n), etc.
3. Following the rough proportion of the word classes in the three dictionaries—20% for adjectives, 60% for nouns and 20% for verbs (see Table 2.1)—we decided to pick one adjective, three nouns and one verb. We had only one adjective (**ugly**) and one verb (**abandon**) after the second step above, and they were automatically chosen as part of our samples.
4. To select three nouns out of 15, we decided to choose those headwords which have only one sense in the paper version of LCD for ease of comparison, which narrowed the number to 11. Next, we decided to use the three most frequent items in the BNC. However, the most frequent among the 11 words, **face**, is divided into two senses in OCD2 where collocates of this word are given in

separate sections. They are given collectively under one sense in MCD and LCD, so we avoided this word, and finally chose the nouns **habit**, **qualification** and **wage**.

4.1. Comparison of the number of collocates across the three dictionaries

To compare the coverage in different dictionaries, we first collected all the collocates given under the headwords **abandon** (v), **habit** (n), **qualification** (n), **ugly** (adj) and **wage** (n) in all three dictionaries. We checked both print and electronic versions for LCD and OCD2 and included those items that are only in the electronic versions. For **abandon**, we only looked at the sense of ‘stop doing/supporting something’ because this is the only sense covered in MCD. Likewise, we only checked the ‘credentials’ sense for **qualification**. For **ugly**, we focused on the ‘unattractive’ sense given as sense 1 in LCD for brevity’s sake. The summaries of the number of collocates are given in Tables 4.1–4.5.

From the quantitative summaries in Tables 4.1–4.5 alone, no particular dictionary can be said to offer more collocates consistently. More extensive research needs to be carried out in order to see how wide a range of collocates each dictionary offers.

Table 4.1 Number of Collocates Given for **abandon** (v)

	OCD2 (sense 2)	MCD	LCD ¹⁾ (sense 2)
abandon + n		8	10
adv	18	12	9
v + abandon	3		2
phrases			1
abandon + prep	2		
(Total)	23	20	22

Table 4.2 Number of Collocates Given for **habit** (n)

	OCD2	MCD	LCD
adj/n + habit	61	21	20
v + habit	19	13	10
habit + v	1		
prep + habit	3		2
phrases	5	4 ²⁾	3
(Total)	89	38	35

Table 4.3 Number of Collocates Given for **qualification** (n)

	OCD2 (sense 1)	MCD	LCD
adj/n + qualification	21	41	15
qualification + n	1		
v + qualification	8	21	6
prep	2		4
and/or		7	
(Total)	32	69	25

Table 4.4 Number of Collocates Given for **ugly** (n)

	OCD2 (sense 1)	MCD (sense 1)	LCD (sense 1)
ugly + n			16
adv + ugly	8	10	5
v + ugly	3		1
phrases			2
(Total)	11	10	24

Table 4.5 Number of Collocates Given for **wage** (n)

	OCD2	MCD	LCD
adj + wage	31	11	10
v + wage	23	15	11
wage + v	3		3
wage + n	32	8	10
phrases	8		6
(Total)	97	34	40

4.2. Frequencies of collocates

We have checked the frequencies of each collocate, except for multi-word items, in the BNC and the Corpus of Contemporary American English (COCA). Part of the result is laid out in Tables 4.6–4.7. In these tables “✓” signifies that the collocate is given in the dictionary in question. The “BNC/COCA Rank” shows the rank of the collocate among the possible collocates in the specified syntactic pattern; for example, “attempt(s)” is the third most frequent item when we search for the nouns within the span of 1–3 words to the right of “abandon” and its inflectional forms in the BNC. The sign “—” means the collocate is used less than five times in the corpus.

The results of our survey shows that the three collocations dictionaries extensively cover frequent collocates, with only a few exceptions such as “claim(s)” for “**abandon** + n,” which is the seventh most frequent item in the BNC. This result may seem quite reasonable because all of these dictionaries are corpus-based and will conform to our expectation that they offer the users frequent collocates of a given word, but we have to remember that there is no dictionary that covers

Table 4.6 Coverage and Frequencies of Collocates for **abandon** + n

	OCD2	MCD	LCD	BNC Rank	COCA Rank
<i>attempt</i>		✓	✓	3	22
<i>belief</i>		✓		25	69
<i>decision</i>			✓	—	—
<i>effort</i>			✓	16	5
<i>faith</i>		✓		41	40
<i>hope</i>		✓	✓	5	9
<i>idea</i>		✓	✓	1	2
<i>plan</i>		✓	✓	2	3
<i>policy</i>			✓	6	13
<i>project</i>			✓	12	10
<i>pretence</i>		✓		15	56
<i>principle</i>		✓	✓	11	16
<i>search</i>			✓	28	44

Table 4.7 Coverage and Frequencies of Collocates for *v* + **habit**

	OCD2	MCD	LCD	BNC Rank	COCA Rank
<i>acquire</i>	✓	✓		10	16
<i>adopt</i>	✓	✓		16	18
<i>alert</i>		✓		—	—
<i>become</i>	✓		✓	6	7
<i>break</i>	✓	✓	✓	7	10
<i>change</i>	✓	✓	✓	5	3
<i>cultivate</i>	✓	✓		—	47
<i>develop</i>	✓	✓	✓	8	6
<i>establish</i>	✓			12	20
<i>form</i>	✓		✓	11	19
<i>have</i>	✓	✓	✓	1	2
<i>kick</i>	✓	✓	✓	9	8
<i>make</i>	✓		✓	4	4
<i>overcome</i>		✓		—	109
<i>quit</i>		✓		—	58
<i>support</i>	✓			—	10

all collocates found in other dictionaries. Also, while it is true that their policy to offer highly frequent collocates for frequent headwords makes sense, the users of collocations dictionaries are likely to use them for encoding purposes more often than for decoding. Considering this fact, users are likely to face problems when they want to know the collocates of infrequent words which will be covered only on a small scale, if any, or when they want to freely express a variety of ideas that may not necessarily be covered within the frequent collocates extracted from corpora. In other words, all our target dictionaries focus on very frequent headwords and collocates, but this could also be a limiting factor for learners.

(Section 4 by Ishii and Kobayashi)

5. Users' perspective

In this section we will highlight the users of collocations dictionar-

ies. Most of us will agree on the effectiveness of collocations dictionaries for learners, but it does not necessarily mean that every learner can use these dictionaries effectively. We conducted a questionnaire survey to make clear how properly learners can use collocations dictionaries and how learners find this special kind of reference work.

5.1. Participants and questions

The participants of our survey are 64 lower intermediate non-English majors at a Japanese university. They are familiar with the notion of collocations in their English classes and take a quiz on collocations in every class.

This survey consists of three sections: A, B and C. In section A, the participants were presented color photocopies of relevant entries in three dictionaries indicated as Dictionary X/Y/Z which corresponds to LCD/MCD/OCD2 respectively, and were asked to write an appropriate expression which is equivalent to a Japanese phrase by referring to the entries. The headwords whose entries our participants had to check to write their answer were carefully chosen; we made sure that the headword and the collocate which was supposed to be an answer were in all three dictionaries. We asked our participants to carefully read all the relevant entries from three dictionaries even if they knew the answer without checking them. Then, the participants were required to decide which dictionary helped most when they answered the question. Also, they were requested to choose or freely write the reason(s) why they thought the dictionary they chose was most helpful. The choices for Questions A1-3/A2-3/A3-3 are based on the result of a previous study (Kawamura and Ishii 2013: 51-52). The original questions and choices are written in Japanese, but their English translations are as follows:

- A1-1: Refer to the entries of **exam** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'I need to retake this exam.': I need to _____.
- A1-2: Which dictionary was most helpful to answer the question

above? (multiple choices allowed)

- a. Dictionary X
- b. Dictionary Y
- c. Dictionary Z
- d. No dictionary was helpful, and I cannot find the right answer.

A1-3: Choose the reason why you found that dictionary helpful. (multiple choices allowed)

- a. The examples are useful.
- b. The amount of information is appropriate.
- c. A meaning is given for each phrase or set of collocates, such as "(=look at different websites)" for *surf the **internet*** in Dictionary X and "have or use the Internet" for the set of collocates *access, be connected to, be linked to, connect to, go on, link to* and *use* for **the Internet** in Dictionary Y.
- d. It is easy to read in terms of the use of boldface, italics, line breaks, colors, and so on.
- e. Others: Write the reason.

A2-1: Refer to the entries of **Internet/internet** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'I ordered the book over the Internet.': I ordered the book _____.

A2-2: (the same as A1-2)

A2-3: (the same as A1-3)

A3-1: Refer to the entries of **health** in each dictionary and write a correct expression in English equivalent to a Japanese expression meaning 'Stress affects our mental health.': Stress affects our _____.

A3-2: (the same as A1-2)

A3-3: (the same as A1-3)

In section B of our survey, we asked the participants to choose the features of the dictionary in question which they find superior/inferior to those of the other dictionaries. The choices are based on the results

of a previous study (Kawamura and Ishii 2013: 52–55) and common for all the following questions:

- B1: Choose the features of Dictionary X you find superior to those of the other two dictionaries. (multiple choices allowed)
- B2: Choose the features of Dictionary X you find inferior to those of the other two dictionaries. (multiple choices allowed)
- B3: Choose the features of Dictionary Y you find superior to those of the other two dictionaries. (multiple choices allowed)
- B4: Choose the features of Dictionary Y you find inferior to those of the other two dictionaries. (multiple choices allowed)
- B5: Choose the features of Dictionary Z you find superior to those of the other two dictionaries. (multiple choices allowed)
- B6: Choose the features of Dictionary Z you find inferior to those of the other two dictionaries. (multiple choices allowed)
 - a. The number of collocates is (not) appropriate.
 - b. A meaning is (not) given for each phrase or set of collocates. (present in Dictionaries X and Y)
 - c. The number of examples is (not) appropriate.
 - d. It is (not) easy to read in terms of the use of boldface, italics, line breaks, colors, and so on.
 - e. Others: Write the feature.
 - f. Nothing in particular.

Finally, in section C, the participants were asked the following question:

- C: Which improvements would make collocations dictionaries easier to use? (multiple choices allowed)
 - a. more collocates
 - b. more examples
 - c. English explanations for proper choice of collocates within each set of collocates
 - d. frequency of each collocate
 - e. Japanese translations or explanations

- f. Others: Write your suggestion.

5.2. Results

5.2.1. Results for section A

The percentages of correct answers to each question are given in Table 5.1. The result for Question A1–1 is very poor (10.9%) mainly because many participants chose a wrong determiner and wrote “retake an exam” while the Japanese expression reads “retake this exam.” The plausible reason for this error is that they found “*retake an exam* (also *resit an exam* BrE)” in LCD and just copied it. If we allow this minor error and focus on the correct choice of a verb collocate, then 60.9% of the participants gave a correct answer, which is indicated with an asterisk in Table 5.1. The results show that more than half of the participants answered the questions correctly in Questions A2–1 and A3–1, and chose a correct verb in Question A1–1. Also, almost all participants wrote a correct part-of-speech word even if they were wrong; we can assume that they at least understood which part of the dictionaries they had to look at. However, considering the fact that the answers are present in all three dictionaries, these percentages should be higher. We need to recognize that not a small number of our participants cannot properly use these collocations dictionaries for some reasons, part of which will be considered below.

The results for questions asking the participants which dictionary was most helpful to answer the questions A1–1/A2–1/A3–1 are summarized in Table 5.2. Note that only those who were correct in the previous question were counted, and those who chose a correct verb but used a determiner that didn’t match the Japanese phrase given are counted separately and shown with an asterisk. (The same policy applies to Table 5.3.) As these results clearly show, most of the participants chose LCD as the most helpful among the three dictionaries.

Table 5.3 summarizes the features which the participants chose as the reason(s) why they thought the dictionary of their choice was the most helpful. The results indicate that many of our participants felt helped by the examples and easiness to read the dictionary. On the

Table 5.1 Percentages of Correct Answers for Each Question

Question	Answers Treated as Correct	Percentage of Correct Answers
A1-1	retake/resit this/the exam	10.9% / 60.9%*
A2-1	on/over/via the Internet/internet	87.5%
A3-1	mental/emotional/psychological/spiritual health	64.1%

Table 5.2 Dictionary Found Most Helpful (Multiple Choices Allowed)

Question	Z [OCD2]	Y [MCD]	X [LCD]	None
A1-2 (retake this exam)	0 (0%)	1 (14.3%)	6 (85.7%)	0 (0%)
A1-2 (retake an exam)*	1 (3.1%)	1 (3.1%)	31 (96.9%)	0 (0%)
A2-2 (over the Internet)	3 (5.4%)	2 (3.6%)	53 (94.6%)	2 (3.6%)
A3-2 (mental health)	1 (2.4%)	8 (19.5%)	34 (82.9%)	2 (4.9%)

(The percentages are of those who gave a correct answer.)

Table 5.3 Reasons for Finding the Dictionary Helpful (Multiple Choices Allowed)

Question	Examples	Amount of Information	Meanings	Easy to Read	Others
A1-3 (retake this exam)	6 (85.7%)	1 (14.3%)	1 (14.3%)	6 (85.7%)	0 (0%)
A1-3 (retake an exam)*	13 (40.6%)	6 (18.8%)	7 (21.9%)	20 (62.5%)	2 (6.3%)
A2-3 (over the Internet)	31 (55.4%)	14 (25.0%)	9 (16.1%)	40 (71.4%)	4 (7.1%)
A3-3 (mental health)	21 (51.2%)	15 (36.6%)	9 (22.0%)	25 (61.0%)	2 (4.9%)

(The percentages are of those who gave a correct answer.)

contrary, our participants did not indicate that the meanings for each set of collocates were such an important factor when they found the answers, which is a little surprising because we expected this feature found in MCD and LCD to be highly helpful to users.

Table 5.4 Strengths and Weaknesses in Each Dictionary (Multiple Choices Allowed)

	Num. of Collocates	Meanings	Num. of Examples	Easy to Read	Others	Nothing
Strengths of LCD	37	27	33	61	10	0
Weaknesses of LCD	20	6	7	1	3	37
Strengths of MCD	33	15	18	3	1	21
Weaknesses of MCD	20	11	7	52	7	5
Strengths of OCD2	33	9	12	6	1	25
Weaknesses of OCD2	19	16	19	52	6	3

(The total number of participants is 64.)

5.2.2. Results for section B

Table 5.4 shows which features of each dictionary were judged to be superior/inferior to the others by our participants. In general, LCD is deemed to be more useful than MCD and OCD2 in many ways, but what is the most striking difference across these dictionaries is that LCD was praised for being much easier to read than the others. We can easily notice that there are at least three advantages in how the information is presented in LCD. First, there is a line break after each item, which yields more white space. Although this also holds true of MCD, the sets of collocates in MCD are larger than the items in LCD in most cases, which makes MCD more crowded with words and with less white space. Second, LCD uses larger font sizes than the other two dictionaries. And third, the syntactic pattern such as “ADJECTIVES” and “VERBS” stand out well in entries. These characteristics give us the impression that this dictionary is much less packed and make it easier to look for the information we want to obtain than in MCD or OCD2. However, it must be kept in mind that in this survey the photocopies of print editions were used, and if the CD-ROM ver-

sion of OCD2 had been offered, the criticism would have been less severe.

As for the number of collocates, there seems to be no big difference in the users' evaluation of the three dictionaries. This may be, however, due to the nature of our survey; appropriate collocates are included in all the dictionaries. Regarding the meanings of each set of collocates, about a quarter to a third of the participants answered that they are useful, although this feature was not regarded to be so important when they tried to find a correct collocate (see Table 5.3). Concerning the number of examples, it seems many users feel that they will benefit if more examples are offered (see also Table 5.5 below).

5.2.3. Results for section C

Table 5.5 presents the result for the question: "Which improvements would make collocations dictionaries easier to use?" As for the numbers of collocates in our target dictionaries, our participants seem satisfied with them, although this could be due to the fact that they didn't experience the lack of the items they were looking for in our survey. On the other hand, around a third to half of the participants answered the other four improvements—more examples, English explanations for proper choice of collocates within each set of collocates, frequency of each collocate and Japanese translations or explanations—would be beneficial, although the last would be impossible in monolingual dictionaries. All of these four possible improvements are related to choosing the best collocate, and they must be related to the not-so-high percentages of correct answers for questions A1-1/A2-1/

Table 5.5 Features Which Our Participants Felt Would Make Collocations Dictionary Easier to Use (Multiple Choices Allowed)

More Collocates	More Examples	English Explanation for Proper Choice of Collocates	Frequency	Japanese Translations/ Explanations	Others
9	28	29	25	26	16

(The total number of participants is 64.)

A3-1. Most of the participants who chose "Others" wrote that the easiness to read is the most important factor.

5.3. Summary

Through our survey, we have revealed three important aspects, among others, in collocations dictionaries from the viewpoint of users. First, the easiness to read weighs importantly with the users. It is hard especially for lower intermediate learners to find an appropriate collocate and check examples written in a foreign language on a cramped page. Second, many users rely on examples to find appropriate collocates. Third, the meanings for each set of collocates in LCD and MCD are helpful to not a small proportion of users. These can be seen as keys for improvement in potential future editions or for their potential competitors.

(Section 5 by Hayashi and Ishii)

6. Conclusion

This paper briefly reviewed and compared three collocations dictionaries for intermediate to advanced learners of English. In Section 2, we compared their headwords giving the actual numbers and showed the advantage of OCD2 in terms of the number of headwords while MCD and LCD have their own characteristic features. Section 3 dealt with the microstructures and compared their different policies especially on the arrangement of collocates, on examples and on labels. Section 4 focused on the collocates and confirmed that in general all three dictionaries offer highly frequent collocates based on corpora. In Section 5, we described how these dictionaries are used by lower intermediate learners and what aspects they find useful or needing improvement. Through our review, it has become clear that every collocations dictionary has its own advantages compared to other competitors, but at the same time they leave room for improvement to be more user-friendly to a wider range of learners.

NOTES

Section 1

- 1) LCD does not differentiate between “collocations” and “collocates” and collectively calls them “collocations.”
- 2) There are two entries of **peep** (v) in the electronic version, one of which has an additional example. The figure of 22,864 is the result of counting the two entries as one item.

Section 2

- 1) We would like to express our gratitude to Seijo University for funding the research for this section through the Grant-in-Aid for Creative Research in 2013–2014.
- 2) Headwords are written in bold face in this paper.
- 3) Different dictionaries employ different policies on how to represent parts of speech. In this paper, we use the simplest type of notations such as “(n)” and “(v)” for all three dictionaries.
- 4) The number of empty headwords is 104, and even if we include them, the total number is 8,522.
- 5) The number for adjectives includes 11 “adj., adv.,” 14 “adj., noun” and one “adj., pron.,” and the number for nouns includes one “noun, exclamation.”
- 6) 113 phrasal verbs are included in this figure.
- 7) The number for adjectives includes 11 “adj, adv” and 19 other similar items that are given more than one part of speech. The same process of lumping together has been carried out for nouns and adverbs as well.
- 8) This number includes 34 other items including five items labeled as “conjunction” and several other types.
- 9) For those items that have more than one part of speech, we used the first part of speech for our counting; for example, an item classified as “adj., noun” was dealt with as an adjective.
- 10) Phrasal verbs in MCD are counted as verbs.
- 11) We used BNC*web*’s “Frequency lists” function.
- 12) OCD2 has 11 headwords labeled as “adj., adv.”

Section 3

- 1) The IPA transcriptions and audio clips for headwords are available on the online version of LCD. Although it is desirable to include phonetic information of the headwords (and also of the collocates), this policy can be justified considering the purpose of collocations dictionaries, i.e., serving learners who want to write naturally.
- 2) For the sake of simplicity, the same notation is used for all three dictionaries. Capitalized letters indicate the position of headwords. N/n, V/v, ADJ/adj and adv stand for noun, verb, adjective and adverb respectively.
- 3) MCD has some other patterns such as “V + with,” “ADJ + in” and “n + of + N.”
- 4) LCD also has some other minor patterns such as “COLOUR” and “SHAPE/POSITION” for **eye** (n).
- 5) LCD also includes predicative adjectives in this category. For example, at the entry

of **prediction** (n), the following expressions can be found: *an accurate prediction*, *a prediction is correct/right*, *a prediction is wrong/incorrect* and *a reliable prediction*.

6) The preface states as follows: “The third question asked (Would a student look up this entry to find this expression?) led to the exclusion of noun collocates from verb and adjective entries” (vi).

7) OCD2 and LCD’s “phrases” category sometimes gives phrases of this pattern; *angry and frustrated/upset* is given in the entry of **angry** in LCD.

8) For some collocations and idiomatic phrases, OCD2 gives a simple paraphrase: “*the burden of proof* (=the responsibility of proving that sth is true),” for instance.

9) The latter two collocations are included only in the electronic version.

10) LCD provides such notes in some entries. For example, at the entry of **research** (n) we can find a note “Don’t say ‘make research’.”

Section 4

1) The number of collocates, not items, are counted; for example, the item **abandon an attempt/effort** is treated as giving two noun collocates. The only exception is the “phrases” category where each item is counted as one phrase; *change/break the habits of a lifetime* is counted as one item, for instance.

2) These four items are given as phrases in the note at this entry; MCD states that it uses notes “when there is a common way of expressing the same idea using a phrase rather than a collocation” (xii).

DICTIONARIES REFERENCED

- LCD: *Longman Collocations Dictionary and Thesaurus*. Harlow: Pearson Education, 2013.
- MCD: *Macmillan Collocations Dictionary*. Oxford: Macmillan Education, 2010.
- OCD2: *Oxford Collocations Dictionary*. 2nd ed. Oxford: Oxford University Press, 2009.
- OALD8: *Oxford Advanced Learner’s Dictionary of Current English*. 8th ed. Oxford: Oxford University Press, 2010.

OTHER WORKS CITED

- Coffey, Stephen. “A New Pedagogical Dictionary of English Collocations.” *Internal Journal of Lexicography* 24 (2010): 328–341.
- Kawamura, Akihiko, and Yasutake Ishii. “Komyunikeeshon Nouryoku Ikusei no Tame no Goi Shidou—Poraitonesu to Korokeeshon no Kanten kara—[Developing Students’ Vocabulary Knowledge for Better Communication Skills: With Special Emphasis on Politeness and Collocation].” *Social Innovation Studies* 8.2 (2013): 37–68.
- Macmillan Education. *Macmillan Education English Language Teaching Catalogue 2013*. 2013.
- Oxford University Press. *Oxford English Language Teaching Catalogue 2013*. 2013.
- Pearson ELT. *The Pearson ELT and Penguin Readers Catalogue 2013*. 2013.

The Motivation for Using Detached Participial Clauses in English¹⁾

HITOSHI KAWABATA

1. Introduction

Sentences like (1) are commonly used in written English:

- (1) Driving down a back road on Cape Cod on a July evening three years ago I saw an unwelcome blue flash in my rearview mirror.¹⁾ (*Newsweek*, April 16, 2012, p. 28)

In this study, the term “detached participial clause”²⁾ (hereafter DP) is used for the underlined clause. Although this term is not typically used, its functional neutrality—i.e., it does not imply an adverbial or adjectival function—is favorable for the scope of this study, which includes so-called non-restrictive participial clauses such as the underlined clause in (2), where the participial clause modifies the noun phrase adjectivally:

- (2) ..., Serb forces had effectively erased history as they destroyed Bosnia’s National Library, built during the waning days of the Austro-Hungarian Empire. (*Newsweek*, April 9, 2012, p. 19)

The DP conveys various semantic meanings, but it is not always clear which meaning a DP actually expresses. In spite of this semantic ambiguity, DPs are used frequently in journalistic articles,³⁾ which presumably endeavor to report news correctly to readers. This suggests that the motivation for using them derives from other than semantic reasons.

The purpose of this study is to answer the question, “What motivates writers⁴⁾ to use DPs in English?” A spectrum of examples of the

DP have been collected from *Newsweek* and *TIME*⁵⁾ to reduce the possibility of using examples whose usage is specific to a certain writer.

In this study, I argue that the motivation for using DPs comes from construing two situations as “one situation.” By “one situation,” I mean that two situations expressed by a DP and a matrix clause have successivity or instantaneity in conscious experience, except in cases where the two clauses describe two aspects of one situation. Section 2 discusses the temporal relation between the two clauses and shows that the DP does not always have temporal proximity to the matrix clause. In Section 3, I suggest another temporal proximity characteristic of DPs: time in conscious experience. Section 4 treats the writers’ motivation in three groups: (i) cases where the successivity is due to the two events occurring successively in the physical world, (ii) cases where the successivity is due to unconscious processes, and (iii) cases where the two clauses describe two aspects of one situation.

2. Tenseless Feature

Radden and Dirven (2007: 202) write: “Notions of time may be coded in language lexically ... and/or grammatically by one of the tenses When a language has tense, this means that its speakers must express the time of each independent situation.” Although a DP does describe a situation, it is tenseless even when the situation has a specific time of occurrence. A reason for this could be that writers see the situation described by the DP as dependent on the matrix clause and construe the two situations as a single one. If this is the case, what kind of dependence do writers recognize between them? Given the fact that participles are tenseless, the most likely answer is temporal dependence. Tomozawa (2003) and Hayase (2002) insist that any meaning of the DP has simultaneity with the matrix clause. However, this view does not seem adequate. As will be shown below, DPs can describe various kinds of time intervals in relation to the matrix clauses.

2.1. Temporal Interpretation

The first candidate for temporal interpretation of the detached pres-

ent participle is simultaneity with the matrix clause. Take the verb *look*, for example. We can look at something while speaking, eating, walking, or doing most other activities.

- (3) Looking deep into his eyes, she inquires, “Do you mean really, really marry you?” (“TILL DEPOSITIONS DO US PART,”¹¹ *TIME*, December 7, 1997)

When the verb lacks tense in the participial form, as in (3), its nature of doability evokes the interpretation of overlapping simultaneity with the matrix clause. In the absence of other clues, this is the most explicit and accessible source for attaining temporal stability.

The same is true of most past participles, but not all. There are some that are associated with specific points of time in the past when the events are known to have occurred; typically, participles indicating birth or production:

- (4) Born in 1948, he is a thoroughgoing Englishman—donnish in his scholarship and vivid in his writing.
(*TIME*, April 16, 2012, p. 40)

When readers encounter the past participle *born*, they naturally imagine a reasonable, if not exact, time interval between the two clauses; for example, a few decades for an adult. Although in our example there is a temporal adverbial, *in 1948*, the past participle *born* has itself the force to make readers imagine a passage of some length of time. This shows that verbs themselves can also provide clues for understanding time intervals between the two clauses.

Moreover, when an interpretation of overlapping simultaneity is not reasonable from the context, readers will recognize that there is a time interval. In (5), it clearly takes a few seconds to walk from the door to the radio.

- (5) Entering the room, I switched [on] the radio.
(Hayase 2002: 175)

Time intervals are not restricted to such small values as in (5).

- (6) Leaving for Geneva with word only that there would be “a

good chance to get something,” Kennerly found on his arrival that he would be the sole photojournalist allowed to observe Reagan and Gorbachev during their meetings (“A Letter from the Publisher,” *TIME*, December 2, 1985, p. 9)

In (6), he must have traveled by airplane, and the time interval between leaving and arriving is hours. These examples demonstrate that a situational relation between the two clauses can also provide a clue for understanding a time interval between them.

Much longer intervals can be described when temporal adverbials are used. In (7), the time interval between the two situations is months:

- (7) Setting sail for the island in the fall of 1740, he reached his destination in the spring of 1741. (Stump 1985: 97)

With the help of temporal adverbials, DPs can describe various kinds of time intervals to the matrix clauses. The important point is that when DPs contain clues for understanding time intervals, the form of a perfect participle, which will be discussed next, is not necessarily required to show a time interval.

2.2. Perfect Participle

Tomozawa (2003) and Hayase (2002) claim that the notion of simultaneity is valid even for the case of the perfect participle. They argue on the grounds that because it expresses perfective aspect, the result of the event described by it remains at its reference time, which is usually equal to the time of the matrix clause. Yet the interpretation of perfective aspect is not always possible, posing a problem to their claim. Radden and Dirven (2007: 206), for example, state that whereas the present perfect always has aspectual meaning, the other perfect forms (the past perfect and the future perfect) *basically* do not contain aspectuality but only indicate anteriority. The latter is also true of the perfect participle.

In principle, the present perfect is not used “with words that refer to a completely finished period of time” (Swan 2005: 441). For example, *be born* cannot be used in the present perfect form. Mayor (2009: 181)

writes: "Do not say ... 'I have been born'." The DP in the perfect form *having been born* is, however, possible, as in (8), where the perfect participle does not express perfective aspect but only anteriority.

- (8) He is now 73, having been born near Vicenza in northern Italy in 1904, and he was one of the first abstract painters in New York in the 1930s ("Veiled in a Strong White Light," *TIME*, June 6, 1977, p. 41)

Next, let us discuss cases where no temporal adverbial is attached to perfect participles.

- (9) a. Eating a hearty breakfast, we prepared for our long journey.
 b. Having eaten a hearty breakfast, we prepared for our long journey. (all from Quirk et al. 1985: 238)

Quirk et al. (1985: 238) write: "[f]rom [(9a)], we understand that the eating and the preparation took place together, while from [(9b)], we understand that the breakfast preceded the preparation." When a simple participle expresses overlapping simultaneity with the matrix clause, as in (9a), it becomes necessary to employ the perfect form to express a time interval between them. We may say this is the main purpose of using the perfect participle. Example (10) corresponds to this case:

- (10) Having destroyed the evidence, he was confident that he wouldn't be arrested. (Hayase 2002: 169)

If the perfect form is replaced by *Destroying*, the DP temporally overlaps with the matrix clause, because it was possible for him to feel confident that he wouldn't be arrested while destroying it.

Finally, let us look at a case where a perfect participle expresses aspectuality but its reference time is not equal to that of the matrix clause.

- (11) The martini is to modern American literature and lore what mead wine was to Norse sagas or claret to 18th-century English literature. Dorothy Parker was perhaps its leading laure-

ate, having given us the immortal quatrain:

I like to drink a Martini
But only two at the most.
Three I'm under the table,
Four I'm under the host.

(*Newsweek*, March 26 & April 2, 2012, p. 17)

Since her quatrain in (11) is still well known, the proper interpretation is that the perfect participle corresponds to the present perfect: its reference time is the time of writing, or now.²⁾

3. Time in Conscious Experience

3.1. Another Temporal Proximity

We have seen that the DP does not always bear temporal proximity to the matrix clause. However, there may be a problem in this understanding of temporal proximity: we presuppose time in the physical world; i.e., we understand temporal relation in terms of the time when events occurred or states existed.

Take (12), for example. It could be said that the two situations in (12) have temporal proximity in the physical world because they are atemporal. However, the example suggests an additional kind of temporal proximity the DP can indicate.

- (12) Surrounded by gardens and woodland, Stormont [i.e., the parliament buildings of The Northern Ireland Assembly] draws locals who go to walk, breathe the fresh air and enjoy mountain views, wetlands, wildlife, flower beds and beautiful lawns.
 (*TIME*, April 2, 2012, p. 56)

Any reader will feel the connection between the DP and the matrix clause in (12) to be natural. Some may have visited nearby parks to enjoy flowers, while others may have watched TV coverage of famous gardens inundated with visitors. They are so familiar with such scenes that they can relate the two clauses in (12) with almost no time for thinking about the relation. This instantaneity in understanding can be related to the temporal feature of the DP.

Let us consider (12) from the writer's point of view. The writer

might have found that there are many visitors whose purpose is not to visit the parliament buildings but to enjoy the outside area. At the moment she shifted her attention to thinking about the reason, she might have been hit by the idea that the buildings are surrounded by attractive gardens and woodlands. This instantaneity in arriving at reasons suggests another temporal feature of the DP. If consciousness is taken into account, the DP has temporal proximity to the matrix clause in the sense of time in conscious experience.¹⁾

3.2. Successivity in Conscious Experience

Psychologist Baars explains conscious experience by using a “theater metaphor.” Baars (1997a: 41–7) compares the working memory of consciousness to a theater stage. The spotlight on the stage corresponds to the focus of consciousness, or attention. The players and the audience correspond respectively to contents of conscious experience (or “inputs” into consciousness²⁾) and a vast number of unconscious mechanisms. The audience can only see the player in the spotlight. Once the spotlight of attention shifts to one of the inputs, its information is broadcast to unconscious mechanisms. We can only consciously focus our attention on the inputs one by one, whereas the input in the spotlight of attention is processed in parallel by unconscious mechanisms (Baars 1997b: 295). In addition, these unconscious processes are triggered automatically by whatever input comes to the attention of consciousness (p. 305).

Consequently, in the case of one input into consciousness and one reaction of an unconscious process to that input, if the process requires almost no time, paying attention to the input leads to an instantaneous reaction from the process. For example, when people see or hear the words *Roman Holiday*, many of them will immediately be reminded of *Audrey Hepburn*. In this case, the film title is an input into consciousness through outer senses, and the actress’ name, whose linkage to the film is stored in memory systems, is instantaneously raised to consciousness by unconscious processes. This successivity or instantaneity resembles the experience of receiving two sensory inputs successively. The difference is that the successivity of the latter is due

to the two events occurring successively in the physical world, whereas that of the former is due to unconscious processes.

4. Construal as One Situation

4.1. Two Inputs

4.1.1. Two Successive Inputs

To describe what writers were or are seeing is a typical usage of the DP, as in (1). It is true that the two situations described in (1) have simultaneity: when the subject *I* saw a blue flash, he was still driving. In terms of conscious experience, however, it is almost impossible to focus attention on both of the situations at exactly the same time. The writer's attention shifted *successively* from driving, or looking ahead, to a blue flash in the mirror. This is true of (13) where DPs are used for describing what is depicted in a picture:

- (13) His pose—one arm raised, brandishing a palette knife, and the other hanging loosely, holding the palette—is exactly that of the *Apollo Belvedere*, the epitome of refined classicism.

(*Newsweek*, August 15, 2011, p. 46)

In (13), we feel the movement of the writer's attention—from one arm with a knife to the other with the palette. Even in the case of a still picture, where every part of it exists at exactly the same time, we cannot grasp the whole picture instantaneously. The focused area of attention is so small that we cannot see all the details of a scene in front of us at the same time.

4.1.2. Two Situations Considered As Successive

Certain common features can be found in cases where two situations are considered as successive even when there is a time interval between them.

- (14) The collision—Dedmon's headlights glowing on Anderson's shirt, the truck accelerating over a curb, and Anderson disappearing beneath it—was caught on a security camera and later broadcast worldwide (*Newsweek*, April 16, 2012, p. 21)

In (14), the scene is named *collision*, which means that the writer construes the series of events as one event existing as a coherent semantic unit.

There is often a purpose or goal underlying the situations described by a DP and a matrix clause. The two clauses usually express a set of means and purpose, or a set of start and goal.

- (15) Going where no first lady has gone before, Michelle Obama plans to get sweaty with strangers in the East Room.
(*Newsweek*, April 9, 2012, p. 7)

In (15), the DP expressed a means Michelle Obama uses for the purpose of getting sweaty with strangers.

- (16) In the aftermath, the FBI conducted more than 200 interviews, charging Dedmon and two others
(*Newsweek*, April 16, 2012, p. 18)

In (16), it is clear that the FBI conducted interviews with the goal of charging the suspects of a crime. It is rather easy to understand that situations containing a means and a purpose have successivity, in that an event as a means is usually continued or repeated until the purpose is achieved.

Examples (6) and (17) illustrate cases where a DP expresses a starting event for a goal, which is usually described by the matrix clause:

- (17) Leaving the lodge, as they strolled together down a forest path on the way to their car, Nitze passed to Kvitsinsky a typed document. ("The Nitze Approach: Hard Line, Deft Touch," *TIME*, January 31, 1983, p. 14)

Situations that express a start and a goal can have a longer time interval between them; for example, hours in (6). However, something will connect the start and goal, as is clearly shown in (17). After leaving the lodge, they *strolled* until they got to the car. Although they might have stopped once or twice on the way, this activity of *walking* would not come to an end until they reached the goal. In the same way, in (6), the activity of *flying* connected Kennerly's leaving the start point and

his arriving at the goal.

Whenever we do something, we always have a purpose or goal—from an easy one to a time-consuming one. What occurs between the start and the goal is a monotonous activity, a typical one being locomotion: walking, driving, flying, and so on. This monotonous, basic activity consumes various durations of time. However, by shifting attention only to the salient events, events at both its start and end, and cutting off the basic activity, we could view the salient events as occurring successively. Therefore, example (7) could also be viewed as successive, although the time interval is months. Similarly, we might also view our life as a monotonous, basic activity. In this case, birth, academic background, and so on, can be considered to be salient events that most people have in common. This would lead to the discussion on characteristic situations in 4.2.2.

4.2. One Input and One Reaction

4.2.1. Cause-and-Effect Relationship

When the DP and the matrix clause express a cause-and-effect relationship, the two situations are almost inevitably connected based on writers' knowledge. Whether the matrix clause describes a cause or a result largely depends on which of the two follows the main flow of the text. However, in most cases it is possible to understand that the situation described by the DP is an instantaneous reaction to the input described by the matrix clause.

When results already exist as facts, it is natural to think about their causes or reasons from the results.

- (18) Lacking advertising, the Floriade [i.e., a flower festival in Holland] is less well known in the U.S., and the contingent of American visitors is small. (*Newsweek*, April 16, 2012, p. 50)

Lacking advertising might not mean that there were no advertisements, but rather that advertising was not enough to make it well known to the American people. Anyone knows that the more advertisements are placed for something, the better known that thing would be, but it is

difficult to judge at the beginning whether advertising is enough or not to accomplish the goal. However, it is rather easy to judge from the result: if it is not well known, this means that advertising was not enough. Therefore, in (18), *Lacking advertising* can be an instantaneous reaction from unconscious processes.

In the following example, the DP is used on the basis of the information described by the preceding sentences:

- (19) “The campaign is a farce. The candidates aren’t talking about the real subjects like employment or debt because they don’t have solutions,” says Hamid Senni

Sensing the public’s discontent, Sarkozy has in the past few days shifted tack to introduce more substance and less rhetoric. (TIME, April 23, 2012, p. 31)

When the writer noticed that Sarkozy had changed his tack, it might have easily triggered the idea that it was due to the public’s discontent. It is unlikely he heard that Sarkozy had sensed the public’s discontent, but anyone who knows the background may be sure that Sarkozy had.

When DPs are used to express a resultative meaning, they describe what were or are inevitably expected as results from the causes described by the matrix clauses. This does not mean that DPs are used only when consequential situations had or have not yet occurred. If the results already exist as facts, the DPs imply that the results match the inevitable expectations.

- (20) When the global economy heats up, demand for oil rises, boosting the price and encouraging producers to pump more.

(TIME, April 9, 2012, p. 28)

- (21) In recent years, slabs of mortar have fallen off the Colosseum, Rome’s nearly 2,000-year-old amphitheater, endangering tourists.

(TIME, April 2, 2012, p. 48)

In these examples, cause-and-effect relationships are so simple that on reading the matrix clauses, even readers would immediately comprehend ideas that are roughly the same as written in these DPs.

Consequential facts described by DPs often coincide with evaluations—how good or bad they are. Relative evaluations, such as *extremely bad*, can immediately be expected from causal situations. As for absolute evaluations, however, it is almost impossible for causal situations to trigger instantaneous reactions that include hard numbers: for example, an exact number of casualties. In (22), the DP expresses an absolute evaluation by using a hard number:

- (22) The economic crisis doesn't seem to have affected the king. His wealth has doubled over the past five years, making him the world's seventh-richest monarch, with a fortune estimated at \$2.5 billion. (*Newsweek*, April 9, 2012, p. 35)

A case in point is the use of *killing*, which is widely used in articles about accidents, as in (23):

- (23) The worst accident in flying history took place when two packed 747 jetliners collided on a runway on Tenerife in the Canary Islands in 1977, killing 583 people. (“Too Close Encounters,”¹⁾ *TIME*, March 27, 2000)

In (23), what can immediately be expected from the fact described by the matrix clause is that there must have been a considerable number of casualties. In (22) and (23), an instantaneous relative evaluation of the fact described by the matrix clause motivated the writers to use a DP for the description of the result, and then the hard evidence was introduced to describe the result as a fact.

4.2.2. Characteristic Situation

Persons or things normally trigger associated memories. Characteristic attributes of persons or things include, for example, name, information about birth or production, academic background, occupation, and location. When people meet persons or experience things for the first time, they might immediately become interested in learning such characteristic attributes of them. This also happens when people read. Therefore, when the contents of such attributes are to be expressed in sentences, this instantaneity can motivate writers to use DPs.

- (24) In early 2010, she shot the thriller *Dream House* opposite Daniel Craig, also known as the 21st century James Bond.
(*TIME*, April 2, 2012, p. 53)

In (24), the name *Daniel Craig* might immediately remind even some readers that he currently plays James Bond.

- (25) Alsop was a giant in a long-lost era of print journalism. He and Stewart, his younger brother and sometime writing partner, were children of Northeastern privilege. Eleanor Roosevelt was a first cousin. Educated at Harvard and Yale, respectively, the Alsops wrote newspaper prose with Henry James-ian flourish and a self-assertiveness born of noblesse oblige.
(*Newsweek*, March 26 & April 2, 2012, p. 55)

In (25), the underlined DP provides not only the Alsops' characteristic attribute but also a successive event extending from their childhood to the time of their becoming journalists, as we have seen in 4.1.2. In addition, it answers the readers' expectations. The fact that the Alsops were children of privilege and Eleanor Roosevelt was their first cousin is certain to make readers expect that they would have graduated from prestigious universities, and things went as expected. On this point, too, the writer was motivated to use the DP.

Table 4.1 shows the frequency of references to characteristic attri-

Table 4.1 Verbs used to describe characteristic attributes

Clause type	Verbs (Number of tokens)
A. Detached past participle	known(5), built(4), called(4), made(3), published(3), dated(2), founded(2), born(1), created(1), dubbed(1), employed(1), handcrafted(1), held(1), hired(1), launched(1), produced(1), said(1), sited(1), titled(1) —Total: 35 (85%)
B1. Non-restrictive relative clause (Passive voice)	based(1), born(1) —Total: 2 (5%)
B2. Non-restrictive relative clause (Active voice)	call(2), christen(1), rename(1) —Total: 4 (10%)

A: Detached past participle is located right after the NP it modifies.

B1: Relative clause is in passive voice and the antecedent is the subject of its VP.

B2: Relative clause is in active voice and the antecedent is the object of its VP.

butes—name, information about birth or production, academic background,²⁾ occupation, and location—by using detached past participles and non-restrictive relative clauses in the examples collected for this study. From Table 4.1 we find that there is a clear tendency to use detached past participles for the purpose of providing the contents of characteristic attributes of noun phrases.³⁾

4.2.3. Violation

With reference to violation in conscious experience, Baars (1997a: 118) writes: “Violations of expectations ... can cause a part of the unconscious context to become conscious and reportable.” In other words, when an input into consciousness violates contextual consistency or expectation in unconsciousness, the violated content instantaneously surfaces to consciousness. This instantaneity can motivate writers to use DPs for describing the violated content. In respect of consistency, this type is diametrically opposed to those discussed so far, but from the viewpoint of the successivity or instantaneity in conscious experience, there is no difference whatsoever between them.

The DP of a concessive meaning corresponds to this type.

- (26) Discovered almost by accident, the substance has revolutionized medicine. (Quirk et al. 1985: 1271)

It might seem that the DP in (26) provides a characteristic attribute, but the fact described by the matrix clause causes surprise, given the accidental discovery of the substance.

In (26), it is rather easy to understand that the DP describes the violated content. This is because the required presupposition is included in our ordinary knowledge of the world. However, this is not always the case. There are instances when providing violated contents only by means of DPs is not enough for readers to fully understand the required presuppositions, as in (27):

- (27) A recent report in *The New England Journal of Medicine* showed that amantadine, a previously approved medication, might have a new use. Originally prescribed to prevent

influenza during outbreaks, it also accelerates improvement in patients with traumatic brain injury.

(*Newsweek*, April 9, 2012, p. 10)

If the first sentence is not provided, readers might interpret the second one as consistent; i.e., they might think it is common that a medicine for a certain disease is effective for others. However, with the help of the first sentence, which indicates that finding a new use of a medicine is worth a report, they can understand that the situation described by the matrix clause in the second sentence is an unexpected and surprising result.

Not surprisingly, this type of DP seems to be seldom used in practice, at least in journalism. In the 498 examples collected for this study, example (27) is the only case. This is probably because correctly conveying inconsistent situations by using DPs is not as easy as conveying consistent ones.

4.3. Two Aspects of One Situation

Finally, we look now into cases where a matrix clause and a DP describe two aspects of one situation.⁴⁾ This claim is basically the same as those made by other researchers, such as Tomozawa (2003) and Hayase (2002). What is used in a matrix clause is a generalized or schematized expression of the situation described by the DP, which is located at the end of the sentence:

- (28) Rodríguez has tried his best, sending 88 résumés to organizations that might require some economics research, but he hasn't gotten a single offer. (*TIME*, April 16, 2012, p. 24)

When the level of abstraction is raised, the expression of a matrix clause can be figurative or metaphorical:

- (29) Single-aisle jets carrying between 120 and 200 passengers, like the 737, are the sweet spot of the airplane business for both Boeing and its European rival Airbus, generating a large part of their profits. (*Newsweek*, March 26 & April 2, 2012, p. 37)

We may say the characteristic of this type is that the matrix clauses express writers' ways—sometimes unique and special to them—of constructing concrete facts described by DPs. Matrix clauses are closely connected to DPs, so that even in this type there might be successivity or instantaneity: the former trigger the latter, or vice versa.

5. Conclusion

If we presuppose time in the physical world, we come to the conclusion that DPs express various kinds of time intervals relative to the matrix clauses. However, if we take into account time in conscious experience, we can see successivity or instantaneity connecting them.

As I have argued, successivity in conscious experience and constructing two situations as one situation are occasions that can motivate writers to use DPs. However, it does not necessarily mean that writers will always use them in those circumstances. Whether they use DPs or not depends on their authorial choice. They can choose to use subordinate clauses instead in order to give logical explanation. In this sense, we may say that the DP is a style for expressing the successivity in conscious experience directly by words.

What contributes to instantaneous reactions from unconscious processes is a kind of knowledge that is quite easily associated with inputs into consciousness. Therefore, a study on frame-like knowledge is indispensable for deeper understanding of the cause-and-effect relationship expressed by DPs. From this perspective, it is also necessary to compare them with other types of construction that express cause-and-effect relationships.

NOTES

Title

1) This is a shortened and slightly revised version of my MA thesis submitted to Tokyo University of Foreign Studies in January 2014. The chapter on previous studies is omitted, where I review the studies of Hayase (1997, 2002, 2011), Kortmann (1991, 1995), Kubota (2002), Stump (1985), Thompson (1983), Tomozawa (2003), Uchida (2002), and Yamaoka (2005).

Section 1

- 1) In the examples throughout this study, the use of underlines is mine.
- 2) This terminology is used by Thompson (1983).
- 3) Tanaka (1998) investigates the use of DPs in *the New York Times* and remarks that the frequency of DPs is 0.049 (about once in 20 sentences) (p. 106). This data is based on the use of present participles only, and he notes that if past participles were included in the search target, the number of the search results would be three times as large as the above number.
- 4) This study treats only written English because, as is often said, the DP is characteristic of written English rather than spoken English (for example, see Thompson 1983: 45–6).
- 5) For this study, 498 examples of the DP have been collected through reading printed issues of *Newsweek* (August 15, 2011, from March 26 & April 2, 2012 to April 23 & 30, 2012) and *TIME* (from April 2, 2012 to April 23, 2012). Table 4.1 is based on these examples. The examples from *TIME* dated before 2011 have been picked from search results from the “TIME Magazine Corpus” (<http://corpus.byu.edu/time/>) and confirmed by the archives stored in the library of Tokyo University of Foreign Studies.

Section 2

- 1) This article is not in the *TIME* Asia edition.
- 2) Jespersen ([1931]/2007: 66) writes:
 ... in speaking of dead people the preterit is necessary, except when the reference is to the result as affecting the present day. Thus we may say: “Newton has explained the movements of the moon” (i.e. in a way that is still known or thought to be correct, ...

Section 3

- 1) Yamaoka (2005) studies detached present participles in English narratives from the perspective of discourse functions, and makes arguments on the functions of the matrix clause and the DP as follows (pp. 35, 85). From the matrix clause, the existence of the narrator can be felt, because by using a pronoun in the third person as a subject and a verb in the past tense, the cognitive act of the story character is objectified from the outside by the narrator. By contrast, from the DP, the character’s consciousness can be felt because by using a non-finite verb without a subject, the markers indicating the existence of the narrator disappear.
- 2) Inputs into consciousness are categorized into three groups: outer senses (e.g., seeing, hearing, feeling), inner senses (e.g., visual imagery, inner speech, imagined feelings), and ideas (Baars 1997a: 42–3).

Section 4

- 1) This article is not in the *TIME* Asia edition.
- 2) No DPs providing academic backgrounds are found.
- 3) The number of type A clauses is 106 of 179, the total number of detached past participles of all locations. The numbers of type B1 and B2 clauses are 27 and 26 respectively. Therefore, the percentages of type A, B1, and B2 clauses are 67%, 17%, and 16% respectively. It is true that in clause type A, the percentage of verbs used to

describe characteristic attributes (85%) is higher than that of clauses (67%), but it also seems possible to argue that there is a general tendency to use detached past participles instead of non-restrictive relative clauses, so that the result shown in Table 4.1 might be just a reflection of this tendency. To discuss this matter, more detailed investigation into the voice of relative clauses is required.

4) This use of the DPs is usually called exemplification or elaboration. On elaboration in general, Halliday and Matthiessen (2004: 396) write:

In elaboration, one clause elaborates on the meaning of another by further specifying or describing it; . . . The secondary clause does not introduce a new element into the picture but rather provides a further characterization of one that is already there, restating it, clarifying it, refining it, or adding a descriptive attribute or comment.

REFERENCES

- Baars, Bernard J. 1997a. *In the Theater of Consciousness: The Workspace of the Mind*. Oxford: Oxford University Press.
- . 1997b. "In the Theatre of Consciousness: Global Workspace Theory, A Rigorous Scientific Theory of Consciousness." *Journal of Consciousness Studies*, 4(4): 292–309.
- Halliday, M. A. K. and Christian M. I. M. Matthiessen. 2004. *An Introduction to Functional Grammar*. London: Hodder Education, 3rd edition.
- Hayase, Naoko. 1997. "The Role of Figure, Ground, and Coercion in Aspectual Interpretation." In Verspoor, Marjolyn, Kee-dong Lee, and Eve Eliot Sweetser (eds.), *Lexical and Syntactical Constructions and the Construction of Meaning*, pp. 33–50. Amsterdam: John Benjamins.
- . 2002. *Eigo-Kobun no Category Keisei (Formation of Constructional Categories in English)*, Chap. 6, pp. 149–92. Tokyo: Keiso Shobo.
- . 2011. "The Cognitive Motivation for the use of dangling participles in English." In Panther, Klaus-Uwe and Günter Radden (eds.), *Motivation in Grammar and the Lexicon*, pp. 89–106. Amsterdam: John Benjamins.
- Jespersen, Otto. [1931]/2007. *A Modern English Grammar on Historical Principles, Part IV Syntax (Third Volume)*. London: Routledge.
- Kortmann, Bernd. 1991. *Free Adjuncts and Absolutes in English: Problems of Control and Interpretation*. London/New York: Routledge.
- . 1995. "Adverbial Participial Clauses in English." In Haspelmath, Martin and Ekkehard König (eds.), *Converbs in Cross-linguistic Perspective*, pp. 189–237. Berlin: Mouton de Gruyter.
- Kubota, Masato. 2002. "Bunshi-Kobun no Bunmyaku-Izonsei (Context Dependency of the Participial Construction)." *Eigo Seinen (The Rising Generation)*, 148(8): 32–3.
- Mayor, Michael (ed.). 2009. *Longman Dictionary of Contemporary English*. Harlow: Pearson Education, 5th edition.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.
- Radden, Günter and René Dirven. 2007. *Cognitive English Grammar*. Amsterdam: John Benjamins.

- Stump, Gregory T. 1985. *The Semantic Variability of Absolute Constructions*. Dordrecht: D. Reidel Publishing Company.
- Swan, Michael. 2005. *Practical English Usage*. Oxford: Oxford University Press, 3rd edition.
- Tanaka, Kenji. 1998. "The New York Times no Naka no Bunshi-Kobun (The Participial Construction in the New York Times)," *Current English Studies*, 37: 95–106.
- Thompson, Sandra A. 1983. "Grammar and Discourse: The English Detached Participial Clause." In Klein-Andreu, Flora (ed.), *Discourse Perspectives on Syntax*, pp. 43–65. New York: Academic Press.
- Tomozawa, Hirotaka. 2003. "Aspect of the Semantics of the English Participial Construction." *English Linguistics*, 20(2): 493–517.
- Uchida, Mitsumi. 2002. *Causal Relations and Clause Linkage: Consequential Participle Clauses and their Use*. Osaka: Osaka University Press.
- Yamaoka, Minoru. 2005. *Bunshi-Ku no Danwa Bunseki (Discourse Analysis of Participial Clauses)*. Tokyo: Eihosha.

The Grammatical Meanings of the Adjectival Participle in English¹⁾

TAKERU ITOKAWA

1. Introduction

The paper aims to reveal the various grammatical meanings of (pre- and postmodifying) adjectival participles in English, breaking them down into eight patterns: (i) the simple¹⁾ premodifying present participle (a *barking* dog); (ii) the compound premodifying present participle (a *studious-looking* girl); (iii) the simple premodifying past participle (a *closed* door); (iv) the compound premodifying past participle (a *well-written* book); (v) the simple postmodifying present participle (for the time *being*); (vi) the compound postmodifying present participle (a window *overlooking the street*); (vii) the simple postmodifying past participle (the people *concerned*); and (viii) the compound postmodifying past participle (a door *held open for him*).

The traditional explanation of the role of the premodifying participle (across these patterns) is that it conveys a meaning of 'permanence' or 'characteristicness'; conversely, the postmodifying participle is said to convey 'temporariness' or an 'actual action' (cf. Quirk et al. 1985: 1242, 1325, 1326, 1328; Bolinger 1967: 3). However, such explanations are based on vague concepts such as 'permanence' and 'temporariness', into which different shades of meaning seem to me to be classified.²⁾

In section 2, I will present my own hypotheses on adjectival participles, which are distinct from earlier explanations in that I deal with adjectival participles based on grammatical categories inherent in the predicate verb, namely, mood, tense, aspect, and voice.

In section 3, we will look at each of the grammatical meanings that

can be conveyed by adjectival participles, using examples from contemporary British novels.

2. Hypotheses

2.1. My grammatical tools for explaining adjectival participles

The most important point about the traditional explanation of the role of adjectival participles in the previous section is that premodifying participles do not refer to actual actions, whereas postmodifying ones do. To indicate reality is a significant aspect of postmodifying participles, and never to indicate reality a noteworthy aspect of premodifying participles.¹⁾

As far as I know, adjectival participles have never been dealt with in a way that is systematically based on grammatical categories connected with the predicate verb—mood, tense, aspect and voice. Because of the relevance of grammatical concepts such as progressive and passive and some relative-clause paraphrases in which different tenses appear, as in De Smet and Heyvaert (2011: 474) and Quirk et al. (1985: 1263–1265, 1326, 1328), adoption of tense, aspect and voice seems clearly reasonable here. However, as far as I know, mood has never been employed in explanations of adjectival participles. The reason for my adoption of this category is that the (especially indicative) mood of predicate verbs is related to the reality of what they describe, which is a key part of what is communicated by postmodifying participles, as stated above. Traditionally, three moods have been admitted in English grammar: indicative, subjunctive and imperative. I think that only the first is relevant to that existence or non-existence of the reality which is expressed by adjectival participles.²⁾ Consequently, in the paper, the term *mood* means indicative mood, and discussion of the presence or absence of mood refers to that mood only. If we take into consideration mood as well as tense, aspect and voice, I hope we will be able to deal with the grammatical meanings conveyed by adjectival participles better and more systematically than ever.

2.2. Hypotheses

I will put forward two hypotheses on pre- and postmodifying participles. The first is as follows:

Hypothesis I: The premodifying participle expresses the connotation of unreality and the grammatical meanings of aspect and voice, while the postmodifying participle expresses the connotation of reality and the grammatical meanings of aspect and voice. The connotation of reality here is taken to be conveyed by the 'grammatical meaning of the (indicative) mood' plus the 'grammatical meaning of tense', and the connotation of unreality by the non-presence of these.

First, let us analyse two examples of the postmodifying participle. (Grammatical meanings attributed to the participles are shown in square brackets [].)

- (1) Who is the man *wandering down the street*? (Quirk et al. 1985: 1326) [mood: *indicative*; tense: *present*; aspect: *progressive*; voice: *active*]
- (2) The car [...] *repaired by that mechanic* [...]. (Quirk et al. 1985: 1265) [mood: *indicative*; tense: *present/past*; aspect: *simple*³; voice: *passive*]
- (2') The car *that will be repaired/is* [...] *repaired/was* [...] *repaired by that mechanic* [...].

The indicative mood indicated by the postmodifying participles in (1) and (2) shows that the actions pointed to by them are actual ones. From the context, we can judge the tense of the participle in (1) to be the present tense,⁴ and the aspect to be the progressive aspect. As the paraphrases in (2') by Quirk et al. (1985) clearly show, the tense of the participle in (2) can be interpreted as present or past tense according to the context; the aspect can be said to be the simple, or non-progressive, aspect. Finally, present participles always carry active voice, as in (1), and past participles of transitive verbs passive voice, as in (2).

Next, we proceed to premodifying participles, which lack the grammatical meanings of (indicative) mood and tense. (Henceforth, the symbol 'φ' means there is no such grammatical meaning in the partici-

ple in question.)

- (3) Survivors were pulled from the *sinking* vessel. (De Smet and Heyvaert 2011: 483) [mood: \emptyset ; tense: \emptyset ; aspect: *progressive*; voice: *active*]
- (4) The *wanted* man was last seen in Cambridge. (Quirk et al. 1985: 1328) [mood: \emptyset ; tense: \emptyset ; aspect: *simple*; voice: *passive*]

Both premodifying participles here do not indicate mood or tense, that is to say, do not point to any actual action. Admittedly, we can often easily interpret the actions referred to by premodifying participles as actual ones; but such interpretations are no more than *inferential* ones based on the function of the anaphoric definite article *the* or on the context in which the premodifying participles occur. (We justly decide that it is an actual fact that the vessel was sinking in (3), but the interpretation is caused by the anaphoric definite article which refers the reader to the description prior to (3) of the actual event.) My view is that premodifying participles *in themselves* never express reality, that is, the grammatical meanings of indicative mood and tense.

Indeed, the argument contrary to my explanation that premodifying participles can be paraphrased into a relative clause (e.g. ‘an *interesting* book’ and its paraphrase ‘a book *which is interesting*’) and therefore can be said to show the grammatical meanings of (indicative) mood and tense seems plausible; but, on my hypothesis, the premodifying participle and the relative clause corresponding to it are different from each other in that the former lacks mood and tense and the latter, which includes a predicate verb, indicates them. Take a ‘*revolving* door’ for instance. Under my hypothesis, the premodifying present participle *revolving* lacks mood and tense, in other words, it does not express what really occurred, occurs, or will occur. Let us suppose a case of extreme kind: if a revolving door never revolved and were scrapped, we could nevertheless still call the door a revolving door. In this sense, we can conclude that premodifying participles in themselves do not denote any actual event or state—that is, they lack mood and tense. As mentioned above, if premodifying participles seem to refer to an actual

action in context, it is only due to inference. De Smet and Heyvaert (2011: 484) point out two readings allowed by the premodifying *exploding* in this sentence: While the *exploding* cigar that was intended to blow up in Castro's face is perhaps the best-known of the attempts of his life, others have been equally bizarre. The first is the 'inherent potential' reading ('the cigar is the kind of cigar that explodes'), and the other, the 'single specific bounded event' reading ('the cigar did explode'). My view is that the second reading is caused by inference based on the anaphoric definite article *the* (which allows the reader to identify the event in context) but not by the grammatical meaning conveyed by the premodifying participle; grammatically speaking, the premodifying *exploding* refers to the actual action on neither reading.

Hypothesis II: The premodifying participle presents an act or state it refers to as a 'characteristic' of the head noun. The postmodifying participle presents an act or state it refers to as an 'accident'⁵ of the head noun.

- (3) Survivors were pulled from the *sinking* vessel. [characteristic]
 (4) The *wanted* man was last seen in Cambridge. [characteristic]

Both premodifying participles—present *sinking* in (3) and past *wanted* in (4)—show characteristics of the head noun.

- (4') jobs *wanted* [accident]

Judging from the way in which Quirk et al. (1985: 1330) present (4'), the noun phrase *wanted jobs* would be unacceptable. The reason why the past participle *wanted* must be placed after the head noun in (4') is given by Hypothesis II: the state of 'being wanted (by people)' is not considered (unless in some special context) by English-speaking people to be a characteristic of, but merely an accidental state of, a *job*. Then, why is (4) acceptable? The reason is that, unlike the *wanted* in (4'), the *wanted* in (4) ('wanted by the police') is thought to be a characteristic of a particular *man*.

3. Mood, tense, aspect and voice as attributed to the adjectival participle

In this section, we will look at each of the grammatical meanings of adjectival participles using examples from contemporary British novels.¹⁾

3.1. Mood

Hypothesis I postulates that premodifying participles do not bear the grammatical meaning of mood at all. Further, as I stated in section 2.1, the only mood that is attributed to postmodifying participles is the indicative mood.

- (5) *Barking* dogs seldom bite. [proverb] [mood: \emptyset ; tense: \emptyset ; aspect: *simple*; voice: *active*]
 (6) The dog *barking next door* sounded like a terrier. (Quirk et al. 1985: 1263) [mood: *indicative*; tense: *past*; aspect: *progressive*; voice: *active*]

The premodifying *barking* in (5) does not indicate dogs' actual act of barking; instead, it expresses a characteristic intrinsic to the head noun: to bark often. By contrast, the postmodifying *barking* in (6) indicates a dog's actual act of barking, taken to be accidental to the head noun.

- (7) Two oil lamps [...] shed a calm creamy light upon the *scratched* [...] surface of what was once a fine rosewood table [...]. [SS: 16] [mood: \emptyset ; tense: \emptyset ; aspect: *perfect*; voice: *passive*]

Note that *scratched* here does not mean the table's past act of scratching itself or something else, but the state resulting from having been scratched.²⁾ That is to say, it lacks reality, that is, the meanings of mood and tense.

- (8) His reward for service *rendered* turned out to be the Czar's icon. [MH: 200] [mood: *indicative*; tense: *past*; aspect: *simple*; voice: *passive*]

Why does this simple past participle *rendered* postmodify the head noun even though the rule is that simple participles premodify the

head noun?³⁾ My answer is that, because the action indicated by the past participle *rendered* must be presented as an actual one, the writer has chosen to place the participle in the position after the head noun.⁴⁾ We can conclude that this *rendered* refers to an actual action of 'his' in the past ('service which was rendered by him'). In this light, the phrase *rendered service* is redundant: we take it for granted that service is rendered, in other words the noun *service* in itself includes the meaning of *rendered* connotatively.

3.2. Tense

As Hypothesis I postulates, premodifying participles do not bear the grammatical meaning of tense, while postmodifying participles represent present or past tense,⁵⁾ depending on the semantic context in which they occur. Let us look at some paraphrases in Quirk et al. (1985: 1263):

- (9) The person *writing reports* is my colleague.
 (9') The person *who will write|will be writing|writes/is writing|wrote|was writing reports* is my colleague.

Depending on the context, we can interpret the participial phrase *writing reports* in (9) as equivalent to one of the relative-clause paraphrases in (9'). From this pair of examples, we easily conclude that postmodifying participles represent tense.

- (10) Miss Kenton would prove the perfect solution to the problem *at present besetting us at Darlington Hall*. [RD: 49] [mood: *indicative*; tense: *present*; aspect: *progressive*; voice: *active*]

Here, the interpretation of the tense of the postmodifying participle is unambiguous: the adverbial *at present* unmistakably shows that the tense is the present.

- (11) [...] he said he thought it might be by one of the Huguenot silversmiths *working in London in the mid-eighteenth century* [...]. [LB: 315] [mood: *indicative*; tense: *past*; aspect: *simple*; voice: *active*]

In contrast to the adverbial *at present* in (10), the adverbial *in the mid-eighteenth century* in (11) demonstrates indubitably that the tense of the participle is the past tense.

3.3. Aspect

3.3.1. Simple aspect

In general, there have been two formal aspects—as realised in predicate verbs—admitted in the English language: the progressive and perfect aspects; however, I will count the simple form of a predicate verb itself as a form of the aspect and henceforth call it the *simple aspect*,⁶⁾ following the aspectual framework of Somiya (2010: 68–69).

- (12) I have heard of various instances of a butler being displayed as a kind of *performing* monkey at a house party. [RD: 35]
[mood: \emptyset ; tense: \emptyset ; aspect: *simple*; voice: *active*]

This *performing* should not be interpreted to imply the progressive aspect ('a monkey which *is performing*') but the simple aspect ('a monkey which *performs*') in this context.

- (13) She may, in the earlier times, have thought of escape; but gradually she fell, as so many *bullied* [...] women do, into a gradual despair. [SS: 158] [mood: \emptyset ; tense: \emptyset ; aspect: *simple*; voice: *passive*]

This *bullied* is interpreted in context as showing simple aspect.

- (14) I find myself now in the attic room of this small cottage *belonging to Mr and Mrs Taylor*. [RD: 159] [mood: *indicative*; tense: *present*; aspect: *simple*; voice: *active*]

To *belong* is a state verb, and this *belonging* thus cannot represent a progressive aspect but must indicate a simple aspect.

- (15) My account is curtailed, but omits nothing of substance and faithfully narrates the actual words *spoken*. [SS: 239] [mood: *indicative*; tense: *past*; aspect: *simple*; voice: *passive*]

Premodifying *spoken* does not refer to any actual act of speaking;

instead it characterises the head noun: *spoken words* are the kind of words that are spoken, not written. In contrast, postmodifying *spoken* in (15) denotes the particular, actual act of speaking in the past.

3.3.2. Progressive aspect

The progressive aspect expresses either the progression of an action

- (3) Survivors were pulled from the *sinking* vessel.
 (16) The policeman *patrolling* was now only a few paces from the consulate [...]. [MH: 162] [mood: *indicative*; tense: *past*; aspect: *progressive*; voice: *active*]

or its iteration.

- (17) [...] I began again to study the *jumping* waters. [SS: 247]
 [mood: \emptyset ; tense: \emptyset ; aspect: *progressive*; voice: *active*]

Postmodifying past participles can show the progressive only in periphrastic form, 'being + past participle' (cf. Quirk et al. 1985: 1265).

3.3.3. Perfect aspect

With premodifying past participles, the perfect aspect refers to the one that conveys a resultant state.⁷⁾

With postmodifying past participles, it conveys a 'resultant state' or contributes to the meaning of 'pluperfect'. Note that pluperfect, or past perfect, comprises past tense and perfect aspect,⁸⁾ and that within adjectival participles, the pluperfect is indicated only by postmodifying past participles, which can show both past tense and perfect aspect. (It follows that premodifying past participles cannot indicate pluperfect because they lack tense.)

- (18) [...] we were confronted by a *closed* gate [...]. [RD: 207] [mood: \emptyset ; tense: \emptyset ; aspect: *perfect*; voice: *passive*]

The perfect aspect of this *closed* conveys the resultant state of being closed; we should note that *closed* here does not refer to anyone's actual action but to a characteristic of the head noun.

- (19) Adam double-checked. It had to be some form of agreement *executed between the Russians and the Americans in 1867*. [MH: 248] [mood: *indicative*; tense: *past*; aspect: *perfect*; voice: *passive*]

The past participle *executed* here implies both past tense and perfect aspect—that is, the pluperfect. The time of the narrative is in the past (i.e., in 1966), while the time of the act referred to is much earlier (1867). This temporal gap is what the pluperfect expresses.

3.4. Voice

Present participles, whether they derive from intransitive or transitive verbs, represent the active voice.

- (20) He [...] looked across the roses at the *assembling* guests. [LB: 56] [mood: \emptyset ; tense: \emptyset ; aspect: *progressive*; voice: *active*]
 (21) [...] the costs of such a trip might still come to a *surprising* amount [...]. [RD: 10] [mood: \emptyset ; tense: \emptyset ; aspect: *simple*; voice: *active*]

Assembling in (20) is the present participle of the intransitive verb *assemble*; *surprising* in (21) is the present participle of the transitive verb *surprise* (and takes no object, as it is a ‘participial adjective’ rather than a present participle proper). Both present participles indicate active voice.

Past participles, if derived from intransitive verbs, indicate active voice, and if derived from transitive verbs, passive voice.

- (22) It had become almost dark, though there was still a little light over the sea where the *sunken* sun was still illuminating the line of white clouds [...]. [SS: 362] [mood: \emptyset ; tense: \emptyset ; aspect: *perfect*; voice: *active*]
 (13) She may, in the earlier times, have thought of escape; but gradually she fell, as so many *bullied* [...] women do, into a gradual despair. [SS: 158]

Sunken in (22) is the past participle of the intransitive verb *sink*; *bullied* in (13) is the past participle of the transitive verb *bully*. The former represents active voice, and the latter, passive voice.

4. Conclusion

What this paper has aimed to do is to present the grammatical meanings conveyed by the adjectival participle. The paper utilises grammatical categories inherent in the predicate verb—mood, tense, aspect and voice—and the concepts of *characteristic* and *accident* to explain differences in meaning between pre- and postmodifying participles. The table below shows the grammatical meanings conveyed by them.

	PREMODIFICATION		POSTMODIFICATION	
	PRESENT PARTICIPLE	PAST PARTICIPLE	PRESENT PARTICIPLE	PAST PARTICIPLE
MOOD	φ	φ	INDICATIVE	INDICATIVE
TENSE	φ	φ	PRESENT/ PAST	PRESENT/ PAST
ASPECT	SIMPLE/ PROGRESSIVE	SIMPLE/ PERFECT	SIMPLE/ PROGRESSIVE	SIMPLE/ PERFECT
VOICE	ACTIVE	ACTIVE/ PASSIVE	ACTIVE	ACTIVE ¹⁾ / PASSIVE

Needless to say, this paper certainly does not exhaust or comprehensively explain this topic. Nevertheless, I hope it throws some light on a challenging grammatical phenomenon. In my view, one grammatical matter that still needs dealing with in connexion with the adjectival participle is the adjectival to-infinitive (e.g. a house *to live in*). Work on differences between adjectival participles and adjectival to-infinitives can rely on the treatment of the former presented here.

NOTES

Title

1) The paper is an abridged version of my master's thesis, presented to the postgraduate school of the Tokyo University of Foreign Studies.

Section 1

1) In the paper I use the term *simple* to mean the use of participles without any arguments/adjuncts to them (e.g. a *sleeping* baby) except for the term *simple aspect*, and the term *compound* to refer to the use of participles with an argument/adjunct to them (e.g.

a baby *sleeping in bed*).

2) For example, *progressive* and *actuality* (or *reality*), even though they are different from each other (the former relates to aspect and the latter to mood and tense), are classified under the same category, *temporariness*.

Section 2

- 1) Cf. Jespersen (1914: 382) and Curme (1931: 64).
- 2) For one reason why adjectival participles do not, or cannot, indicate subjunctive mood, see Biber et al. (1999: 632).
- 3) On simple aspect, see section 3.3.1.
- 4) For the factors in determining tense in postmodifying participles, see Quirk et al. (1985: 1264) and Huddleston and Pullum (2002: 162).
- 5) In the paper I use the term *accident* in its philosophical meaning: 'a property of a thing which is not essential to its nature' (s.v. *accident* n. 3 in *Concise Oxford English Dictionary* (Twelfth edition)). For example, the premodifying past participle *rotten* in 'a rotten apple' indicates a 'characteristic' of an apple, whereas the postmodifying past participle *eaten* in 'The food *eaten* was meant for tomorrow' (Quirk et al. 1985: 1265) refers to an 'accident' of the food.

Section 3

- 1) To collect examples of the adjectival participle I have used four present-day British novels:

Jeffrey Archer. *A Matter of Honor (MH)*. New York: St. Martin's Press, 2005.

Alan Hollinghurst. *The Line of Beauty (LB)*. New York/London: Bloomsbury Publishing, 2004.

Kazuo Ishiguro. *The Remains of the Day (RD)*. London: Faber and Faber, 1989.

Iris Murdoch. *The Sea, The Sea (SS)*. London: Triad Grafton Books, 1980.

After each example quoted from one of these novels is shown the abbreviated title and the page on which the example appears. For example, [SS: 16] indicates page 16 of *The Sea, The Sea*.

- 2) Cf. Bolinger (1967: 9).
- 3) As the following examples show, there are cases where compound participles premodify the head noun: There is of course no refrigerator, which is dismaying to a *fish-eating* man [SS: 14]; Adam began to devour the *freshly cooked* food [MH: 247].
- 4) Another example of a simple postmodifying participle is 'being' in the idiomatic phrase *for the time being*. The phrase can refer to the present, the past, or the future (cf. Jespersen 1931: 91); therefore, the present participle *being* must be placed after the head noun, whose position allows interpretation of tense in my hypothesis.
- 5) In the paper I admit two tenses: the present and the past.
- 6) For example, in the sentences *She dances well* and *She danced well*, both of the predicate verbs *dances* and *danced* are taken to be the simple aspect.
- 7) Note that here, the 'resultant state' includes the so-called 'statal passive', which is appropriately explained by Quirk et al. (1985: 169–170). In the paper I treat adjectival past participles meaning 'resultant state', whether they correspond to the perfect form

of the verb or to the statal passive, as denoting perfect aspect.

8) With predicate verbs, pluperfect is expressed by the periphrastic form *had* + past participle, in which the first *had* itself denotes the past tense, and the combination of auxiliary *have* + past participle, the perfect aspect.

Section 4

1) Compound past participles derived from intransitive verbs normally cannot post-modify the head noun. For exceptions, see Quirk et al. (1985: 1265).

REFERENCES

- Archer, Jeffrey. *A Matter of Honor*. New York: St. Martin's Press, 2005. Print.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad, and Edward Finegan. *Longman Grammar of Spoken and Written English*. Harlow: Longman, 1999. Print.
- Bolinger, Dwight. "Adjectives in English: Attribution and Predication." *Lingua* 18 (1967): 1–34. Print.
- Curme, George Oliver. *A Grammar of the English Language III: Syntax*. Boston: Heath, 1931. Print.
- De Smet, Hendrik, and Liesbet Heyvaert. "The meaning of the English present participle." *English Language and Linguistics* 15.3 (2011): 473–498. Print.
- Hollinghurst, Alan. *The Line of Beauty*. New York/London: Bloomsbury Publishing, 2004. Print.
- Huddleston, Rodney, and Geoffrey K. Pullum. *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press, 2002. Print.
- Ishiguro, Kazuo. *The Remains of the Day*. London: Faber and Faber, 1989. Print.
- Jespersen, Otto. *A Modern English Grammar on Historical Principles: Part II*. London: George Allen & Unwin, 1914. Print.
- Jespersen, Otto. *A Modern English Grammar on Historical Principles: Part IV*. London: George Allen & Unwin, 1931. Print.
- Murdoch, Iris. *The Sea, The Sea*. London: Triad Grafton Books, 1980. Print.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. London: Longman, 1985. Print.
- Somiya, Kiyoko. "Eigo no asupekuto ni tsuite" ["On the Aspects in English"]. *Gogaku kenkyusho ronshu* [*Journal of the Institute of Language Research, the Tokyo University of Foreign Studies*] 15 (2010): 59–80. Print.

投稿規定

(1) 投稿は岩崎研究会会員に限る。但し、非会員であっても論文審査委員から推薦のあった場合は特別に認める。(2) 論文の内容は未発表のものに限る。(3) 用語は英語に限り、原則として native check を受けたものとする。(4) 注 (note) は後注とし、章ごとに通し番号を付ける。(5) ギリシャ字、ロシア字以外の特殊文字はできるだけローマ字化してほしい。音声記号は国際音声学協会 (IPA) 所定のものを用いる。(6) 引用文献: 書式は MLA Style に従う。(7) 枚数: 論文はワープロ原稿で、1行はアルファベットの小文字で70字、450行以内。A4判のハードコピー1部にCD-Rを添える。(8) 原稿はすべて論文審査委員による審査の上採否を決定する。共同執筆論文を別として、論文の掲載は毎号1人1篇とする。(9) 都合により短縮を求められることがある。印刷上の体裁および論文の掲載年度については編集委員に一任する。(10) 抜刷は20部までを無料で、別に本誌1部を呈上する。(11) 原稿は随時受付ける。(12) なお、詳細は別に定める。

論文審査委員

赤須 薫	土肥 一夫
岩崎 春雄	馬場 彰
浦田 和幸	東 信行
斎藤 弘子	増田 秀夫
清水あつ子	宮井 捷二
投野由紀夫	山田 茂

編集後記 本誌の発行が諸事情により遅れていましたが、ようやく会員の皆様に届けられることになりました。これは、論文の執筆者は言うまでもなく、編集に携わった方々、特に赤須先生の努力の結果であることを記したいと思います。

(2014年9月 S.M.)