Unit 19 L2 acquisition of grammatical morphemes

(Case study 3)

19.1 Introduction

This unit returns to the issue of language learning as discussed in units unit 10.8 and 16 via an interlanguage study on the basis of a learner corpus. We will also explore error tagging and problem-oriented annotation introduced in units 4.4.8 and 4.4.9.
In this case study, we will use a corpus of learner English to investigate the data produced by L2 learners of English. The description of learner language is of particular interest to second language acquisition (SLA) research. The interest in learner language stems from the assumption that the researcher may gain insights into the process of second language acquisition by exploring the L2 productions of a language learner. If we have a better understanding of the second language acquisition process we can apply the findings to a variety of practical aspects of language teaching: syllabus design, materials development, task design, language testing, and so on.
A number of different approaches have been taken to the description of learner language. Ellis (1994: 44) identified four major approaches:

- the study of learners' errors
- the study of developmental patterns
- the study of variability
- the study of pragmatic features

The study of learners' errors was undertaken quite intensively in the late 1960s and 1970s after Pit Corder (1967) made the significant claim that L2 learners, like L1 learners, were credited with a 'built-in-syllabus', which guided their language acquisition. Selinker (1969) coined the term *interlanguage* to refer to the special mental grammars which it was assumed that learners constructed during the course of their language acquisition. Interlanguage theory treated learner behaviour, including their errors, as rule-governed.
While initially associated strongly with error analysis, interlanguage analysis based upon error analysis went out of fashion in the 1980s as a number of methodological and theoretical problems with it were identified. Ellis (1994: 73), for example, pointed out that error analysis did not provide a complete picture of how learners acquire an L2 because it described learner language as a collection of errors. As a consequence

of criticisms such as this, more and more attention was paid to the entirety of learner language. Central to this enterprise is the description of developmental patterns of interlanguage.

Dulay and Burt (1973) were among the first to conduct an empirical study of the acquisition order of the grammatical features of English. They studied the order of acquisition of grammatical morphemes (such as *-ing* and *the* that play a greater part in structure than content words such as *dog*), which was first investigated by Roger Brown in L1 acquisition (Brown 1973). Throughout their papers, Dulay and Burt claimed that L2 acquisition proceeds quite systematically and that the acquisition order is not rigidly invariant but is remarkably similar irrespective of the learners' L1 backgrounds, age and/or medium of production. Since then, more than fifty L2 morpheme studies have been reported, using data from a variety of L1 backgrounds and analysis procedures (see Larsen-Freeman and Long 1991; Ellis 1994 for a review). Criticisms of the methodology utilized in the early morpheme studies are well known (see Long and Sato 1984). However, as Larsen-Freeman and Long (1991) noted, despite admitted limitations in some areas, the morpheme studies provide strong evidence that interlanguages exhibit common accuracy/acquisition orders. Contrary to what some critics have claimed, so many studies have been undertaken with sufficient methodological rigour which show sufficiently consistent general findings that the commonalities can no longer be ignored (*ibid*: 92). The aim of this study is to verify some of their findings by using the corpus-based approach.

Recently there has been a growing awareness that it is necessary to investigate learner language by collecting a large amount of learner performance data on computer, so-called learner corpora (see unit 7.8 and 10.8). The term learner corpus was first used for Longman's learner dictionaries, in which the information on EFL learners' common mistakes was provided based upon the Longman Learners' Corpus. Following from this, a project called the International Corpus of Learner English (ICLE) was launched as a part of the ICE (International Corpus of English) project (Granger 1998) in 1990 specifically to collect L2 data (see unit 7.8). From these beginnings the interest in learner corpora has grown, and at the time of writing more than a dozen projects constructing learner corpora are under way around the world (see unit 7.8).

In this case study, we will revisit a once popular topic of SLA research, acquisition studies of English grammatical morphemes and see how learner corpora can shed new light on this old area of study. There are a couple of reasons why we chose morpheme studies as our primary topic for this investigation. Firstly, as Ellis (1990) noted, morpheme acquisition studies were a kind of performance analysis in the sense that they aimed to provide a description of the L2 learner's language development and

looked not just at deviant but also at well-formed utterances (Ellis 1990: 46). Performance analysis provides a basis for investigating the following important questions:

- Is there any difference between the order of instruction and the order of acquisition?
- Is it possible to alter the 'natural' order of acquisition by means of instruction?
- Do instructed learners follow the same order of acquisition as untutored learners or a different order? (Ellis 1990: 139)

Learner corpora, if used properly within a suitable research design, can prove to be an effective tool which can be used to answer these interrelated questions by providing the evidence of learner language in a more systematic and comprehensive way. Secondly, although there are many criticisms of morpheme studies (e.g. Hatch 1978; Long and Sato 1984), morpheme acquisition order studies are still a good starting point if one wishes to see how effective learner corpora can be in describing interlanguage.

19.2 Morpheme studies: a short review

In the early 1970s, it was discovered that English speaking children learn grammatical morphemes in a definite sequence (Brown 1973). Dulay and Burt (1973) decided to replicate the study with L2 learners. They asked Spanish-speaking children learning English to describe pictures, and checked how often the children used eight grammatical morphemes in the right places in a sentence. The results showed that L2 learners have a common order of difficulty for grammatical morphemes, as shown in Table. 19.1:

Table. 19.1 An accuracy order of grammatical morphemes (Dulay and Burt 1973)

Order	Morpheme	Example
1	plural *-s*	book*s*
2	progressive *-ing*	John is go*ing*
3	copula *BE*	John *is* here
4	auxiliary *BE*	John *is* going
5	articles	*the* books
6	irregular past tense	John *went*
7	third person *-s*	John likes books
8	possessive *-s*	John'*s* book

One of the problems for the rank orders that Dulay and Burt observed is that they disguise the difference in accuracy in use between various morphemes. For instance, a morpheme with a one percent lower accuracy of usage than another morpheme is

given a different ranking in just the same way as a morpheme that is used 25% less accurately would be. To overcome this problem, Krashen (1977) proposed a grouping of morphemes (see Fig. 19.1). He claimed that it was 'a natural order supported by the longitudinal and cross-sectional, individual and grouped SL findings. Items in the boxes higher in the order were regularly found (80-90%) accurately supplied in obligatory contexts before those in boxes lower in the order' (Krashen 1977: 151).

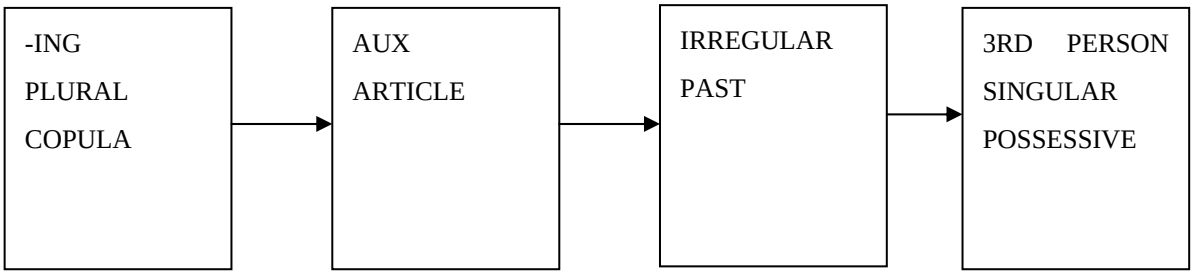

Fig. 19.1 The natural order for L2 acquisition proposed by Krashen (1977)

The results were used to claim that there was a more or less invariant order of acquisition which was independent of L1 background and age. Although this order was slightly different from that found for the same morphemes in L1 acquisition research, it provided evidence in favour of the existence of universal cognitive mechanisms which enabled learners to discover the structure of a particular language (see Ellis 1994 for a detailed review).

Although there were some stern critics of the morpheme studies as noted above, the interest in morpheme acquisition grew to the extent that different approaches to the study of morpheme acquisition emerged: target-like use analysis of morphemes, as opposed to obligatory context analysis only (Pica 1982; Lightbown 1983), morpheme acquisition in different L2 contexts (Fathman 1978; Makino 1980; Sajavaara 1981), and by learners with different L1 (Mace-Matluck 1977; Fuller 1978) and L2 backgrounds (Bye 1980; van Naerssen 1986). The different types of studies clearly made different contributions to the study of grammatical morpheme acquisition, though they were all focused on the same general question. Larsen-Freeman and Long (1991: 92), based on their review of this literature, concluded that those studies provided strong evidence of a developmental order for L2 grammatical morpheme acquisition.

Given that there is some merit in exploring the acquisition of such morphemes by learners, an obvious place to start such an investigation is in a learner corpus. That is what we will do now in this case study.

19.3 The Longman Learners' Corpus

The learner corpus data used for this study is a subcorpus from the Longman Learners' Corpus (i.e. see unit 7.8). We selected this corpus for our study because it is one of the few learner corpora which is publicly available for research and each of its components per L1 background is large enough to allow one to extract errors and provide data from learners at different proficiency levels.

In this study, we will use the subcorpus that is composed of the Japanese EFL learners' written composition data in order to see whether the acquisition order of grammatical morphemes is the same as the one found by previous studies. While this study examines the morpheme acquisition in Japanese learner data, the methodology and techniques described in the following sections should apply to other learner groups as well. Before we can undertake the interlanguage study, we will first need to annotate the corpus data for parts-of-speech and error types.

19.4 Problem-oriented corpus annotation

Since we need to find all of the correct and incorrect instances of relevant morphemes in the learner data, this case study will involve a large amount of problem-oriented manual tagging work (see 4.4.9). The basic procedure of data processing is as follows. First, the original corpus data will be pre-processed so that we can convert the header information into a format suitable for use with WordSmith (version 4). Second, the whole subcorpus will be tagged using a POS tagger. Third, we will manually tag morphological errors using the POS information as a guide. Then the accuracy rate of the usage of individual grammatical morphemes will be obtained based on the number of the error tags and relevant POS tags. Finally, the results will be compared with previous findings.

19.4.1 Basic formatting of the texts

Since some of the corpus data was compiled in the 1980s, the header information was formatted in the way similar to the COCOA format (see unit 3.3). In order to process this data properly for use with WordSmith, we need to convert the header section into the SGML format (see unit 3.3), i.e. with the opening tag <head> and the ending tag </head>. This is necessary because by default WordSmith does not distinguish the header apart from the main text.

In this example, we will select the Japanese learner group (in the file folder 'J' in the Longman Learners' Corpus). The subcorpus contains approximately one million words in total. Table 19.2 shows the typical format of a corpus header, with glosses added by us to explain these header elements.

Table 19.2 The corpus header

Element	Example value	Gloss
<RF>	20522	Document reference number
<CO>	JAP	Source country code
<LA>	JAP	Student language & national variety code
<LE>	IN	Student level code (IN =intermediate)
<EN>	CLA	Environment code (CLA=classwork)
<TT>	2	Task type (2=free essay)
<TV>	AmE	Target language variety
<PA>	My friend and I went to…	The beginning of the essay

We will first show you how to convert the COCOA format into the SGML format. As each corpus file starts with the <RF> (i.e. reference number) tag, we can insert the <head> tag before <RF> and insert the </head> tag between the <TV> (target language variety) element and the first occurrence of <PA> (paragraph) tag. There are several different ways to do this. The simplest, yet most time-consuming way is to use the 'find and replace' function of a text editor. If you are using Microsoft Word, go to *Edit – Replace* after opening a corpus file. Enter <RF> in the 'Find what' text box and <head>^p<RF> in the 'Replace with' text box (here ^p means a new line). Click on the 'Replace' button to insert the start tag <head>. Similarly, replace **the first occurrence** of <PA> with </head>^p<PA>. After these operations, save the file as a text file (i.e. 'Text only' rather than 'Word document'). The result should look like the example below:

<head>
<RF> 20522
<CO> JAP
<LA> JAP
<LE> IN
<EN> CLA
<TT> 2
<TV> AmE
</head>
<PA> My friend and I went to …

After creating the corpus header for all of the files, we will combine these files into eight large files, each for a proficiency level, as shown in Table 19.3. Note that in some subcorpora of the Longman corpus, there are a number of files in which the proficiency level is unknown or incorrectly coded. In these cases the proficiency level is labelled as 'XX' or 'ZZ'. If your study of a certain subcorpus contains such files, you can discard them. Since we have to add the corpus header to all of the files in the

Japanese subcorpus (1,667 files), we need to find a more time and labour-saving way to do the task. One possible way is to write a so-called 'Word macro' for this purpose if you are good at programming (see the online help of Microsoft Word for a description of what a macro is or how to record a Word macro); alternatively you can download a tool such as PowerGREP from the Internet to help you. However, combining these small files into eight larger files according to their different proficiency levels presents yet another difficulty for most readers. As such, we have written a simple program that will help readers to add corpus headers and combine files at one go. The program also inserts the start tag <body> and the end tag </body> to indicate the corpus data proper. After you have downloaded the program from our companion website, simply place it in the file folder for the subcorpus you have selected (in the case of Japanese data, the file folder is J of the Longman Learners' Corpus) and double click the program file. You will have eight files named in Table 19.3, each containing marked up corpus data for the relevant proficiency level (files of which the proficiency level is labelled as 'XX' or 'ZZ' are discarded automatically by the program). This tool can not only help you to pre-process the Japanese data used in this case study, it is also useful when you undertake your own mini-projects using the Longman Learners' Corpus.

Table 19.3 Proficiency levels encoded in the Longman Learners' Corpus

File name	Proficiency level	New group
llc_BE.txt	Beginning	Elementary
llc_EL.txt	Elementary	
llc_PI.txt	Pre-intermediate	Intermediate
llc_IN.txt	Intermediate	
llc_UI.txt	Upper-intermediate	Advanced
llc_AD.txt	Advanced	
llc_PR.txt	Proficient	Proficient
llc_AS.txt	Academic studies	
	Degree-course	

With the SGML corpus header, tools such as WordSmith can be used to search the main body of text while excluding the data in the header. To facilitate our analysis of morpheme errors, we will put the eight files into four groups as indicated in Table 19.3 ('New group'): elementary, intermediate, advanced and proficient.

19.4.2 Error tagging: a proper way and a dirty way

In this section, we will show you two different methods of extracting the error information on grammatical morphemes. One is to annotate the entire corpus and then extract error tag frequencies. The other way is to retrieve concordance lines using WordSmith and mark the lines with errors temporarily for counting. While both

methods can yield the accuracy rate for each morpheme, each has its own advantages and disadvantages. The former will be more useful in terms of replicability and reusability, but it is quite time-consuming. The latter, on the other hand, is time-efficient, but the results are only valid as long as the processed data is stored on your computer – there is no way for the results to be used for secondary purposes. Both methods involve POS tagging. In this case study, we tagged the data with CLAWS (see unit 4.4.1 and the Appendix). Note that the POS tagged files were renamed as *llc_be.pos.hrz.txt*, etc.

Table 19.4 shows a tagset of grammatical morphemes with which we will annotate our data. We have chosen to study these grammatical morphemes because they were studied in Brown (1973), Dulay and Burt (1973) and Krashen (1977). By studying the same set of grammatical morphemes, we will be able see if their claims still hold in the light of corpus data.

Table 19.4 List of morpheme tags

Morpheme	Correct tag	Error tag
article	<ART>	<ER_ART>
possessive -*s*	<POS>	<ER_POS>
3rd person singular -*s*	<3PS>	<ER_3PS>
irregular past	<IRPST>	<ER_IRPST>
auxiliary *BE*	<AUXBE>	<ER_AUXBE>
plural -*s*	<PL>	<ER_PL>
copula *BE*	<COP>	<ER_COP>
progressive -*ing*	<PROG>	<ER_PROG>

To ensure the comparability between Dulay and Burt's study and ours, we will tag each text according to the criterion set for the Bilingual Syntax Measure proposed by Dulay and Burt. In other words, we will only look at the 'obligatory context', i.e. contexts that require the obligatory use of grammatical morphemes in samples of learner language. We will convert the relevant tags on the basis of the following rules:

(1) Tagging verb *BE* as <COP> or <AUXBE>:

a) Look for every occurrence of *BE* verb (*_VB*)

b) If those *BE* verbs are followed by either verbs with a progressive marker (*_VVG) or a past participle marker (*_VVN), then assign the tag <AUXBE>;

c) If those *BE* verbs are followed by adjectives (*_J*) or nouns (*_N*), then assign the tag <COP>.

(2) Assigning the tag <PL> to all the nouns with the tag NN2 (i.e. *_NN2);

(3) Assign the tag <POS> to all the items tagged as GE (i.e. *_GE);

(4) Assign the tag <PROG> to all the words tagged as VVG (i.e. *_VVG);

(5) Assign the tag <3PS> to the words tagged VVZ (*_VVZ)

(6) Assign the tag <ART> to *the*_AT, *a*_AT1 and *an*_AT1 (*_AT*)

(7) <IRPST> should be tagged by looking at each verb labelled VVD. Both regular and irregular verbs were labelled VVD, so there is no way of distinguishing them other than a manual analysis.

The task of searching for and replacing relevant POS tags with the morpheme tags can be done semi-automatically (e.g. using WordPad) in cases (2) – (6) with the help of POS tags. For (1) and (7), you need to look at each case separately and manually correct the tags. For this, you can use a text editor with a search function (e.g. Notepad, Wordpad or Microsoft Word). We do not have space to illustrate this procedure step by step here. If you use MS Word, remember again to save the edited file as a text file.
After these tags are automatically assigned, errors can be manually tagged for each example retrieved using the above morpheme tags. Here we should remember at least three types of error notations:

1) *overuse error*: This is <ER_ART CF="">a</ER_ART> my dog.
2) *omission error*: This is <ER_ART CF="a"></ER_ART> dog.
3) *misformation error*: This is <ER_ART CF="a">an</ER_ART> dog.

Note that by specifying the value of the attribute CF ('correct form'), the morpheme error tags can specify types of target modification errors (overuse, omission or misformation) without creating extra tags. If one wants to extract overuse errors only, one can do so by extracting the lines that contain the string *CF=""*.
After tagging is done, we can calculate the accuracy rate with which the morphemes were actually used in our corpus data. We will again follow the measurement method adopted by Dulay and Burt, namely looking only at the obligatory context, so that the information on overuse of the forms will not be used even though the annotated corpus itself contains such information.
Readers may have already noted that the conventional 'proper way' of error tagging is extremely time and labour-consuming. However, the time and labour which you have invested will be repaid because the annotated data is reusable. For those who do not wish to invest a large amount of time in error-tagging their learner data, we will introduce a dirty yet quick way to complete the task, which takes advantage of the SET function of WordSmith version 4. As an example, we will give you a step-by-step demonstration of how to tag errors related to articles (*the, a* and *an*).

1. Open WordSmith 4 and choose texts (Fig. 19.2).

2. In order to avoid searching the header section, we will need to limit our search to just the body of the text. To do this, choose *Settings – Adjust settings*. The *Tools Settings* window will open (Fig. 19. 3). Choose *Tags & Markup* and click on the *Only Part of File* button at the bottom. Then type in <*head*> and </*head*> in the blanks of *Sections to Cut Out*. Then press *OK* (Fig. 19. 4).

3. Also click on the *Text & Languages* tab, add the underscore (_) in the *Characters within word* field. This will ensure that the word plus POS information *the_AT*, for instance, is recognized as one word. If you do not add this information, the Concord function will not provide any information on collocations or clusters. So make sure you properly set this option (Fig. 19. 5).

4. Go back to the main window and choose *Concord* (Fig. 19. 6). Select *File – New*. The search window will then open. In order to find article errors, let us search for all the instances of singular nouns first. Type in **_NN1* in the search box and then press *OK* (Fig. 19.7).

5. You will get the search results (Fig. 19.8).

6. In order to check the missing articles, sort the concordance lines by the left context. Choose *Edit – Resort*. Then select L1 (the first word on the left of the node word) for the main sort, L2 and L3 for the second and the third sorts. Make sure you activate the sort by ticking the box on each tab menu (Fig. 19. 9).

7. The sorted concordance lines will appear (Fig. 19. 10).

8. Here you should check each concordance line to see if there is an article error. Move the cursor to the column *SET*. When you find an error, type 'E' in the same line on the *SET* column. By doing so, you can mark the concordance lines for errors (Fig. 19.11).

You will have to make a realistic decision here. Since there are more than 160,000 concordance lines, it would be nearly as time-consuming as the 'proper way' to check every line. As such we would suggest that you error-tag concordances at random, for instance, every 10th line (make the adjustment in *WordSmith Settings - Concord*), if you wish to reduplicate this study as a warming up exercise for your mini-project at the end of this case study. This random sampling will allow you to tag a few thousand examples in a reasonable time.

9. After the manual error tagging, you can re-sort the concordance lines by the *SET* order (Fig. 19.12).

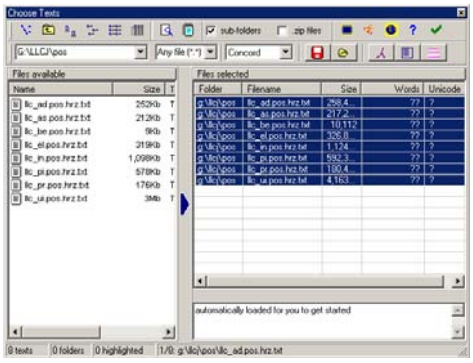

Fig. 19.2 Choosing files

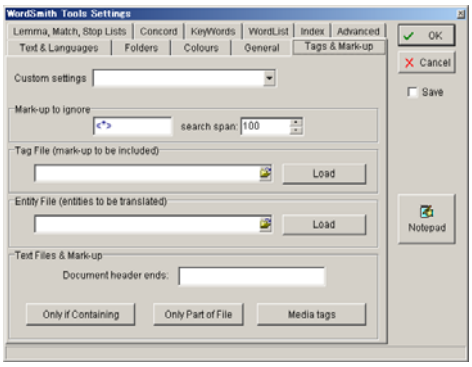

Fig. 19.3 Adjusting settings

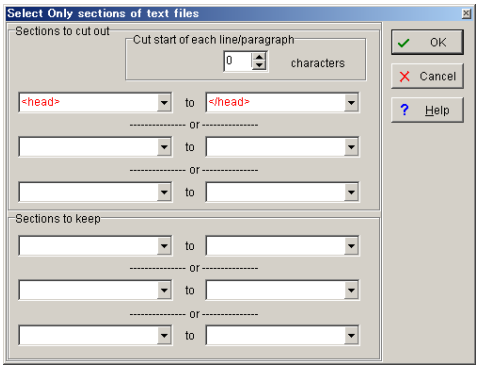

Fig. 19. 4 Choosing only part of file

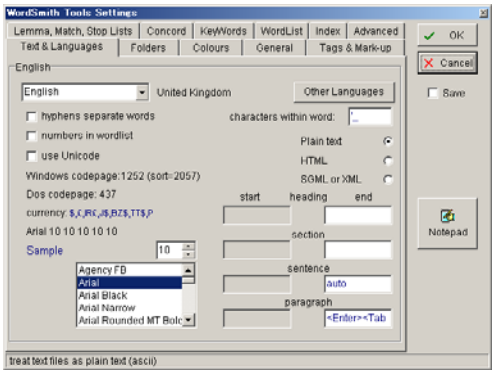

Fig. 19. 5 Defining a word

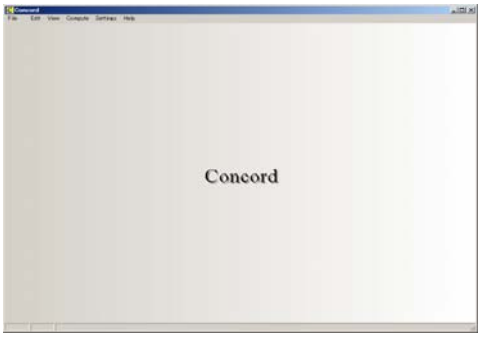

Fig. 19. 6 Concord

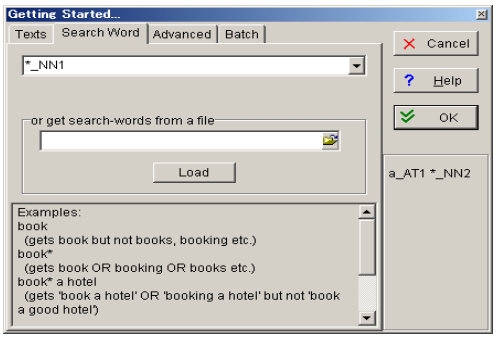

Fig. 19. 7 The search window

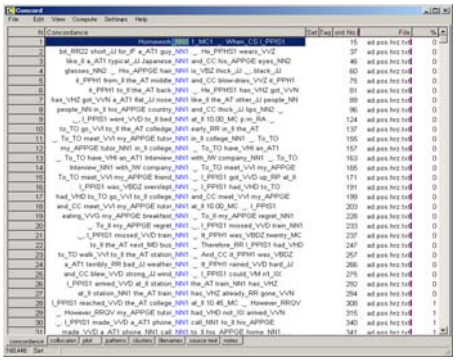

Fig 19.8 The search results

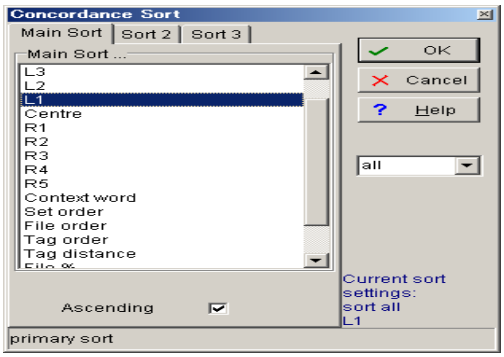

Fig. 19.9 The sort window

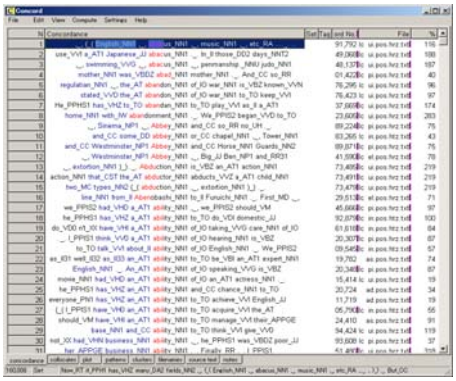

Fig. 19.10 Sorted by the left context

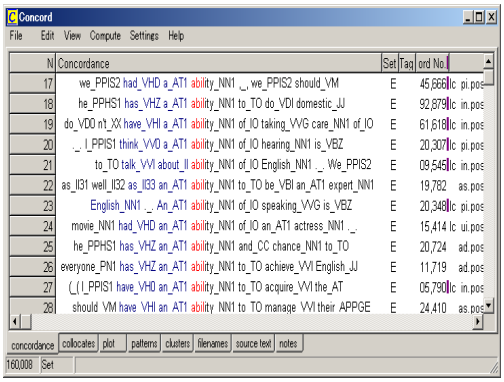

Fig. 19.11 Marking errors

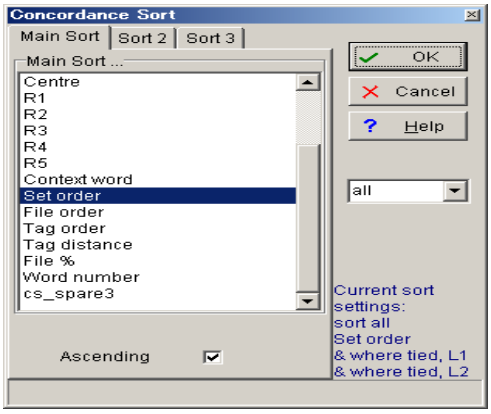

Fig. 19.12 Sort by the *SET* order

This will give you the number of article errors for singular nouns. You can do the same for plural nouns. Make sure that you will get both the number of correct cases and the number of incorrect cases.

Using the random sampling procedures, we have error-tagged 2,000 concordances for articles and each of the other morphemes under consideration. Table 19.5 shows what to search for each grammatical morpheme and Table 19.6 gives the overall frequencies of correct and incorrect forms of each morpheme.

Table 19.5 POS tags for each morpheme

Morpheme	POS tags to search	Comments
copula *BE*	VB*	Right-sort, then delete all the cases in which the VVG and VVN tags follow *BE* verbs.
auxiliary *BE*	VB*	Right-sort, then retain the lines which have VVG and VVN tags after *BE* verbs
plural *-s*	NN*	NN1 for omission errors
progressive *-ing*	VVG; VB* + VV*	
3rd person singular *-s*	VVZ	
irregular past	VVD	

Table 19.6 Overall frequencies of morpheme errors

Grammatical morpheme	Data type	Elementary		Intermediate		Advanced		Proficient	
		Correct	Error	Correct	Error	Correct	Error	Correct	Error
Copula *BE*	Fre.	1885	115	1929	71	1908	92	1895	105
	%	94.25	5.75	96.45	3.55	95.40	4.60	94.75	5.25
Plural *-s*	Fre.	1598	402	1622	378	1630	370	1773	227
	%	79.90	20.10	81.10	18.90	81.50	18.5	88.65	11.35
3rd person *-s*	Fre.	1417	583	1394	606	1538	462	1760	240
	%	70.85	29.15	69.70	30.30	76.90	23.10	88.00	22.00
Possessive *-s*	Fre.	1538	462	1515	485	1879	121	1882	118
	%	76.90	23.10	75.75	24.25	93.95	6.05	94.10	5.90
Article	Fre.	1259	741	1404	596	1204	796	1590	410
	%	62.95	37.05	70.20	29.80	60.20	39.80	79.50	20.50
Irregular past	Fre.	1643	357	1590	410	1581	419	1671	329
	%	82.15	17.85	79.50	20.50	79.05	20.95	83.55	16.45
Auxiliary *BE*	Fre.	1773	227	1920	80	1737	263	1843	157
	%	88.65	11.35	96.00	4.00	86.85	13.15	92.15	7.85
Progressive *-ing*	Fre.	1444	556	1654	346	1796	204	1862	138
	%	72.20	27.80	82.70	17.30	89.80	10.20	93.10	6.90

19.5 Discussion

The accuracy rate can be easily computed by dividing the frequencies of correct forms by the total of frequencies of correct and incorrect forms shown in Table 19.6. The state of acquisition is defined as ‘90% correct’ in the same way as Dulay and Burt’s Bilingual Syntax Measure.

Table 19.7 shows the results of the accuracy order of the eight grammatical morphemes, where the figures are expressed in proportions. According to this result, the Japanese-speaking learners of English have least difficulty with copula *BE*, which is basically the same as the order proposed in Dulay and Burt’s (1973) and Krashen’s (1977) studies. The most difficult item for them is the article (*the*, *a* and *an*). This late acquisition of the article system is similar to the findings with Polish learners reported by Botley and Uzar (1998). As neither Japanese nor Polish has an article system, it can be reasonably speculated that L1 knowledge can affect the acquisition process in both cases.

Table 19.7 Accuracy order of eight grammatical morphemes

Grammatical morphemes	Elementary	Intermediate	Advanced	Proficient
Copula *BE*	94.2	96.3	95.5	94.7
Auxiliary *BE*	89	96.1	86.7	92.5
Possessive *-s*	76.7	76.2	94.8	95.2
Progressive *-ing*	72.1	82.3	89.8	94.3
Plural *-s*	80	81	81.4	88.5
3rd person *-s*	70.8	69.6	76.7	89.4
Irregular past	82.3	79.6	78.9	83.7
Article	63	70.2	60.2	79.6

Among the eight morphemes, copula *BE*, auxiliary *BE*, possessive *-s* and progressive *-ing* reached a 90% accuracy rate and can be regarded as 'acquired' items. However, the other four morphemes did not reach an accuracy rate high enough to be considered acquired, even in the proficient student group.

Fig. 19.13 shows diagrammatically the comparison of our results with the order observed by Dulay and Burt (1973) and Krashen (1977). The noteworthy difference is that articles, which show the lowest accuracy rate in all of the morphemes, are the most difficult items for Japanese learners. Since the Japanese language does not have the notion of articles attached to nouns, the proper use of articles should be very difficult for them to acquire. Possessive *-s*, in contrast, is the item which is relatively easier for Japanese learners, as reflected by its higher rank order in relation to the order observed by Dulay and Burt, and Krashen. This is perhaps because Japanese has a particle *no*, which is similar to the English possessive *-s*, as in *John no hon* 'John's book'.

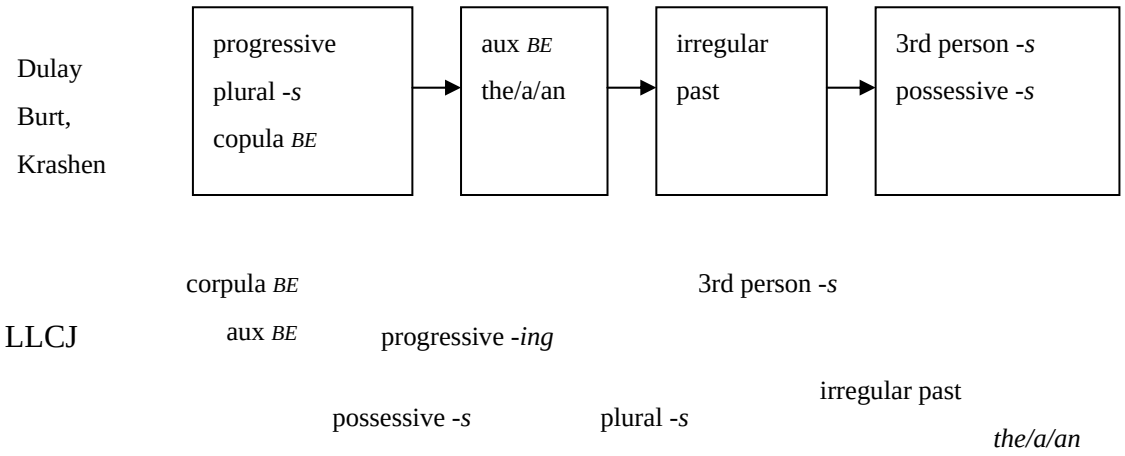

Fig. 19.13 Comparison of morpheme acquisition orders

The results indicate that there are some differences between the order of acquisition of grammatical morphemes proposed previously and the order found on the basis of learner corpus data in this study. What is the implication of this finding? One possible implication is that one may argue that the concept of a universal order of acquisition needs to be reconsidered. A large-scale corpus-based analysis will shed more light on this issue. It would be interesting if one could compare the results for the Japanese learners with those of other L1 groups represented in the Longman Learners' Corpus (French, German, Spanish, and so on). The other possible implication is that one may argue that the universal acquisition order should hold and the fact that it does not may lead us to examine carefully the quality of our learner corpus data. Since the Longman learner data was collected in an opportunistic way, the criteria for determining the subjects' proficiency level are sometimes subjective. Learners' proficiency levels as

graded in the Longman Learners' Corpus do not always match the quality of actual compositions. If the corpus design had been better, the results might have been different. This is merely a possibility, but it is a timely reminder that the results derived from a corpus are, to some extent, only as good as the corpus itself (cf. unit 10.15). A badly designed corpus, or indeed, even a well designed corpus, when used for a purpose it is not designed for, may provide misleading results.

19.6 Unit summary and suggestions for further study

This study exploited a commercial learner corpus called the Longman Learners' Corpus by examining how Japanese learners acquire grammatical morphemes such as the ones investigated by Dulay and Burt (1973) and others in the 1970s. These influential studies claimed that L2 learners acquired grammatical morphemes in a universal order independent of the developmental path of learning. This study largely verified that claim using large learner corpus data. It also introduced problem-oriented manual error tagging techniques and the newly released WordSmith 4.

As the Longman Learners' Corpus contains data from learners of English from various L1 backgrounds and at different proficiency levels, it has made it possible to investigate the features of interlanguage across different L1 backgrounds and developmental stages. Readers are advised to use the techniques introduced in this case study to analyze the order of acquisition of English grammatical morphemes on the basis of the components produced by learners with their L1 backgrounds.