RETHINKING TRANSLATION UNIT SIZE:
AN EMPIRICAL STUDY OF AN ENGLISH-JAPANESE NEWSWIRE CORPUS

BY

FUMIKO KONDO

A THESIS SUBMITTED TO
THE UNIVERSITY OF BIRMINGHAM
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF ENGLISH
COLLEGE OF ARTS AND LAW
THE UNIVERSITY OF BIRMINGHAM
NOVEMBER 2009

UNIVERSITY OF BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

ABSTRACT

The translation unit has been regarded as an elusive notion in linguistics. The literature shows that there seems to be little agreement regarding, in particular, their identification and size. This study attempts to rethink these two central issues of translation units with the help of a parallel corpus: the ARC (the Alignment of Reuters Corpora), an English-Japanese newswire corpus. The main achievements of this study are: the identification of five variables associated with translation unit size; the establishment of an unbiased, reproducible identification method; and, the demonstration that translation pairs (i.e. translation units and their equivalents) are ideal for contrastive analysis. The identification method, 'the one-equivalent principle', established in this thesis is justified linguistically by a thorough, systematic review of the relevant literature, and empirically using nine case studies. The target words of the case studies were the most frequent content words in the ARC: *market*; *government*; *year*; *economic*; *new*; *foreign*; *said*; *told*; and, *expected*. The examination of translation unit size, as well as non-translation units and translation pairs, shows that parallel corpora, and the one-equivalent principle, are powerful tools for understanding the nature of translation units.

For Shoji, Akemi, and Oliver

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LISTS OF EXAMPLES

LISTS OF FIGURES

LISTS OF TABLES

1. Introduction

The translation unit is a concept that linguists created; its purpose is to describe, systematically, how translators carry out their tasks. This term has been useful and convenient in translation pedagogy, as well as in machine translation. For example, in a translation lesson, a lecturer might claim that: 'Word-for-word translation does not always produce a good translation. A good translation is one that is carried out by translating translation units'. Or, in an introductory module on machine translation, students might be taught: 'Computers can recognise words separated by spaces in text and replace the words into the corresponding words in another language; however, such a translation process is unlikely to produce a readable result. On the other hand, if computers are able to detect translation units in a text, and replace them into the corresponding items in another language; this translation should be as readable as one created by human translators'. Such introductory lessons are easy to understand and the logic has a neat clarity.

However, if we then suppose that a student goes on to ask further questions, such as, 'How can we find such translation units then?', or 'What do these units look like?'; the lecturer is obliged to concede that, 'That is a long story!'. Although translation phenomena become easier to explain in an abstract way by using this term, there is a risk that the translation unit alone ends up bearing, and possibly subsuming, many of the complex mysteries of translating. The translation unit is an 'elusive' notion, which 'is not easy to describe, categorise, and formalise' (Lefevere, 1993). Indeed, within linguistics, there seems to be little agreement on the identification and size of translation

units; the literature shows that the methods of identification are diverse, and the outcomes (i.e. the identified translation units) are very variable. The elusiveness of the translation unit concept may be due to the use of methods that approach the discourse via language data subjectively sampled, and therefore of limited usefulness for drawing inferences about processes, such as translation unit formation, which may characterise the wider discourse community in which one is interested.

To address this situation, this thesis attempts to investigate what a translation unit is and what it looks like with the help of a corpus. This thesis, therefore, is situated in corpus linguistics. Among the several different definitions of corpus linguistics, I align myself with Teubert (2005): 'Corpus linguistics is … an insistence on working only with real language data taken from the discourse in a principled way and compiled into a corpus'. The literature on translation unit size often uses language data; however, most of these studies do not belong to corpus linguistics *sensu* Teubert (2005). For example, the language data seen in Barkhudarov's (1993) landmark study on translation units, discussed in 2.3, are not 'real', but introspective (i.e. 'language data constructed by linguists', Olohan, 2004). There is no indication of where the examples in Barkhudarov's study were taken from. Analysing examples outside of discourse means his work should not be considered part of corpus linguistics, as defined here.

Similarly, the translation unit studies by Alves and Gonçalves (2003) and Lörscher (1991), discussed in 2.2, would not be categorised as corpus linguistics research either. Linguistic investigation *sensu* Teubert (2005) should be carried out on real language data (i.e., for translation unit analyses, original and translated texts); however, Alves and

Gonçalves (2003) and Lörscher (1991) primarily analysed materials, such as translator's keyboard and vocalised activities, which cannot be viewed as real language data, that is, as wholly representative of the discourse community of which insights are sought (Teubert, 2005; see details in 2.2.2). Even the studies by Toury (1986), Teubert (2001), and Vinay and Darbelnet (1995), discussed in 2.3, are hard to define as corpus studies on translation units in a strict sense. Their corpora were not compiled from the discourse in a 'principled' way. Whilst Teubert (2005) does not go as far as to clarify what is meant by 'principled', my interpretation is that language data in corpora should be selected in an unbiased manner. The corpora of Toury (1986), Teubert (2001), and Vinay and Darbelnet (1995) contain only one or a few paragraphs. It is often the case that if a corpus contains a few texts, the selection of them necessarily tends to be subjective, and therefore biased. Therefore, their corpora are not ideal examples of what corpus studies can achieve above other approaches to linguistic problems.

The current thesis achieves the above suggested requirements of corpus linguistics: it works only with real language texts (i.e. original and translated texts) compiled systematically into a large corpus. This study, therefore, uniquely demonstrates what corpus study can contribute to the problem of translation unit identification and size.

1.1. Organisation of the thesis

This thesis is organised as follows: Chapter 2 will present a systematic review of translation unit size. By saying 'systematic', I mean that the choice of the reviewed literature was not biased. All the accessible literature was examined; these were the

studies found by using, (a) two search engines, the ISI Web of Knowledge and ebrary (search word was 'translation unit'), and, (b) several core reference books such as the Dictionary of Translation studies (Shuttleworth and Cowie, 1997), the Encyclopedia of Language and Linguistics (Brown et al., 2006), and the Routledge Encyclopedia of Translation Studies (Baker, 1998); all of the articles and books mentioned in the retrieved articles and reference books as further reading were reviewed as well.

Chapter 3 will validate the identification method of translation units used in this thesis. Since this study establishes a new method to extract translation units, the reasons why the existing methods were not used are clarified. The next three chapters will be dedicated to case studies. The studies are divided by word classes of target words. Translation units of the noun target words will be identified in Chapter 4; translation units of the adjective target words in Chapter 5; and, translation units of the verb target words in Chapter 6. Lastly, all the findings are gathered and interpreted in Chapter 7, followed by the conclusion chapter, Chapter 8.

Also, as a technical note, I add that this thesis follows the standard citation rules for the University of Birmingham: a customised version of the Harvard System of referencing and bibliography[1]. One notable trait of these rules is that citations in the text do not include page numbers.

---

[1] http://www.i-cite.bham.ac.uk/Harvard.shtml (accessed on 19 November 2009)

1.2. Terms

It is useful to state clearly, at this early stage, that the three terms, translation unit, correspondence, and equivalence, are different in this thesis. First, a translation unit is a 'minimal stretch of language that has to be translated together, as one unit' (Newmark, 1988b). Translators carry out their tasks by these units. Studies of translation units usually aim to identify such translation units; translation units are therefore an outcome of such study. Second, equivalence refers to 'the relationships which exist between SL [=source language] and TL [=target language] texts or smaller linguistic units' (Shuttleworth and Cowie, 1997). Therefore, one can say that a pair of translation units in a language and its corresponding item in other language has equivalence. The corresponding item in the TL is called a 'translation equivalent' in this thesis. Many linguists indeed use the term 'equivalent' for it: e.g. 'equivalent' (Barkhudarov, 1993); 'TL equivalent' (Newmark, 1988b); and, 'translation equivalent' (Teubert, 2004b). On the other hand, some theorists use other terms such as: 'TL segment' (Toury, 1986); 'corresponding segment' (Toury, 1986); and, 'TL unit' (Vinay and Darbelnet, 1995; Shuttleworth and Cowie, 1997). Taken into account that equivalence is found in translated texts, the term 'translation equivalent' is more accurate.

Lastly, the term 'correspondences' refers to the pairs of words or constructions 'which are found to correspond' between original and translated texts (Johansson, 2007). They are not identical to the pairs of translation units and equivalents (cf. Kondo, 2007). Studies of correspondences usually aim to identify similarities and differences between the correspondences. Therefore, correspondences are not outcomes, like translation

units; rather, they are starting points. For example, English *person* and Norwegian *person* are correspondences (Johansson, 2007). They are examined to see whether they are the closest match between the two languages or not. What this study concerns is translation units, not correspondences.

1.3. The representation of Japanese data

Lastly, it is worth mentioning about the representation of Japanese data in this thesis. There are two issues to clarify here: romanisation and word segmentation. First, the Japanese language has a 'notoriously complex writing system', according to Quint (2003). It uses 'Chinese characters (KANJI) in combination with two separate forms of the phonetic syllabic script known as KANA; the names of the latter are *hiragana* and *katakana* [original emphasis]' (Grootaers, 1983). Taking an example from the ARC, the sentence 労働市場がタイトである 'labour market is tight' consists of: four kanji characters (e.g. 労働市場 'labour market'); four hiragana characters (e.g. が 'subject market' and である 'is'); and, three katakana characters (e.g. タイト 'tight'). However, for the sake of clarity and readability for non-Japanese speakers, all the Japanese examples will be romanised in this study. There are three recognised systems of romanisation, the Hepburn, Kunreisiki, and Nipponsiki systems (Itasaka, 1983; Coulmas, 1996; Unger, 2006), and the Hepburn system was used to represent Japanese data in the current study. This was because the Hepburn system is the most traditional and is used very widely within English texts (Totten III, 1983; Coulmas, 1996; Jorden, 2000; Unger, 2006).

There are, however, some disagreements on what the Hepburn system actually is; in particular, the usage of macrons (to indicate long vowels, e.g. *ē*, *ā*, and *ō*) is a main cause of variation. For example, 労働市場 'labour market' can be romanised into *rōdō shijō* or *rōōdōō shijōō* in studies where macrons are used (e.g. **Kodansha Encyclopedia of Japan**, 1983; **Basic Japanese-English dictionary**, 1986; **Japanese For Busy People**, 1994; Kamiya, 2001; Bunt, 2003; Unger, 2006; Tranter, 2008). On the other hand, in studies where macrons are not used, the same word is romanised into *roodoo shijo* or *roudou shijou* (e.g. Terakura, 1983; Uchida and Nakayama, 1993; Tsujimura, 1996; Taira, 1998; Batchelder and Ohta, 2000; Cipris and Hamano, 2002; Iwasaki, 2002; Ohara et al., 2003). There are some other variables that create minor variations: the usage of apostrophes (to separate the *n* sound from a following vowel, e.g. *shin'ai* 'dear'), and the spelling of *m* (before *b*, *m*, or *p* consonants, e.g. *shimbun* 'newspaper'). These can cause many possible romanisations for the same data even within the Hepburn system.

Comprehensive guidelines for romanisation can help to alleviate these problems of the Hepburn system. One such set of guidelines is presented in Iwasaki's study (2002, see the Appendix 1). By using these guidelines, different researchers should produce the same romanisations. For example, the romanisation of 労働市場 'labour market' is *roodoo shijoo*. Due to this repeatability, this thesis followed Iwasaki's guidelines. Macrons were not used; instead, a sequence of two vowels, e.g. *ee*, *aa*, and *oo* was used for long vowels. Apostrophes are used, which is useful for distinguishing homographs: Iwasaki's examples (2002) are *shin'ai* 'dear' and *shinai* 'do not do'.

Word segmentation is another thorny issue in the representation of Japanese data. As one can see in the above Japanese example (労働市場がタイトである 'labour market is tight'), there is no space between the words in the sentence. Reflecting this nature, the romanised sentence becomes *roodooshijoogataitodearu*. This, however, makes the sentence very hard to understand, even for native Japanese speakers. This is because, as Kaiser (2003) points out, word boundaries are often obvious where kanji items start and kana items end; 'kanji … would typically appear after a word boundary' and 'kana tend to appear at the end of a word rather than the beginning'. Without the help of kanji and kana, the romanised sentence seen above provides few clues to word boundaries. Such a representation is obviously useless for readers; therefore, in this study, all the Japanese data were segmented by smaller units.

Segmentation of Japanese data was made using the software ChaSen (Matsumoto et al., 2002) in the ARC (Utiyama and Isahara, 2003). This helped to make the segmentation in this thesis consistent; such consistency would be harder to maintain without this software. This is because there is no consensus on the division of romanised Japanese examples into individual words (Itasaka, 1983). Compound words, such as 労働市場 'labour market', are the main issue, in that segmentation is hard to agree on, for example, (a) *roodooshijooo* 'labour market', (b) *roodoo shijoo* 'labour' and 'market', and (c) *roodoo-shijoo* 'labour-market' are all possible segmentations. Verbs are not easy to segment either, e.g. 見られている 'is being looked'. The possibilities are: (d) *mirareteiru* 'is being looked'; (e) *mirare teiru* 'be looked' and 'progressive marker'; and, (f) *mi rare teiru* 'look', 'passive marker', and 'progressive marker', etc. The software

8

ChaSen segments 労働市場 into (b) *roodoo shijoo* and 見られている into (g) *mi rare te iru*. Therefore, no subjectivity is involved in dividing words.

Lastly, there are three distinctive Japanese symbols in news texts: the comma 、; the period 。; and, the quotation mark 「」. These were replaced in the representations by the English comma, period, and quotation mark respectively.

2. Translation unit size

It would not be an overstatement to say that there is substantial disagreement concerning translation unit size in the literature (Figure 2.1). Some theorists maintain that translation units can occur at almost all language unit lengths, i.e. the morpheme, single word, phrase, clause, sentence, paragraph, and whole text (Newmark, 1988b; Barkhudarov, 1993). On the other hand, some argue that translation units are restricted to only one of these lengths (Goldman-Eisler, 1972; Bassnett-McGuire, 1980; Bell, 1991; Isham and Lane, 1993; Alves and Gonçalves, 2003). Other theorists claim that translation units can occur at a selection of language unit lengths (Tancock, 1958; Nida and Taber, 1969; Toury, 1986; Lörscher, 1991; 1993; Vinay and Darbelnet, 1995; 1996; Teubert, 2001; Barbosa and Neiva, 2003; Livbjerg and Mees, 2003). This chapter aims to clarify these diverse and overlapping viewpoints on translation unit size by presenting a systematic review of the literature. It will be different from existing reviews, which discuss translation units in general (Shuttleworth and Cowie, 1997; Malmkjær, 1998; Luo, 1999; Zhu, 1999; Munday, 2001; Hatim and Munday, 2004; 2006; Kenny, forthcoming). The current review sharpens its focus on the matter of size; in particular, it will attempt to clarify which are the key factors associated with the disagreement.

Figure 2.1. Selected views on translation unit size

	Morpheme	Word	Phrase	Clause	Sentence	Paragraph	Text
Alves and Gonçalves (2003)			■				
Barbosa and Neiva (2003)		■	■	■	■	■	
Barkhudarov (1993)	■	■	■	■	■	■	■
Bassnett-McGuire (1980)							■
Bell (1991)				■			
Goldman-Eisler (1972)				■			
Isham and Lane (1993)				■			
Livbjerg and Mees (2003)		■	■	■			
Lörscher (1991)			■	■			
Newmark (1988b)	■	■	■	■	■	■	
Nida and Taber (1969)					■	■	
Tancock (1958)		■			■	■	
Teubert (2001)		■	■	■			
Toury (1986)			■	■			
Vinay and Darbelnet (1995)	■	■	■	■			

2.1.   Disagreement in the literature

The disagreement on translation unit size can be said to be due to disagreement on what 'translation unit' actually means. There are two main senses in which the term translation unit is used. One refers to cognitive units; these are typically defined as 'segments of the source text … [on which] the translator's focus of attention is directed' (Alves and Gonçalves, 2003). The other refers to a lexical unit; it is usually defined as a 'minimal stretch of language that has to be translated together, as one unit' (Newmark, 1988b). This distinction is argued by Malmkjær (2006) as well. Based on a traditional division of the translation process and its products, the cognitive viewpoint can be called a 'process-oriented' translation unit; whilst the lexical translation unit is 'product-oriented'. Malmkjær's claim (2006) is that this dichotomy causes a difference in the methods of identification of translation units only, without commenting on

11

whether these methods would lead to the identification of different units or not. The process-oriented unit can be identified by observing the mental processes occurring in a translator's mind; whilst the product-oriented one can be identified by observing original and translated texts (Malmkjær, 2006). If this is the case then one might also wonder whether these different definitions result in the recognition of different translation units of different sizes.

Bennett (1994) is one linguist who highlights this association; his definitions are closely related to translation unit size. Distinguishing cognitive from lexical translation units, Bennett (1994) coins two terms: 'translation focus' for the cognitive translation unit ('the section of text which the translator focuses on at any one time'); and, 'translation atom' for the lexical translation unit ('the smallest segment that must be translated as a whole'). His claim regarding the association between translation unit type and size is that the former type of unit tends to be larger than the latter. Since Bennett (1994) puts forward little linguistic evidence for this claim, I examined the translation unit literature to see whether his claim can be supported or not. The examined literature comprised: Tancock (1958); Nida and Taber (1969); Goldman-Eisler (1972); Bassnett-McGuire (1980); Toury (1986; 1995); Newmark (1988b; 1988a); Lörscher (1991; 1993; 1996); Bell (1991); Barkhudarov (1993); Isham and Lane (1993); Vinay and Darbelnet (1995); Shuttleworth and Cowie (1997); Malmkjær (1998; 2006); Danielsson (2001); Teubert (2001; 2002; 2004b); Alves and Gonçalves (2003); Barbosa and Neiva (2003); Livbjerg and Mees (2003); Chang et al. (2006); and, Bernardini (2007). These were all the studies of translation unit size able to be accessed within the resources of this thesis (discussed in 1.1).

2.2. Cognitive unit

2.2.1. Definitions

In order to examine Bennett's claim (1994), the literature was first examined to see if there was actually such a considerable distinction between the two types of definitions: that is, between cognitive and lexical translation units. If the difference between the two is large, their attributes, such as size, are likely to be different, supporting Bennett's claim (1994). This section, therefore, will explore how cognitive translation units are defined in the literature. The studies examined for this aim include: Lörscher (1991; 1993; 1996); Malmkjær (1998; 2006); Alves and Gonçalves (2003); and, Livbjerg and Mees (2003).

The cognitive sense of a translation unit has three main characteristics, as shown in the definitions in Table 2.1. First, as mentioned in 2.1, it is concerned with the translator's mental activities. All the definitions in Table 2.1 include terms relating to this: 'mind' (Malmkjær, 2006); 'attention' (Lörscher, 1991; 1993; 1996; Malmkjær, 1998; Alves and Gonçalves, 2003); and, 'doubt' (Livbjerg and Mees, 2003). Second, the cognitive translation unit is a unit of the original text. The theorists in Table 2.1 always maintain that the unit is a segment of 'SL [= source language] text' (Lörscher, 1991; 1993; 1996) or 'source text' (Malmkjær, 1998; Alves and Gonçalves, 2003; 2006). Third, the cognitive translation unit is an inseparable unit in translating. It is usually defined as a unit to render 'as a whole' (Lörscher, 1991; 1993; 1996; Malmkjær, 1998). Put very generally, therefore, the cognitive translation unit is an inseparable unit of the original text at which the translator's focus of attention is directed.

13

Table 2.1. Selected definitions of cognitive translation units

Theorists	Definitions[2]
Alves and Gonçalves (2003)	'segments of the source text, independent of specific size or form, to which, at a given moment, the translator's focus of attention is directed'
Livbjerg and Mees (2003)	'any word or phrase in the text, or any aspect of such a word or phrase, which is verbalised by any single participant and for which he or she expresses any degree of doubt about its proper translation'
Lörscher (1991)	'an SL text segment which, at a given point in time, s/he puts into his/her focus of attention in order to render it into the TL *as a whole* [original emphasis]'
Lörscher (1993)	'the SL text segments which a subject extracts and puts into his/her focus of attention in order to render them into the target-language *as a whole* [original emphasis]'
Lörscher (1996)	'the SL text segments which the subjects extract and put into their focus of attention in order to render them into the target-language *as a whole* [original emphasis]'
Malmkjær (1998)	'the stretch of source text on which the translator focuses attention in order to represent it as a whole in the target language'
Malmkjær (2006)	'the stretch of the source text that the translator keeps in mind at any one time, in order to produce translation equivalents in the text he or she is creating'

2.2.2. Identifications

Identifying such translation units is very hard to achieve. It requires access to the translator's mind, which is obviously not directly possible. Many attempts at identifying such translation units have been made with the help of technology. Alves and Gonçalves (2003), for example, carried out an experiment in which they asked four translators to render English texts into Portuguese on computers; all the translators'

---

[2] SL (i.e. source language) refers to the original language; TL (i.e. target language) refers to the translated language (Munday, 2001).

keyboard activities during the tasks were recorded by the programme, Translog. The recordings, in particular, uncovered the pause intervals, i.e. when translators typed and paused. For example, in the experiment using the sentence, *By 1998, the year of the U.S. Embassy bombings in Africa, he had acquired the lean, wolfish look of a revolutionary*, the keyboard recordings showed that the translators rendered the nominal phrase, *the U.S. Embassy bombings in Africa*, at once stretch; they typed the Portuguese translation without pauses. On the contrary, there were substantial pauses before and after translating this nominal phrase. Alves and Gonçalves interpreted such substantial pauses as the boundaries between translation units and identified *the U.S. Embassy bombings in Africa* as a translation unit. Another nominal phrase in the sentence, *the lean, wolfish look of a revolutionary*, was also regarded as a translation unit since it was segmented by such pauses.

One might wonder whether the other items in the sentence (such as *by 1998*, *the year of*, *he had acquired*) can be regarded as translation units or not. Although all the three expressions seem to be divided by pauses, Alves and Gonçalves (2003) show little interest in them. This is because their concern was not with the items that translators rendered easily, such as *by 1998*. Rather, they are more concerned with the items that translators have difficulties of rendering, such as *U.S. Embassy bombings in Africa*, where there may be problems of ambiguity. Most of the translators wondered if there were bombings to one or more than one embassy, and so how the phrase should be rendered. This predilection for such difficulties is indeed what studies on cognitive translation units mainly focus on; they tend to be 'problem-oriented', according to Kenny (forthcoming). One extreme example of this trend can be seen in Livbjerg and

Mees's definition (Table 2.1); for them, a translation unit is a unit which translators have problems in rendering. Based on their definition, then, none of the phrases, *by 1998*, *the year of*, and *he had acquired* would be regarded as translation units; these items were translated without any problems, which was confirmed by the keyboard recordings and interviews after the tasks (Alves and Gonçalves, 2003).

The predilection for translation problems is reasonable, if one's interest is the translator's mind. This is because the translation problem is, indeed, a repository of the mental processes of how translation units were determined by translators. For example, Lörscher (1991) carried out an experiment in which he asked translators to render the German clause, *In jüngerer Zeit sind mehrfach Versuche unternommen worden*, into English. Lörscher (1991) used a think-aloud method (i.e. asked translators to verbalise his/her thoughts during translation); the translator verbalised that '*In recent times*, ehm (2s [= pause of two seconds]), *many attempts have been made*' (italics indicate the translation of the original text). The translator did not vocalise any problems, which means that the task was carried out very smoothly. What the two-second-pause indicates is that the original clause was divided into two translation units: *In jüngerer Zeit* and *sind mehrfach Versuche unternommen worden*. This analysis succeeded in identifying the translation units in a relatively straightforward manner. However, this result is less attractive to most cognitive linguists. The translator divided the original clause into two translation units too quickly to verbalise how and why s/he did it, as if it was an automated process. The analysis gives little clues to such mental processes determining the translation units. Text which can be translated smoothly, like the above example, is not a good resource for observing mental activities.

On the other hand, texts of translation problems provide much information about mental activities. For instance, in the experiment on the sentence, *By 1998, the year of the U.S. Embassy bombings in Africa, he had acquired the lean, wolfish look of a revolutionary*, most translators found it difficult to understand what *wolfish* meant (Alves and Gonçalves, 2003). At the interview after the task, one of the translators said, "I had some doubts with vocabulary for example, wolfish. I didn't find it in a dictionary, so I used a reconstruction using context" (Alves and Gonçalves, 2003). This indicates how the translator established a translation unit of *wolfish*; s/he could not understand and render the meaning of *wolfish* as it was, therefore, the wider context was reconstructed to allow the finding of a suitable translation unit. More specifically, another translator mentioned that 'I had to use the dictionary to search for better terms for some words like acquired the lean, wolfish look and so on. I used literal translation many times, paraphrasing things like *adquiriu o gusto, a face selvagem, palidez mórbida* and so on' (Alves and Gonçalves, 2003). Again, it shows the process of determining a translation unit of *wolfish*; the translator could not render *wolfish* as it was, therefore, the larger unit *acquired the lean, wolfish look* was considered as a translation unit. Examining these data of translation problems makes it possible to reveal how translation units were taking shape in the translator's mind. Translation problems, thus, are valuable resources for cognitive translation units.

To sum up, the cognitive translation unit is concerned with the translator's mental processes during translation, i.e. how the translators determine and create translation units in their mind. This makes the research problem-oriented, since translation problems are posited to provide an indirect insight into mental processes. One

application of such studies would be translation pedagogy; cognitive translation units can instruct novices how to render expressions which do not have established equivalents in another language.

2.3.   Lexical unit

2.3.1. Definitions

So, what about the lexical translation units? Is it considerably different from the cognitive unit? Or, is the difference rather insignificant? This section is devoted to such questions, and aims to find out, exactly, what a lexical translation unit is. The examined studies include: Newmark (1988b; 1988a); Barkhudarov (1993); Toury (1995); Vinay and Darbelnet (1995); Shuttleworth and Cowie (1997); Malmkjær (1998; 2006); Danielsson (2001); Teubert (2001; 2002; 2004b); Chang et al. (2006); and, Bernardini (2007). Table 2.2 shows all the definitions of lexical translation unit to be found in the literature.

Table 2.2. Selected definitions of lexical translation units

Theorists	Definitions
Barkhudarov (1993)	'the *minimal* language unit in the source text that corresponds to an equivalent in the target text [original emphasis]'
Bernardini (2007)	'bilingual collocation pairings'
Chang et al. (2006)	'a Translation Unit is a single or multiword unit in a language. Together with its context information, a Translation Unit is translated into another language unambiguously'
Danielsson (2001)	'A unit of translation (UT) can be defined as a relation between UMs [= units of meaning] or sets of UMs.'
Malmkjær (1998)	'the target-text unit that can be mapped onto a source-text unit.'
Malmkjær (2006)	'translation units are pairs of source and target text segments'
Newmark (1988a)	'the source-language unit which can be recreated in the target language without addition of other meaning elements from the source language'
Newmark (1988b)	'minimal stretch of language that has to be translated together, as one unit'
Newmark (1988b)	'the smallest segment of an SL text which can be translated, as a whole in isolation from other segments.'
Shuttleworth and Cowie (1997)	'the linguistic level at which ST is recodified in TL.'
Teubert (2002)	'the text segments that are translated as a whole'
Teubert (2004b)	'We do not translate single words in isolation but units that are large enough to be monosemous, so that for them there is only one translation equivalent in the target language, or, if there are more, then these equivalents will be synonymous. I call these units translation units'
Toury (1995)	'the linguistic-textual unit in the original text within which the translator tended to work'
Vinay and Darbelnet (1995)	'the smallest segment of the utterance whose signs are linked in such a way that they should not be translated individually'
	'lexicological units within which lexical elements are grouped together to form a single element of thought'

The definitions in Table 2.2 allow one to recognise three characteristics of lexical translation units. First, the lexical translation unit is again a unit of the original text: 'source-language unit' (Newmark, 1988a); 'segment of an SL text' (Newmark, 1988b); 'unit in the source text' (Barkhudarov, 1993); and, 'unit in the original text' (Toury, 1995). It is true that there are two other minor claims. One is that the lexical translation unit is a unit of the translated text (Malmkjær, 1998); another is that the lexical translation unit is a pair of original and translated expression (Danielsson, 2001; Malmkjær, 2006; Bernardini, 2007). However, few researchers have yet to subscribe to these positions. Second, the lexical translation unit is an inseparable unit in translating: a unit to translate 'as a whole' (Newmark, 1988b); 'as one unit' (Newmark, 1988b); and, a unit which 'should not be translated individually' (Vinay and Darbelnet, 1995). These two characteristics are actually in common with the definitions of cognitive translation units.

The third characteristic is distinctive: the lexical translation unit shows little interest in mental processes. None of the definition in Table 2.2 includes words denoting mental activities such as 'mind' or 'attention'. What the lexical translation unit is concerned with is the equivalence relationship between the translation unit in one language and its correspondents in another language; so we have, 'corresponds' (Barkhudarov, 1993); 'linked' (Vinay and Darbelnet, 1995); 'recodified' (Shuttleworth and Cowie, 1997); 'recreated' (Newmark, 1988a); and, 'mapped onto' (Malmkjær, 1998). The main interest in these studies is on whether translation equivalence is established on a word-for-word basis; and, if not, then on a what-unit-for-what-unit basis. Such equivalence is usually examined by comparing original and translated texts. Put very

generally, therefore, the lexical translation unit is an inseparable unit of the original text which corresponds to its equivalent in a translated text.

It is worth mentioning that the translation unit in this sense is usually regarded as a social phenomenon. When the German clause, *In jüngerer Zeit sind mehrfach Versuche unternommen worden*, was rendered into *In recent times many attempts have been made*, as discussed in 2.2, the original clause was instantly divided into the two translation units: *In jüngerer Zeit* and *sind mehrfach Versuche unternommen worden* (Lörscher, 1991). Why did this happen? It is because the translator was familiar with both phrases *In jüngerer Zeit* and *sind mehrfach Versuche unternommen worden* and knew their English equivalents. In other words, he or she had known the translation pairs: (a) *In jüngerer Zeit* and *In recent times*; and, (b) *sind mehrfach Versuche unternommen worden* and *many attempts have been made*. So, how and why had he or she known this information?

Teubert (2001) claims that established translation pairs, such as *In jüngerer Zeit* and *In recent times*, are stored in 'a universe of discourse' ('the sum of all communication acts ever uttered by members of a language community'). All the translation pairs (i.e. translation units and their equivalents) 'have been proposed, tested, and established over time' in this discourse. The new translation pair is proposed, '[o]nly when translators are faced with the lack of a "ready" equivalent' (Barkhudarov, 1993); it goes into the discourse, will be tested by other community members, and might be established if they accept it. This universe of discourse is, therefore, a repository of translation pairs. And,

it is this discourse that linguists have to examine in order to identify lexical translation units.

2.3.2. Identifications

Considering that 'a universe of discourse' is a repository of the lexical translation units and their equivalents, it would be ideal if there was a corpus consisting of all the translations ever rendered. However, such a corpus can only exist, unfortunately, as the linguist's equivalent of a Borgian library! Therefore, smaller parallel corpora ('made of originals in language A and their translations into language B' (Bernardini, 2007)) are analysed as samples of such discourse.

One key piece of research on translation units using a small data set is by Vinay and Darbelnet (1995); this work could be considered a precursor to corpus-based analyses. Vinay and Darbelnet (1995) analysed a French academic paragraph and its English translation, and identified translation units and their equivalents. The first examined sentence had six translation units, *Si nous songeons/ à une 'ère nouvelle',/ c'est dans la mesure/ où nous pouvons ressentir,/ sinon définir,/ une 'ère à dépasser'* (/ means the boundaries of translation units); each translation unit corresponds to the English equivalent: *If we speak of/ a new era/ we can do this only to the extent/ that we can feel/ or identify/ an era that is past* (again, / is the boundaries of the translation equivalents). The identification was based on 'the degree of cohesion' between the elements; the unit is regarded as a translation unit when the items are either (a) fixed expressions which are translated together, i.e. idioms, or (b) less fixed but translated together, i.e. phrases

(1995). This does not deny the possibility of single words as translation units; the single word can be regarded as a translation unit when the item is translated alone (further relevant discussion will be made in 3.3.1).

Another piece of reseach carried out with a small, subjectively chosen, data-set is by Toury (1986), consisting of an English passage and its 27 different Hebrew translations; Johansson (2007) calls this type of corpus a 'multiple translation corpus', which has originals and their translations by more than one translator, rather than a 'parallel corpus', which is a common term still. Toury's identification of translation units was based on the 'no leftover' principle (1986; 1995). He assumes that all the elements in the original texts are 'represented by' their counterparts in translation (1986). Vinay and Darbelnet (1995), in fact, also used this criterion, apparently unconsciously, in the above example sentence, *Si nous songeons/ à une 'ère nouvelle',/ c'est dans la mesure/ où nous pouvons ressentir,/ sinon définir,/ une 'ère à dépasser'* (*If we speak of/ a new era/ we can do this only to the extent/ that we can feel/ or identify/ an era that is past*). If one thinks of *c'est dans la mesure*, there is no expression in the English sentence to be semantically and syntactically equivalent (according to the online translation service, AltaVista [www.altavista.com], *c'est dans la mesure* means 'it is in measurement'). Therefore, *c'est dans la mesure* could have been leftover without being linked to any English items. Vinay and Darbelnet (1995), however, matched *c'est dans la mesure* with *we can do this only to the extent* since the meaning of *c'est dans la mesure* is, to some extent, represented by this phrase. This is exactly what the 'no leftover' principle suggests one should do. It gives flexibility when one matches translation pairs syntactically and/or semantically different. Such flexibility is necessary, since a

translation equivalent in a target language is not always identical in its structure and exact meaning to its translation unit in the source language; '[w]hat is expressed in one language … could be expressed in another language in quite different ways' (Johansson, 2007).

Both the degree of cohesion and the no leftover principle are useful criteria for identifying lexical translation units and their equivalents. However, these criteria inevitably involve manual examination of the original and translated texts; this is rather a time consuming method of analysis, and would not be possible as data sets get larger. Seeing that the lexical translation unit is regarded as a social phenomenon, linguists are keen to analyse translation units in corpora larger than the ones used by Vinay and Darbelnet (1995) and Toury (1986). What methods are available for this purpose, then?

Wang (2006) identified translation units in a large parallel corpus consisting of more than 10 million words ('approximately 5.6 million English words and 4.6 milion Chinese characters'): the Hong Kong Legal Document Corpus (HKLDC). She focused on frequent two-word phrases which were composed of an adjective and a noun, e.g. *straight line*. In particular, she extracted thirty of the most frequent of these phrases and analysed if they were translation units or not. The criterion used was the monosemous principle, which was suggested by Teubert (2004b); '[w]e do not translate single words in isolation but units that are large enough to be monosemous, so that for them there is only one translation equivalent in the target language, or, if there are more, then these equivalents will be synonymous. I call these units translation units'. An item can be regarded as a translation unit, (a) if all examples of the item in the corpus were rendered

into one single translation equivalent, or, (b) if all examples of the item in the corpus were rendered into several but synonymous translation equivalents. By targeting frequent phrases, Wang (2006) managed to use this criterion and extracted translation units out of the large parallel corpus, the HKLDC. Danielsson (2001) is another linguist who has carried out a quantitative analysis of translation units, but from a Swedish-English perspective. Her method is also based on frequent phrases (further relevant discussion will be made in 3.3.4).

2.4.   What causes the disagreement?

2.4.1.   The dichotomy between the definitions

The similarities and differences between the cognitive and lexical translation units should by now be obvious. Both refer to an inseparable unit of the original text in translating. However, the cognitive translation unit represents a unit towards which the translator's focus of attention is directed. In other words, the translation unit is regarded as a mental phenomenon, i.e. individual translators invent translation units in their minds each time they render part of a text, which means that the individual mind is a resource of translation units in this sense. For the purpose to access this resource, recordings of translator's thoughts and activities during translation (i.e. keyboard activities, interviews, and think-aloud protocols) are indispensable; they have to be used for the identification of translation units. These studies are useful for understanding translation strategies, and so teaching them to the novices.

On the other hand, a lexical translation unit represents a unit which corresponds to its equivalent in translated text. Since the translation unit in this sense is not regarded as a mental phenomenon, no special recording is needed for identification. Rather, the unit is regarded as a social phenomenon, and the universe of discourse is a repository of lexical translation units. Therefore, parallel corpora, as samples of such discourse, can be used for the identification of translation units. A careful comparison of the original and translated texts in a given corpus is the key method for identification. There are some criteria, such as the degree of cohesion (Vinay and Darbelnet, 1995), the no leftover principle (Toury, 1986; 1995), units of meaning (Danielsson, 2001), and the monosemous principle (Teubert, 2004b), which have been conceived in order to help in identification. Studies on lexical translation units, therefore, aim to extract translation pairs which have been established over time in the discourse. An application would be in the compilation of bilingual dictionaries and glossaries, and for translation memory (i.e. 'a database … of a pair of source and target text segments' (Reinke, 2006)).

Is this dichotomy between theoretical translation unit types related to the issue of translation unit size? As discussed in 2.1, Bennett (1994) clearly expresses this association between translation unit size and research perspective. According to him, cognitive translation units are larger, i.e. at any size between single word and an entire text; 'a translator may go from focusing on the translation of one single word to that of a phrase, then a clause and sentence, and finally an entire text' (1994). On the other hand, lexical translation units are more manageable and smaller, e.g. between morphemes and sentences (Bennett, 1994); morpheme refers to 'a grammatical unit which is used to constitute words' (Hartmann and James, 1998). This is because 'no translator can work

with any text other than the very shortest as an undivided UT [=unit of translation]' (Bennett, 1994). This trend is, in fact, consistently found within other relevant case studies.

Studies on lexical translation units suggest that their translation units are, as Bennett (1994) argues, found at some linguistic length between the morpheme and the sentence (Figure 2.2). Toury (1986), for example, as discussed in 2.3, identified lexical translation units based on 27 Hebrew translations of a written English passage; the translation units identified were mainly phrases and clauses. Therefore, Toury's analysis (1986) overlaps with Bennett's (1994) claims on lexical translation unit size. Vinay and Darbelnet (1995) examined translation unit size from the English-French perspective, arguing that translation units can be either fractions of words, single words, and several words. By saying 'a fraction of a word', these authors mean the morpheme; 'several words', can be assumed by their examples, e.g. *nocks and crannies* and *to report progress*, to include phrases and clauses.

Similarly, Teubert's corpus analysis (2001) agrees with Bennett's claim on lexical translation unit size (1994). Teubert's examined texts were 'French newspaper texts and their English or German translations, which contain about a thousand words in each language' [personal communication]. The analysis was rather unique in the sense that Teubert used a bilingual dictionary; he examined if one could translate a word by looking up its equivalent in a comprehensive bilingual dictionary. Teubert concluded that translation units consist of 'a single word or of several words'. Due to the different corpora and the methodologies used, the results of these empirical studies all differ slightly. However, all of their analyses validate Bennett's view on the size of lexical

translation units: the units are likely to be between the morpheme and the sentence.

Figure 2.2. Lexical translation unit size and definitions[3]

	mo.	w.	ph.	cl.	sent.	par.	tex.
Bennett (1994)		■	■	■	■		
Newmark (1988b)		■	■	■	■		
Teubert (2001)		■	■				
Toury (1986)			■	■			
Vinay and Darbelnet (1995)	■	■	■	■			

Unsurprisingly, given the breadth of possibilities, studies on the cognitive sense of translation units do suggest that their translation units are at some linguistic length between a single word and an entire text (Figure 2.3); but do they show any indication of differences in size between cognitive and lexical translation units? Lörscher (1993) maintains, from empirical studies, that these translation units are phrases, clauses, and sentences. He also points out that smaller units, such as words and morphemes, can cause bad translations, which frequently happens in translations carried out by novice translators: they tend to focus their attention on smaller units. Alves and Gonçalves (2003) identified only two translation units in their experimental text: *U.S. Embassy bombings in Africa* and *the lean, wolfish look of a revolutionary*. However, both units are phrases, which also lends supports to the idea that cognitive translation units might be bigger that lexical ones. Livbjerg and Mees (2003), as one can see from their definition of the translation unit (Table 2.1), clearly state that the unit is either a word or a phrase. Again, a size difference between cognitive and lexical translation units is hinted at, on average.

---

[3] mo. = morpheme level; w. = single word level; ph. = phrase level; cl. = clause level; sent. = sentence level; par. = paragraph level; tex. = text level.

Figure 2.3. Cognitive translation unit size and definitions

	mo.	w.	ph.	cl.	sent.	par.	tex.
Alves and Gonçalves (2003)			■				
Barbosa and Neiva (2003)		■	■	■	■		
Bennett (1994)		■	■	■	■	■	■
Goldman-Eisler (1972)				■			
Livbjerg and Mees (2003)		■	■				
Lörscher (1991)			■	■			
Nida and Taber (1969)						■	■
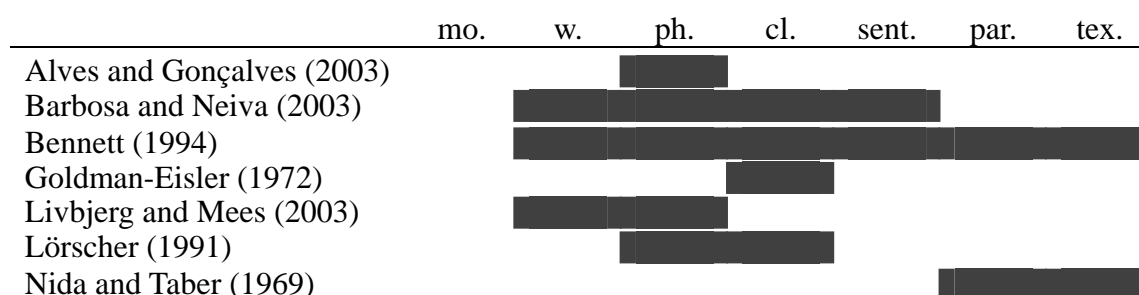
Research by Nida and Taber (1969) is also worth mentioning here. It is fair to note first that Nida and Taber do not use the term 'translation unit', but they use another term, the 'focus of attention in translation'. Seeing that most of the cognitive translation units are defined by the focus of attention (Table 2.1), it is appropriate to discuss their claim of translation unit size here. Nida and Taber maintain that the focus of attention in translation should be at the level of the paragraph and total discourse, which are much higher levels than Lörscher (1993), Alves and Gonçalves (2003), and Livbjerg and Mees (2003) suggest.

The literature, therefore, on balance, seems to validate Bennett's claim (1994) that translation unit size is closely related to whether researchers are working from a cognitive or lexical unit perspective. On average, the lexical unit has a range of smaller sizes and the cognitive unit has a range of larger sizes; this tendency is seen by comparing the distribution of categories in Figures 2.2 and 2.3. This would explain why some linguists define the lexical translation unit as the 'smallest' unit (Table 2.2). Newmark (1988b), for example, defines it as 'the smallest segment of an SL text which can be translated, as a whole'; similarly, Barkhudarov (1993) defines it as the 'minimal'

unit, Vinay and Darbelnet (1995) as the 'smallest', and Toury (1986) as a 'manageable' unit. The smallness is apparently associated with the lexical sense of the translation unit. By defining translation units as 'smallest', linguists manage to cut off the cognitive sense of the translation unit, and thereby imply that their translation units are lexical ones. On the other hand, no cognitive linguists define their translation units as the 'smallest' units (Table 2.1), reinforcing the finding that smallness is associated only with the lexical sense of the translation unit.

As discussed so far, it can be seen that there are two standpoints where linguists use the term 'translation unit': the cognitive and the lexical. This dichotomy is closely related to the issue of translation unit size. The cognitive sense of the translation unit is likely to be larger; on the other hand, the lexical one is likely to relate to smaller units (2.4.1). However, this dichotomy alone cannot explain the wide diversity of views on translation unit size held among linguists (Figure 2.1). A careful comparison of diverse case studies makes it clear that there are three more possible key variables associated with the size issue: the language pair being considered (2.4.2); the type of text (2.4.3); and who is doing the translating (2.4.4).

2.4.2. Language pairs

Translation unit size is also related to which language pair one is considering. Teubert argues that '[w]hat is a translation unit in relation to one target language does not have to be one in relation to another' (2002). One of his examples is the phrase *false alarm* (Teubert, 2004a). *False alarm* is a translation unit, when it is examined from the

30

English-German perspective. This is because *false alarm* has to be translated as a whole into the German equivalent *blinder Alarm*; if *false* and *alarm* are rendered separately, the translation is likely to be *falscher Alarm*, which is not suitable for referring to the meaning of *false alarm*. On the other hand, the same phrase *false alarm* cannot be regarded as a translation unit when it is examined from the English-French perspective. Translators are likely to succeed in translating *false alarm* into the French equivalent *alarme fausse* by rendering *false* and *alarm* separately. Therefore, the translation unit sizes are different in these cases. In the English-German comparison, the translation unit was *false alarm*, which was a phrase. On the other hand, in the English-French comparison, the translation units were *false* and *alarm*, which were single words. The different language pairs caused different translation unit sizes.

An empirical study by Goldman-Eisler (1972) demonstrates this relation between language pairs and translation unit size effectively. Her study is unusual: in order to identify translation units, Goldman-Eisler (1972) used spoken translations, i.e. simultaneous interpretations. Interpreters usually cannot start the translation right after they recognise the first word of the original text. They have to wait for a few more words coming after the first word, and then translate the chunk. This unit that interpreters have to wait for is a translation unit, in the sense that it is an inseparable unit for translating (Goldman-Eisler terms it an 'ear-voice span', i.e. a 'segment which the interpreter needs to monitor before he can start encoding' (1972)). The study shows that interpreters tend to wait for verbs before starting translation. The translation unit in the French-English comparison (translated from French to English), therefore, tends to be the clause, consisting of a subject and a verb. On the other hand, the unit in the

German-English comparison (translated from German to English) tends to be the longer clause consisting of a subject, a verb, and an object. This is due to the position of verbs in German: the 'verb frequently follows the object' (Goldman-Eisler, 1972), which makes the translation units larger than those from the French-English comparison.

The two theorists above make themselves very clear that translation unit size is related to the language pair being considered. So, a question which arises here is: which language pair is associated with a smaller or larger translation unit? Vinay and Darbelnet (1995) point out that, (a) this is a matter of taxonomical and historical background of a given language pair; and more specifically, that (b) translation units tend to be single words when the two languages are from the same family and culture. Since there is little multi-linguistic evidence in Vinay and Darbelnet's study (1995), the literature was examined to see if this view was tenable or not. Figure 2.4 shows all the empirical studies on specific language pairs, relating to the English language; the studies are sorted by the language pairs for the sake of the discussion.

Figure 2.4. Translation unit size and language pairs

	Language pairs	mo.	w.	ph.	cl.	sen.	par.	tex.
Goldman-Eisler (1972)	English-German				■			
Lörscher (1991)	English-German			■	■			
Goldman-Eisler (1972)	English-French				■			
Teubert (2001)	French-English		■	■				
Alves and Gonçalves (2003)	English-Portuguese			■				
Barbosa and Neiva (2003)	English-Portuguese		■	■	■			
Toury (1986)	English-Hebrew			■				

Figure 2.4 contains four language pairs: English-German; English-French; English-Portuguese; and English-Hebrew. No translation units were found at the levels of morpheme, paragraph, or text in the empirical studies. All the translation units were identified only at the levels of word, phrase, clause, and sentence, no matter which language pair the linguists used. Moreover, the sizes of phrase and clause are dominantly seen in Figure 2.4; the translation units at the levels of single word and sentence are rather rare.

First, the language pairs which are not from the same language family in Figure 2.4 are English-French, English-Portuguese, and English-Hebrew (see the simplified Indo-European family tree in Figure 2.5). The three relevant studies, i.e. Goldman-Eisler (1972), Alves and Gonçalves (2003), and Toury (1986), show that their translation units were not at the single word level, as Vinay and Darbelnet (1995) maintain in their point (b) above. Second, the language pair which shares some historical background is English-French. This is due to the 'periods of bilingualism' in English history, allowing the language to share some vocabulary with French (Vinay and Darbelnet, 1995). The case study carried out by Teubert (2001) proves that this is the case; his translation units were identified at the level of single words; point (a) above of Vinay and Darbelnet (1995) was supported.

Figure 2.5. Language family: Indo-European languages (abridged from Trask, 2000)
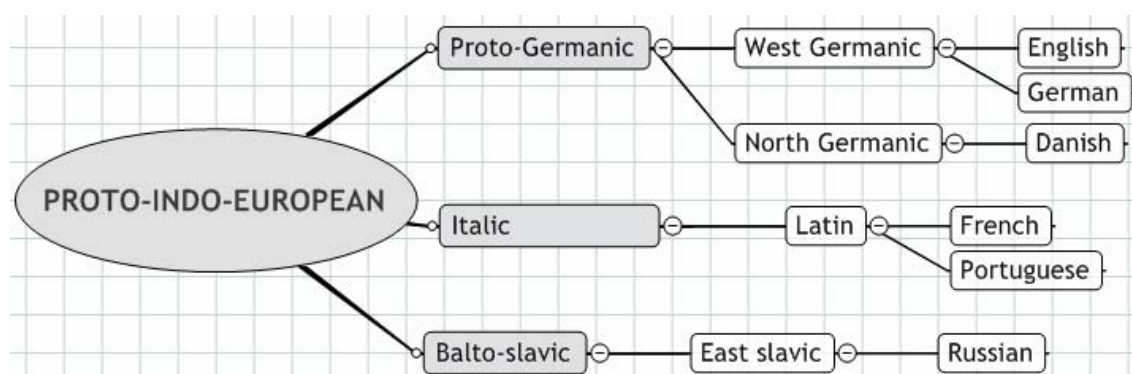
However, there are two researches which disagree with point (b) of Vinay and Darbelnet (1995): Goldman-Eisler (1972) and Lörscher (1991). They examined data from an English-German comparison. Taxonomically speaking, this is the most closely related pair in Figure 2.4; both are from the same language group, the Germanic language group (Durrell, 2006). However, their identified translation units were not single words. Perhaps, this claim about single words (point (b) above; Vinay and Darbelnet, 1995) is rather strict, in the sense that they specify the translation unit size. One might argue that it would be more tenable if one modifies the claim slightly; for example, we might claim, more broadly, that the translation units tend to be *smaller* when the two languages are from the same family and culture. In fact, Koller takes this relative view; 'a translation between unrelated languages will usually involve larger units than if SL and TL [= target language] are closely related' (cited and translated in Shuttleworth and Cowie, 1997).

This relative view accords better with Figure 2.4. The empirical studies suggest: (a) the translation units in the English-French pair are relatively the smallest; (b) the translation

units in the English-Portuguese pair are relatively the largest; and, (c) the translation

units in the English-German pair are larger than the ones of the English-French pair and

smaller than the ones of the English-Portuguese. All these findings concur with the

relative view. There is one study for which this view is not quite appropriate.

Translation units in the English-Hebrew perspective (Toury, 1986) were found to be at

the same level as translation units in the English-German perspective (Lörscher, 1991):

phrases and clauses. From the language taxonomical point of view, English-Hebrew is

the most distant language pair in Figure 2.4; Hebrew is not an Indo-European language,

but a Semitic language (Hackett, 2006). On the other hand, English-German, is the

closest language pair; as discussed above, both are from the German language group in

the Indo-European language family (Ross, 2006). These two language pairs have no

striking difference regarding the sizes of translation units. Further empirical studies on

this language pair would be of great interest.

Therefore, the literature validates the claim that translation unit size is associated with

the language pair used for the identification. This association is likely to be a matter of

taxonomical and historical background of a given language pair. If the two languages

are closely related, the translation units are expected to be smaller. If the languages are

not closely related, then translation units are expected to be larger. By saying 'smaller'

or 'larger', the translation unit sizes usually range from the single word to the sentence,

as far as the empirical evidence is concerned.

2.4.3. Text types

Some theorists argue that translation unit size is decided by the type of text. Sager (1994) is one of these, maintaining that 'size … depends on the nature of the document'. There are many ways to classify the nature of a document. Tancock (1958), for example, distinguishes between texts that have 'overtones' and 'continuity of style' from the ones that have 'the least flavour'. Literature belongs to the former type; according to Tancock (1958), their translation units tend to be larger (i.e. longer sentences or short paragraphs). On the other hand, business letters or reports of scientific experiments belong to the latter type; their translation units tend to be smaller (i.e. single words). Other linguists who have written about the relationships between text type and translation unit size are Bassnett-McGuire (1980) and Barkhudarov (1993); in particular, the translation unit in poems is their concern. According to Bassnett-McGuire (1980), translation units in this text type are likely to be the line, verse (i.e. a 'small number of metrical lines' (**Oxford English Dictionary**, 2000)), and stanzas (i.e. a 'group of lines of verse' (**Oxford English Dictionary**, 2000)); similarly, Barkhudarov (1993) maintains that translation units in poetry are likely to be the whole text. These studies suggest that a significant size difference exists between the translation unit in technical documents and more creative writing such as poetry.

Other text types have been examined in several empirical studies (Figure 2.6). The research by Teubert (2001) was carried out with a corpus of newspaper texts; translation units found were at the level of words and phrases [personal communication]. Similarly, Alves and Gonçalves (2003) also examined translation units in news text from the news

magazine *Newsweek*. Translation units were found at the level of phrases; this overlaps with the study by Teubert (2001). Vinay and Darbelnet's research (1995) was based on an academic article. Translation units found in this analysis were mainly at the levels of phrases and clauses (1995). Both text types (i.e. news text and academic articles) make translation units smaller than those found in literature. Tancock's 'flavour' theory (1958) is compatible with the text types used in these studies; the less 'flavour', style or originality a text has, the smaller its translation units are.

Figure 2.6. Translation unit size and text type (i)

	Text types	mo.	w.	ph.	cl.	sen.	par.	tex.
Alves and Gonçalves (2003)	news magazine			▮				
Teubert (2001)	newspaper texts		▮	▮				
Toury (1986)	TV scripts			▮	▮			
Vinay and Darbelnet (1995)	academic article			▮	▮			

Translation unit size in different text types is summarised in Figure 2.7. Translation units in technical documents are associated with a smaller size of translation unit, whilst those in prose texts are associated with relatively larger sizes of translation unit. Translation unit size in news text and academic articles lies between the two. Figure 2.7 helps explains why Barkhudarov (1993) maintains that such a wide range of translation unit sizes are possible: the phoneme, morpheme, word, phrase, clause, sentence, and the whole text (1993). Considering that he discusses this size issue for the sake of translation pedagogy, his diversity is valid; he includes all the possibilities of translation unit size for novices who may have to render any text type, i.e. technical documents, news texts, academic articles, or literature. Regarding the levels of phoneme and morpheme, Barkhudarov (1993) points out that translation units are rarely identified at these levels. The phoneme as a translation unit happens only when the names of people

and places are translated; the morpheme as a translation unit occurs only when languages are morphologically similar (Barkhudarov, 1993). Since his main concern is translation pedagogy, it is reasonable that his view on translation unit size is exceptionally diverse, in order to cover all the possibilities.

Figure 2.7. Translation unit size and text types (ii)

morpheme	word	phrase	clause	sentence	paragraph	text

t.d.
news text
academic articles
literature
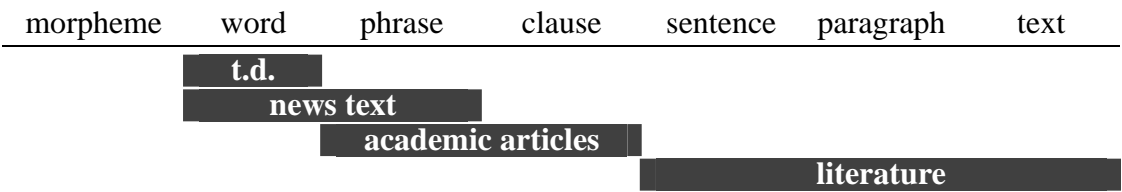
2.4.4. Translators

The translator can be an important variable associated with translation unit size, as well. Several studies have been carried out in order to find out the differences in translation unit sizes between experienced and inexperienced translators (Figure 2.8).

Figure 2.8. Translation unit size and translator experience

Studies	Translators	mo.	w.	ph.	cl.	sen.	par.	tex.
Toury (1986)	Inexperienced							
Barbosa and Neiva (2003)	Inexperienced							
Lörscher (1991)	Inexperienced							
Alves and Gonçalves (2003)	Inexperienced							
Toury (1986)	Experienced							
Barbosa and Neiva (2003)	Experienced							
Lörscher (1991)	Experienced							
Vinay and Darbelnet (1995)	Experienced							
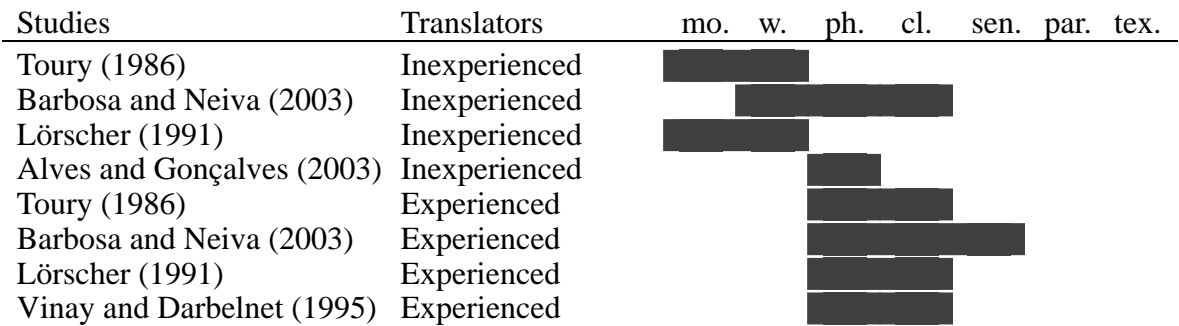
For example, Toury (1986) used his data of English-Hebrew and compared translated texts produced by the two different types of translators. The translation units that the

38

inexperienced translators used were smaller (i.e. morphemes and words) than the units that the experienced used (i.e. phrases and clauses). Based on observations on English-German, Lörscher (1991) agrees with this view, arguing that non-professional translators use smaller translation units (i.e. single word or smaller than that) and professional ones use larger units (i.e. phrase, or larger than that). Barbosa and Neiva (2003) carried out an experiment, from the English-Portuguese perspective, examining the difference between inexperienced and experienced translators using the think-aloud method (see 2.2 concerning this method). A translation unit was segmented by 'a break in the translation flow' in the recording, and the units formed by inexperienced translators were found at the level of single words, phrases, and clauses; while the units by experienced translators were found at the levels of phrases, clauses, and sentences (2003), which coincides with both Toury's (1986) and Lörscher's (1991) views. These empirical case studies are summarised in Figure 2.9.

Figure 2.9. Translation unit size and translators

| morpheme | word | phrase | clause | sentence | paragraph | text |
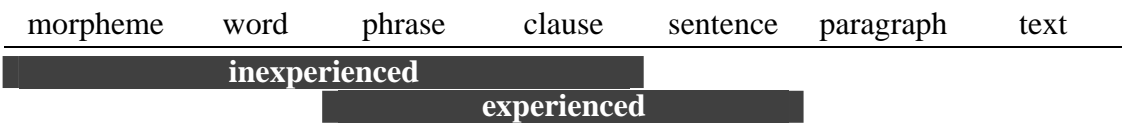
Some experiments were carried out using data rendered by only inexperienced translators. The analysis by Alves and Gonçalves (2003) is one of these. Their results showed that translation units used by inexperienced translators were at the level of phrases; this accords with Figure 2.9. On the other hand, some experiments were carried out using data rendered only by experienced translators. Vinay and Darbelnet's research

39

(1995), for example, shows that the translation units of experienced translators were dominantly phrases and clauses; none of the smaller units (i.e. morpheme and word) were recognized as dominant translation units, again supporting Figure 2.9. Therefore, the literature clearly shows that translation unit size is closely related to how experienced the translators are. Smallness is associated with inexperienced translators (i.e. morpheme, word, phrase, and clause), while the translation units that experienced translators use are relatively larger (i.e. phrases, clauses, and sentences).

2.4.5.   Implications

So far, I have compared the literature on translation unit size and found that there are four associated variables. Figure 2.10 shows the comparative relations between all the variables and translation unit size. The arrow represents a continuation of the major sizes of translation unit in the following order: morpheme; word; phrase; clause; sentence; paragraph; whole text.

Figure 2.10. Translation unit size and associated variables

Smaller translation unit   Larger translation unit

←——————————————————————————————→

Definition: lexical   Definition: cognitive
Language pair: closely related   Language pair: unrelated
Text type: non-literary   Text type: literary
Translator: inexperienced   Translator: experienced
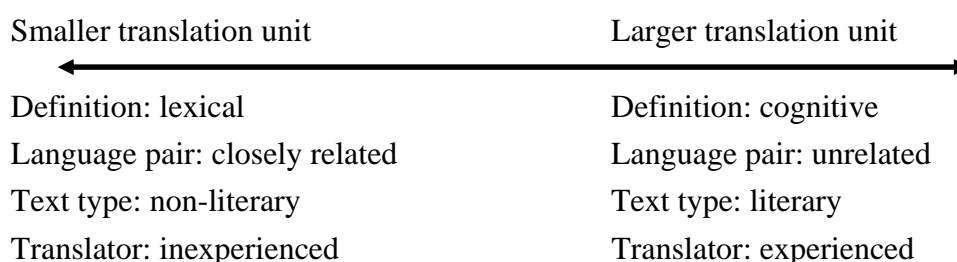
First, whether translation units are defined as lexical or cognitive affects how small or large the units are. Lexical translation units tend to be smaller; cognitive ones tend to be larger. Second, whether the two languages are closely related or unrelated affects how

small or large a translation unit is. The former unit tends to be smaller; the latter unit between unrelated languages tends to be larger. Third, whether the language data consists of non-literary (e.g. technical documents, news text, academic articles) or literary texts (e.g. poem and prose) affects the size of the translation unit. The unit tends to be smaller in the former texts but larger in the latter text types. Lastly, whether language data are rendered by inexperienced or experienced translators affects the size as well. Inexperienced translators tend to use smaller translation units and experienced translators larger ones.

Figure 2.10 shows the intricate relationship between the four variables and translation unit sizes. It is these variables that cause the wide diversity of views on translation unit size held among linguists (Figure 2.1). The diversity occurs because each theorist focuses on different variables. Some theorists are interested in lexical translation units and identify their sizes (Toury, 1986; Vinay and Darbelnet, 1995; Teubert, 2001); some are interested in cognitive units and identify their sizes (Nida and Taber, 1969; Lörscher, 1991; Alves and Gonçalves, 2003; Livbjerg and Mees, 2003). It is not surprising that their results are different. Similarly, some argue the size issue by examining data from closely related languages (e.g. English-German (Lörscher, 1991), English-French (Vinay and Darbelnet, 1995), English-French-German (Goldman-Eisler, 1972; Teubert, 2001), English-Russian (Barkhudarov, 1993)); on the other hand, some identify their translation units in data of unrelated languages (e.g. English-Portuguese (Alves and Gonçalves, 2003; Barbosa and Neiva, 2003) and English-Hebrew (Toury, 1986)). Their claims on translation unit size are not necessarily identical at all.

41

Also, discussion on translation unit size easily ends up with disagreement if each theorist focuses on different text types. It is unlikely that an agreement will be reached between theorists who examine translation units in news texts (Teubert, 2001; Alves and Gonçalves, 2003) and in literature (Tancock, 1958; Bassnett-McGuire, 1980). Moreover, some theorists analysed translation units in data consisting of several different text types. For example, Lörscher's data (1991) includes both academic articles and news texts from *The Times* newspaper. The result regarding translation unit size is likely to differ from an analysis which focuses only on either of these text types. Similarly, some experiments were carried out using translation data from inexperienced translators (Alves and Gonçalves, 2003); and some were by experienced translators (Vinay and Darbelnet, 1995). Their views on translation unit size are likely to differ, as well.

What this investigation clearly shows is that the four variables need to be stated clearly if one wants to argue over translation unit size. It is necessary to choose one definition of translation unit, one language pair, one text type, and one level of translators experience, otherwise, the analysis ends up with results that cover a large range of sizes. For example, if one examines translation unit size in mixed text types, the result is likely to tell a wide range of translation unit size (e.g. Tancock (1958), Barkhudarov (1993), Newmark (1988b), shown in Figure 2.1 and Figure 2.11). Since it covers almost every possible translation unit size, such a finding is rather disappointing in the sense that the analyses fail to pin down a few particular sizes of translation unit. It also makes it very hard to conclude what the precise effects of changing one variable in isolation might be, and to determine the linguistic bases of such changes.

Figure 2.11. Selected studies and variables

	Definitions	Languages	Text types	Translators
Alves and Gonçalves (2003)	Cognitive	En-Por	News magazine	Inexperienced
Barbosa and Neiva (2003)	Cognitive	En-Por	-	In/experienced
Barkhudarov (1993)	Lexical	En-Rus	Mixed	Experienced
Bassnett-McGuire (1980)	-	-	Poem, prose	Experienced
Bell (1991)	-	-	-	Experienced
Goldman-Eisler (1972)	Cognitive	En-Fr	Mixed	Experienced
Livbjerg and Mees (2003)	Cognitive	Dan-En	Newspaper	Inexperienced
Lörscher (1991)	Cognitive	En-Ger	Mixed	Inexperienced
Newmark (1988b)	Lexical	-	Mixed	Experienced
Nida and Taber (1969)	Cognitive	-	-	Experienced
Tancock (1958)	-	-	Mixed	Experienced
Teubert (2001)	Lexical	Fr-En	Newspaper	Experienced
Toury (1986)	Lexical	Heb-En	TV script	In/experienced
Vinay and Darbelnet (1995)	Lexical	En-Fr	Academic text	Experienced

The four variables could be very useful for contrastive studies. If one compared two studies which focused on translation units from the same definition, examined in the same text types, rendered by the same level of translators, but in different language pairs, one could identify how similar or different the languages are, and codify these differences between the pairs. One of the closest matches in Figure 2.11 is between Livbjerg and Mees (2003) and Teubert (2001). They both examined their translation units in newspaper texts. However, the former defined the unit from the cognitive point of view; while, the latter defined the unit from lexical point of view. Also, the translators' qualities were different between the two studies. If their variables were the same except for the language pair, they could have drawn conclusions on how different or similar Danish and French are from the point of view of English; that is, what features of a language pair determine, or influence, translation unit size.

2.5. Conclusion

This chapter has demonstrated that there is little agreement on translation unit size in the linguistics literature. In order to clarify this issue, this chapter was dedicated to reviewing the relevant case studies. What this comprehensive review has demonstrated is that there are four important variables associated with translation unit size, and that they are at the root of the disagreements. For a productive discussion of translation unit size, it is indispensable that studies clearly state with which variables they are concerned.

3. Data and methodology

The aim of this thesis is to investigate translation unit size in the English-Japanese comparison. Before embarking upon empirical analysis, there are some methodological issues that need to be discussed clearly. First, the research question needs to be revisited and refined, along with the four variables discussed in Chapter 2: cognitive or lexical sense; language pair; text type; and, translator's quality. Second, the corpus data used for this study will be validated. In particular, both advantages and limitations will be clearly recognised. Third, the methods of translation unit extraction will be clarified. The existing methods are discussed in detail, followed by a discussion on why this study did not utilise any of these. Finally, an alternative method for extracting translation units will be introduced, with its advantages and limitations.

3.1. Research question

As stated clearly in Chapter 1, this thesis aims to find out how large or small the translation units are in the English-Japanese comparison. This question, however, now appears to be too broad to bring productive results to the discussion of translation unit size (discussed in 2.4.5). Therefore, this research question needs to be re-shaped and specified along four operational variables. First, this study defines a translation unit as a lexical unit. This is because this study is interested in translation pairs that have been established over time in the discourse, not ones which were invented by a few translators in their minds (see the relevant discussion in 2.4.1). Also, the current study attempts to exploit parallel corpora for identifying translation units. Due to the use of a

corpus, lexical translation units, rather than cognitive translation units, are more promising to focus on, since they should be visible on the text surface. Second, this study focuses on the English-Japanese comparison (translating from English to Japanese). The choice of this language pair is motivated by my own personal interest and expertise; I myself am a bilingual speaker of the two languages, and have carried out cross-linguistic research on English and Japanese (Kondo, 2000; 2006). The parallel corpus used in this study is the Alignment of Reuters Corpora (ARC) at the National Institute of Standards and Technology (NIST) in Japan; the details will be discussed in 3.2. This choice of corpus determined the last two variables. The text type concerned in this study is newswire articles; the translators are experienced.

The modified research question, therefore, is: how large or small are translation units in English-Japanese newswire texts, rendered by experienced translators, from a lexical point of view? The relevant literature suggests several possible translation unit sizes: morpheme; single word; phrase; clause; sentence; paragraph; and, whole text (Chapter 2). This thesis will present empirical research and investigate which one of these possible translation unit sizes is commonly found in the corpus. Also, whether the translation units are restricted to only one size, or are a selection of these possible lengths, is another topic of investigation.

So far, I have used the term 'translation unit' as a hyponym of cognitive and lexical translation units, i.e. a term covering both cognitive and lexical translation units. I have made myself clear regarding whether a unit is a 'cognitive translation unit' or a 'lexical translation unit', where this classification was needed in the discussion. From this point

forward, since all the translation units to which I will refer are lexical translation units only, I use the term 'translation unit' to refer to the lexical unit only.

3.2.　Corpus data

The ARC was the parallel corpus used for identifying translation units in this study (Table 3.1). The main benefit of this corpus was its directionality (i.e. whether a corpus contains texts translated from language A to language B and/or from language B to language A (Altenberg and Granger, 2002)). The ARC consists of texts written in English and their Japanese translations (Ian Soboroff, personal communication).

Directionality is the most crucial feature when one chooses a corpus for identifying translation units. This is simply because the translation unit is a unit in the original text (discussed in 2.3). If one wants to identify English translation units, one has to use parallel corpora consisting of English original texts and their Japanese translations; while, if one wants to identify Japanese translation units, one has to use parallel corpora consisting of Japanese original texts and their English translations. This study belongs to the former group; English translation units and their sizes are concerned. The ARC is, at present, the largest parallel corpus for this direction available in the public domain (Table 3.1): 1.8 million words for English texts and 2.5 million morphemes for Japanese texts (counted by ParaConc (Barlow, 1995), see the details of ParaConc in 3.4.6). All

the article texts contained within the corpus were published in 1996 and 1997[4] (Utiyama and Isahara, 2003).

Table 3.1. Large parallel corpora of English and Japanese

	ARC	EJTAD[5]	JENAAD[6]
Directionality	Eng. ⇨ Jap.	Jap. ⇔ Eng.	Jap. ⇨ Eng.
Size (English)	1.8 million words	0.6 million words	4.9 million words
Size (Japanese)	2.5 million morphemes	1.0 million morphemes	6.1 million morphemes
Content	Reuters Newswire	Fiction and non-fiction	Broadsheet newspaper

The EJTAD (English-Japanese Translation Alignment Data) was another possible corpus for this research; however, it was not used. As shown in Table 3.1, the directionality was not appropriate for this study. It is a collection of bilingual texts of three different translation directions: (a) English originals (e.g. *A Christmas Carol* by Charles Dickens) and their Japanese translations; (b) Japanese originals (e.g. *Botchan* 'Master Darling' by Soseki Natsume) and their English translations; and (c) Japanese and English translations of original texts written in other languages (e.g. *The Little Match Girl* by Hans Christian Andersen). Considering that this study is concerned with English translation units, only the subcorpus of (a) is suitable for this research. It was possible to extract this subcorpus out of the whole of EJTAD; however, this was not worthwhile, since the data would be much smaller than the ARC. The identification method in this study (see 3.4) requires quantitative data; therefore, the EJTAD was not preferable.

---

[4] http://trec.nist.gov/data/reuters/reuters.html (accessed on 19 November 2009)
[5] http://mastarpj.nict.go.jp/~mutiyama/align/index.html (accessed on 19 November 2009)
[6] http://mastarpj.nict.go.jp/~mutiyama/jea/index.html (accessed on 19 November 2009)

The JENAAD (Japanese-English News Article Alignment Data) was another possible corpus for this study; however, it was not used either. Due to the data directionality (i.e. from Japanese to English, shown in Table 3.1), the use of this corpus would have required a change in my research question; Japanese translation units would have to have been investigated, instead of English ones. When the size of the JENAAD was considered, this option was actually attractive; the JENAAD is a much larger corpus than the ARC (Table 3.1). However, the research question presented above remained. This was because of my on-going interest in cross-linguistic comparisons of translation unit size; how does translation unit size in one language pair differ from the one in another language pair?

For this cross-linguistic aim, it is more promising to focus on English translation units rather than Japanese ones. There are many parallel corpora consisting of English originals and their translations in other languages; 'English is probably the most widely translated language in the world' (Baker, 1992). This gives ample possibility to examine English translation units, and their sizes, in many language pairs, and compare them. On the other hand, there are far fewer parallel corpora consisting of Japanese originals and their translations in other languages. This provides much less opportunity for examining Japanese translation units and their sizes in several language pairs, and comparing them. This thesis, therefore, focuses on English translation unit size, since this is more beneficial towards cross-linguistic projects on translation unit size.

The term 'parallel corpora' has been used so far to refer to corpora consisting of 'source texts in language A and translations in language B' and/or 'source texts in language B and their translations in language A' (Olohan, 2004). This usage of the term is

supported by many linguists (Teubert, 1996; Barlow, 2000; Kenny, 2001; Hunston, 2002; Danielsson, 2003; Olohan, 2004; Teubert, 2004b; Bernardini, 2007). However, it is worth noting that there are some linguists who use the term 'parallel corpora' differently. For example, Johansson (2007) uses the term in a broader sense, referring not only to the parallel corpora *sensu* Olohan (2004) but also to the comparable corpora, i.e. corpora 'made of originals in language A and comparable translations into language A' (Bernardini, 2007)). Hartmann and James (1998) also use this broad definition of the term 'parallel corpora'. In order to avoid confusion, it is essential to state clearly that this thesis uses the term 'parallel corpora' *sensu* Olohan (2004) only.

3.2.1. Advantages

The directionality and size of the ARC texts are not the only factors for which the ARC was chosen for this study. Two other variables, text type and translator's quality, also meant that the ARC was the preferred corpus. First, the ARC consists of one single text type: newswire texts. As discussed in 2.4.3, translation unit size is related to the text type the corpus contains. The ARC allows one to identify specific characteristics of translation units in newswire texts. This advantage becomes clearer if one compares the ARC with the EJTAD. The latter corpus contains texts of various text types (Table 3.1), such as literary works, e.g. *Through the Looking Glass: And What Alice Found There* by Lewis Carroll, and academic articles, e.g. *Discourse on the method of rightly conducting the reason, and seeking truth in the sciences* by Rene Descartes. Translation units in literary texts are likely to be larger than the ones in academic texts (2.4.3);

therefore, it is hard to interpret the results if one examines translation unit size in such a mixed corpus.

Moreover, the EJTAD contains another text type as well: spoken texts such as academic lectures, e.g. *RMS lecture at KTH* by Richard M. Stallman, and political inaugural speeches, e.g. *Inaugural Address* by John F. Kennedy. In the systematic review of the relevant literature, there is no concrete evidence proving that translation unit size in written text is smaller or larger than that in spoken text. However, in the case that this might be another variable associated with size, a corpus which consists of both written and spoken is not ideal for identifying translation unit size. The EJTAD, therefore, is not suitable, unless the data are controlled by creating a subcorpus of a single text type; the ARC does not need such manipulation.

Second, the ARC was the preferred corpus due to the translator's quality. The data in the ARC were rendered by professional translators only. All the translation in the ARC must be carried out by translators whose competence meets the requirements to work for a highly regarded newswire company such as Reuters. As discussed in 2.4.4, translation unit size is associated with whether corpus data were rendered by experienced or inexperienced translators; the translation unit size is larger in texts of experienced translators and the translation unit size is smaller in texts of inexperienced translators. Therefore, it is better that the study of translation unit size is examined in corpora which contain either translation texts by experienced or inexperienced translators. The ARC was desirable for this aspect as well.

51

There was one technical aspect that made the ARC stand out as well. The data in the ARC are already aligned at the sentence level. This means that a sentence in the source text and its corresponding sentence in the target text are paired up and ready for processing in a parallel concordancer, which is an essential tool to analyse large parallel texts. It is true that the data can be aligned by computer programmes automatically, if the data one wants to use is not aligned. However, such automatic alignment needs manual checking, which can be time-consuming, especially if the data are translations between unrelated languages (Olohan, 2004). Translation texts between English and Japanese are likely to belong to this case. According to Utiyama and Isahara (2003), one-to-one correspondence between English and Japanese sentences in news articles is exceptional; some sentences are omitted or added in translation texts in order to 'fill cultural gaps'. Correct automated alignment is, therefore, rarely guaranteed. One can easily lose large amounts of time, checking and fixing the wrong alignments. Thus, data that are already aligned are a precious resource; they allow one to start analysis immediately.

Diachronic consistency of translation equivalence is another benefit of the ARC. The ARC contains data published in 1996 and 1997 only. Therefore, the translation units identified in this corpus are contemporary ones as well as their corresponding equivalents. On the other hand, if one uses the EJTAD, the translation pairs (i.e. translation units and their equivalents) tend to have diachronic differences. For example, the EJTAD has texts of *The Black Cat* by Edgar Allan Poe: English original text published in 1843 and the Japanese translated text published in 1951 (according to the website of Aozora bunko, the Japanese equivalent of Project Gutenberg, from which the

texts were taken[7] for the EJTAD). The outcomes of such analysis are: translation units of the mid-1800s and their equivalents of the mid-1900s. Seeing that this diachronic inconsistency might influence the sizes of translation units, the ARC was the preferred corpus for this study.

3.2.2. Limitations

One limitation is that the ARC is a collection of specific sentence pairs. The ARC is composed of two sets of alignment data: one-to-one and one-to-many sentence alignment data. The former consists of sentence pairs in which one English sentence, e.g. *A copy of the speech was released in Sydney*, corresponds to one Japanese sentence, e.g. *Enzetsu genkoo wa shidonii de happyoo sa re ta mono* 'the speech copy is the one released in Sydney'. On the other hand, the latter consists of sentence pairs in which one English sentence, e.g. *Abnormal weather in South Africa has flooded a desert, blocked roads with snow and delayed harvesting of maize crops, government and agriculture officials said on Tuesday*, corresponds to two or more Japanese sentences, e.g. *Minamiafurika de wa ijoo kishoo ni yori, sabaku de koozui ga hassee shi, yuki de dooro ga shadan sa re, toomorokoshi no shuukaku ga chien shi te iru. Seefu oyobi noogyoo kankee sha ga akiraka ni shi ta* 'In South Africa, due to abnormal weather, desert is flooded, road is blocked with snow, harvesting of maize crops is delayed. Government and agriculture officials said'. This means that the ARC does not contain the following sentence alignment pairs: one-to-zero, i.e. an English original sentence and no correspondence in Japanese, due to omission; zero-to-one, i.e. no English original

---

[7] http://www.aozora.gr.jp/cards/000094/card530.html

sentence and one added Japanese sentence; zero-to-many, i.e. no English original sentence and many added Japanese sentences; and many-to-many, i.e. many English sentences and their many corresponding Japanese sentences. Moreover, if some long sentences are divided up, or some short sentences are combined into one new sentence, such data are not included in the ARC. In other words, translation units identified in the ARC are the ones of special sentence alignment data.

Another limitation is related to the first point. Since it is a collection of specific sentence pairs, the ARC does not allow the user to trace back to the original texts. Supposing that an English article is composed of five sentences, E1, E2, E3, E4, and E5, and its Japanese corresponding article is composed by four sentences, J1, J2, J3, and J4: E1 corresponds to J1 and J2; E2 was omitted in Japanese; E3 corresponds to J3; and, E4 and E5 correspond to J4. Then, the ARC contains only two sentence pairs from this article pair: (a) E1 and J1 and J2, i.e. a one-to-many sentence alignment, and (b) E3 and J3, i.e. a one-to-one sentence alignment. The other sentences, E2, E4, E5, and J4, are discarded and are not seen in the ARC. If one examines E1 and wants to see what E1 is followed by in the original text, the ARC does not allow one to do so; the corpus does not contain E2. This means that investigation beyond the sentence level is not possible in the ARC. Whether translation units are identified at the level of paragraph and text cannot be examined using this corpus.

The other limitation is the size of the ARC: 1.8 million words for English original texts and its Japanese translations of 2.5 million morphemes. This might not be large enough to identify a considerable amount of translation units. Danielsson (2001) used the

54

Swedish-English parallel novel corpus containing 0.5 million words per language and extracted translation units. Since she managed to extract only a few translation units from the corpus, Danielsson (2001) points out that much larger corpora are preferable for identifying translation units. The ARC is about 4 times larger than her parallel corpus, if the English texts are compared. I will come back to this point after the analyses and discuss whether a parallel corpus of approximately 4 million words in total is enough for studies identifying translation units or not in 7.5.3.

One might add one more limitation: mistranslations. It is true that all the translated data in the ARC are not guaranteed to be good translations; it is possible that it consists of some bad translations. Therefore, the ARC might yield some 'bad' translation units. However, this is unlikely to happen in this study. This is because the current study set up a filter to eliminate unwelcome translation data. The filter was designed (in 3.4.2), based on the nature of bad translations; wrong translations do not 'tend to be repeated' (Teubert, 2004b). Therefore, this was not regarded as a major limitation.

The ARC, thus, has three limitations. First, the ARC is a collection of paired English and Japanese sentences of only one-to-one and one-to-many relations, which gives a limited, special platform for examining translation unit sizes. Second, the ARC allows the user to investigate within the sentence; context investigation beyond the sentence level is impossible to achieve. Finally, the size of the ARC is potentially smaller than the ideal. Despite such limitations, the ARC is still the most beneficial resource for identifying translation units and their sizes for this thesis, due to the unrivalled advantages discussed in 3.2.1.

3.3. Existing methods

The methods for identifying (lexical) translation units were briefly introduced in 2.3: the degree of cohesion (Vinay and Darbelnet, 1995); the no leftover principle (Toury, 1986; 1995); units of meaning (Danielsson, 2001); and the monosemous principle (Teubert, 2004b). However, none of these was used in this thesis in a strict sense. This subsection will discuss these existing methods with their drawbacks. In particular, the reasons why the methods were not implemented in this study will be clarified.

3.3.1. Degree of cohesion

The degree of cohesion method was suggested by Vinay and Darbelnet (1995). This method relies on one attribute of a translation unit; the translation unit has 'cohesion between the elements' within it (1995). Therefore, the identification focuses on judging whether an item has a close relationship with other item(s) or not. There are apparently

three degrees of cohesion strength between elements. First, if an item is a highly coherent unit, such as idioms, then it is likely to be a translation unit. Their example was the French idiom *avoir lieu*; it is regarded as a translation unit since it has strong cohesion between *avoir* and *lieu* from the translation point of view. *Avoir lieu* is an inseparable unit in a sense that it is rendered as a unit into its English equivalent, *to take place* (1995). Second, if an item has weak cohesion between elements, such as most fixed phrases, then it is likely to be a translation unit as well. The French phrase *un hiver rigoureux* was their example for this type; it was regarded as a translation unit, since it has some cohesion between *un*, *hiver*, and *rigoureux*, but this is of a weaker type. *Un hiver rigoureux* is a relatively inseparable unit, although not idiomatic, and is rendered as a unit into its English equivalent *a severe winter* (1995). Lastly, if a given item does not have any cohesion between elements, then it is likely to be regarded as a single word translation unit. For example, *commençant* in the sentence *elle aura duré environ 30 ans, commençant au début des années 60* has no cohesion with its surrounding words; *commençant* is alone rendered into *starting* in *it will have lasted some 30 years, starting in the early 60s*. Therefore, *commençant* is regarded as a single word translation unit.

However, judging the degree of cohesion is not often straightforward, of which Vinay and Darbelnet (1995) are themselves aware. In particular, it is hard to detect units of weak cohesion and no cohesion. An example sentence from the ARC demonstrates this difficulty: *By and large, moderate growth is still the most likely scenario so far* (rendering into Japanese, *Zentai teki ni mi te, yuruyaka na keezai seechoo toiu no ga, ima no tokoro mottomo kanoosei no takai shinario da* 'Observing generally, moderate

growth is the most likely scenario so far'). Two translation units are easily recognized: *by and large* and *so far*. They are regarded as highly coherent units, since (a) *by and large* acted as an inseparable unit, i.e. *by*, *and*, and *large* were not separately rendered into Japanese; in other words, there are no words denoting the meanings of *by*, *and*, and *large* separately in the translation, and (b) similarly, *so far* acted as an inseparable unit, i.e. *so* and *far* were not separately rendered, either.

On the other hand, the rest of the sentence (i.e. *moderate growth is still the most likely scenario*) is hard to analyze. *Moderate growth*, for example, does not exhibit cohesion as strongly as *by and large* and *so far* do. Then, would you say that it is a weak coherent unit or a unit without coherence? The Japanese counterpart of *moderate growth*, *yuruyaka na keezai seechoo*, does not help us to answer to this question; one cannot decide whether the words *moderate growth* were rendered together or separately. Some may claim that *moderate growth* is a translation unit; it has weak cohesion between *moderate* and *growth*, and was rendered as a whole into *yuruyaka na keezai seechoo*. Another could claim that *moderate growth* has no cohesion between the words, and was rendered separately, i.e. *moderate* corresponding to *yuruyaka na* and *growth* corresponding to *keezai seechoo*; therefore, *moderate* and *growth* would be two single word translation units. The judgment tends to be biased, relying on the researcher's interpretation of what cohesion is, and their knowledge of the language pair under investigation.

The crucial problem of Vinay and Darbelnet's identification method (1995), thus, is that they do not clearly explain what cohesion actually refers to. For the highly coherent

units, they imply that cohesion is defined semantically and syntactically; '[t]he unity of meaning is very clear and is often marked by a syntactic characteristic' (1995). Both *by and large* and *so far* belong to this group. However, when it comes to units of weak cohesion and no cohesion, Vinay and Darbelnet list examples of such units without much explanation. Without the clear guideline, this identification is not helpful to linguistis who want to use this identification method as a reseach tool. The irreproducibility is fatal.

One might wonder if Halliday and Hasan's (1976) definition of cohesion – 'Cohesion occurs where the INTERPRETATION of some element in the discourse is dependent on that of another [original emphasis]' – can be an alternative guideline, and make Vinay and Darbelnet's identification method feasible. Unfortunately, this is not the case. What Halliday and Hasan define as cohesion differs from the definition of Vinay and Darbelnet. The former pair refer to 'text-forming relations' (Halliday and Hasan, 1976); for example, cohesion lies between the two items, *the admiral* and *him*, in the sentence '*If you happen to meet the admiral, don't tell him his ship's gone down*'. Without having *the admiral*, it is not possible to interpret *him*. Such relations, forming meaning in the text, are called cohesion by Halliday and Hassan (1976), which is obviously not applicable to Vinay and Darbelnet's identification of translation units; the words *the admiral* and *him* cannot be a translation unit. These items do not act as a single unit in the translation process.

3.3.2. No leftover principle

The no leftover principle was suggested by Toury (1986; 1995). This principle is based on his assumption regarding translation; all the elements in the original texts are 'represented by' their counterparts in a translation (1986). Therefore, when translation units are examined under this principle, one has to focus on matching all the elements in the original texts with all the elements in the translated texts, without having any leftover elements. In other words, this principle denies the existence of zero correspondence, i.e. items of no counterpart between source and target texts (Johansson, 2007).

For example, *mi te* 'observing' in the above example (3.3.1) is an added item in translation. If one looks at the original text, there are no words denoting the meaning of *mi te*: *By and large, moderate growth is still the most likely scenario so far* (rendered into *Zentai teki ni mi te, yuruyaka na keezai seechoo toiu no ga, ima no tokoro mottomo kanoosei no takai shinario da* 'Observing generally, moderate growth is the most likely scenario so far'). Since the principle denies zero correspondence, one has to find where *mi te* was rendered from. What one can do is: to go above the word level in the text; to form a larger unit of *mi te* at the phrase level, i.e. *zentai teki ni mi te* 'observing generally'; and then, to identify its translation unit, i.e. *by and large*. Even though *mi te* has no direct counterpart in the original text, it is possible to find the translation unit of *mi te*.

Another example is *the* in the same example; it is an omitted item in translation. The original sentence has *the*, but the Japanese translation has no word denoting the meaning of *the*. Since the principle denies zero correspondence, it assumes that the meaning of *the* was rendered somewhere in the Japanese sentence. Again, in order to find what *the* was translated into, one goes above the word level, forms a larger unit of *the* at the phrase level, i.e. *the most likely scenario*, and identifies the translation equivalent, i.e. *mottomo kanoosei no takai shinario*. The translation pair, *the most likely scenario* and its equivalent *mottomo kanoosei no takai shinario*, can then be extracted without problems.

However, the principle is sometimes troublesome. *Still* in the same example belongs to this case. *Still* is an omitted item in translation; the original sentence has *still*, but the Japanese translation does not have words denoting the meaning of *still*. *Still* cannot be analysed as *mi te* and *the* were. One can try to go above the word level and form a larger unit of *still* at the phrase level, but what could be the unit at the phrase level: *is still*, or *still the most likely scenario*? *Still* does not have a close relationship with any of the other elements in the sentence, like *the* does to the noun phrase *most likely scenario*. *The* is dependant on *most likely scenario*; while, *still* is an element of solitude at the phrase level. In order to form a larger unit of *still*, one needs to go up to the clause level and form a unit, i.e. *moderate growth is still the most likely scenario so far*, and identify its translation equivalent, i.e. *yuruyaka na keezai seechoo toiu no ga, ima no tokoro mottomo kanoosei no takai shinario da*. However, this makes the whole sentence a translation unit.

Some cases like *still* are better to be leftover as items of zero correspondence. According to Baker (1992), omission often happens if 'the meaning conveyed by a particular item or expression is not vital enough to the development of the text'. An item can simply be omitted without being rendered anywhere in the translated text. In such cases, it is rather pointless trying to find the counterpart, which does not exist in the text. Unless the item forms a larger unit at the phrase level like *mi te* and *the*, such items should be leftover, and labeled as zero correspondences. Otherwise, as discussed in the example of *still*, the principle causes one to regard the whole clause as a translation unit. A principle that denies zero correspondences in translation is an unrealistic one to follow.

In order to avoid potential confusion, it is useful to note that the terms 'zero correspondence' and 'non-equivalence' are used differently in this study. Both refer to the phenomenon where there is no direct counterpart between source and target texts (Baker, 1992; Johansson, 2007); however, the perspective differs. The former is the phenomenon as judged by linguists; while, the latter is the phenomenon as judged by translators. For example, the items *mi te*, *the*, and *still* are examples of zero correspondence in the above instances. Both original and translated texts were compared and it was concluded that these units had no direct counterparts. On the other hand, non-equivalence is found when translators try to render texts. *The* is a good example. It has no equivalents in Japanese; therefore, it causes difficulty for translators. A translator may give up trying to render *the* by omitting it, then zero correspondence of *the* would occur (this is the case of the above example from the ARC). The Genius English-Japanese dictionary (2001) states that this happens often. Another translator

may try to render it, then zero correspondence of *the* would not occur, since *the* would have its counterpart in translation. Since this thesis is not concerned, in the main, with translator's strategies, the term 'non-equivalence' is not needed as often as the term 'zero correspondence'.

3.3.3. Monosemous principle

The monosemous principle was suggested by Teubert (2004b). This principle is based on one attribute of a translation unit; a translation unit is monosemous, that is, it has only one meaning (Teubert, 2004b). This becomes clearer by comparing translation units with single words. Single words cannot often be translation units; this is because they are ambiguous in their meanings and, therefore, difficult to render without the context. For example, how can a translator render the preposition *by* without the context? On the other hand, the phrase *by and large* is easy to translate as it is. It has a clear meaning unlike the single word *by*; therefore, a translator can render it without consulting any additional context. The items like *by and large* are translation units in the sense that they are inseparable units for translation. The crucial feature of such units, therefore, is their monosemy. Taking this attribute further, then, if one identifies a unit which has one clear meaning such as *by and large*, this unit is likely to be a translation unit. This is the basic idea of the monosemous principle.

Such identification of monosemous items is not easy to carry out. However, Teubert (2004b) suggests that it is possible if one examines given items from the translation perspective (i.e. how a given item is rendered in parallel corpora); a monosemous item

63

is likely to be rendered into the same translation however many times it occurs, because the meaning is always the same. For example, the item *by and large* is likely to be rendered into the same translation every time it occurs; due to its single, unambiguous meaning. Wherever it appears, the meaning of this item remains the same; therefore, the corresponding items in another language should be the same as well. On the other hand, it is unlikely that the single word *by* always corresponds to the same item in another language, due to the ambiguous meaning of *by*. This sounds a promising method for recognising translation units. The judgment of translation units is unbiased, relying only on the recurrence of translations (i.e. how often a given item is rendered into the same translation) in a corpus. It is fair to say, however, that this is quite a strict criterion. Even *by and large*, that is guaranteed to be a translation unit by Shuttleworth and Cowie (1997), cannot meet this criterion in the ARC. It occurred twice, and the translations were different, e.g. *zentai teki ni mi te* 'observing generally', and *zentai teki ni* 'generally'. The strict monosemous principle would filter this item out: it would not be regarded as a translation unit.

Regarding this point, Teubert (2004b) loosens the principle by setting up one additional condition for translation units, taking account of the cases like *by and large* in the ARC: synonyms in translations. That is: even though a given item is rendered into several translations in a researcher's corpus or corpora, if the translations are synonyms, the item can be regarded as a translation unit (Teubert, 2004b). With this filter it is possible to make the above item, *by and large*, a translation unit as far as the ARC data is concerned, since the two translations (i.e. *zentai teki ni mi te* 'observing generally' and *zentai teki ni* 'generally') can be regarded as synonyms. However, this additional

criterion causes some problems as well (see below). For the sake of convenience, the translation pair identified by this additional criterion is called 'one-to-many monosemy' (referring to the monosemous relationship between one translation unit and its many synonymous equivalents), distinguishing it from the pair identified by the stricter principle called 'one-to-one monosemy' (referring to the monosemous relationship between one translation unit and its single equivalent).

One example of one-to-one monosemy is *legal adviser* from the English-Chinese study of Wang (2006). *Legal adviser* has only one translation equivalent in the HKLDC (see the details of the corpus in 2.3). Wang (2006) examined thirty examples of *legal adviser* in the corpus and all the examples of *legal adviser* were rendered into *fa lü gu wen* (法律顾问). Therefore, *legal adviser* was regarded as a monosemous unit; hence, it was a translation unit. One example of one-to-many monosemy is *necessary modifications*. Wang (2006) examined again thirty examples of *necessary modifications* in the HKLDC; four translation equivalents of *necessary modifications* were found: (a) *bi yao de bian tong* (必要的变通); (b) *bi xu de bian tong* (必需的变通); (c) *xu yao de bian tong* (需要的变通); and, (d) *bi xu de xiu gai* (必需的修改). Wang (2006) interpreted these as synonymous; therefore, under the monosemous principle, with the additional synonymy criterion, *necessary modifications* was regarded as a monosemous unit; hence, it was concluded that it was a translation unit. Needless to say, if an item has several translation equivalents in a corpus, and the equivalents are not synonyms, then, the item is not regarded as a translation unit. For example, *conclusive evidence* was not regarded as a translation unit as its Chinese translation equivalents found in the corpus,

que zheng (确证) and *bu ke tui fan de zheng ju* (不可推翻的证据), are not synonyms (Wang, 2006).

The monosemous principle gives clear guidance on how to identify translation units; in particular, the identification is straightforward in the case of one-to-one monosemy. This principle, however, has a methodological weakness in the case of one-to-many monosemy. The identification involves examining if the translation equivalents are synonymous or not. This is a hard task to achieve systematically. For example, as noted above, there are four translation equivalents of *necessary modifications* found in the HKLDC. They are highly likely to be synonyms to some extent; since they were all rendered from *necessary modifications*; it is inevitable that they share some aspects of the meaning of *necessary modifications*. One possible way to examine such synonymy would be to ask English-Chinese translators; one can extract sentences where (a) *bi yao de bian tong* (必要的变通) was used from the corpus, replace (a) *bi yao de bian tong* into (b) *bi xu de bian tong* (必需的变通), and see if they think such replaced items are odd in the context. However, such questionnaires are always difficult to interpret. The judgment is intuitive and biased; it is often the case that the outcomes would be a mixture of opinion (i.e. some say that they are synonyms and some disagree with it).

Comparing the case of *necessary modifications* with the case of *conclusive evidence* is more confusing. There are two translation equivalents of *conclusive evidence* in the corpus: *que zheng* and *bu ke tui fan de zheng ju*. They were rendered from the one

single item *conclusive evidence*; therefore, they are again very likely to be synonyms to a certain degree. However, the analysis (Wang, 2006) showed that *conclusive evidence* was not a translation unit. This was because one translation equivalent *que zheng* has a literal meaning of 'factual evidence'. On the other hand, the other translation equivalent *bu ke tui fan de zheng ju* has a literal meaning of 'impossible overthrown evidence' (Wang, 2006). Due to the slight semantic difference caused by the different word combinations, the two equivalents were regarded as non-synonyms. However, this judgment is not quite consistent, if one compares this analysis with the analysis of *necessary modifications*. If the different word combinations cause them to be non-synonymous, then, one wonders why this does not apply to the case of *necessary modifications*, as well. Despite different word combinations, all the four equivalents (*bi yao de bian tong*, *bi xu de bian tong*, *xu yao de bian tong*, and *bi xu de xiu gai*) were regarded as synonyms. Wang's (2006) research shows that the judgement of synonyms is hard to systematise, and therefore the resulting analyses can often only be rather intuitive. Such subjective judgments can become inconsistent as data-sets get larger. Again, this is the matter of irreproducibility.

3.3.4. Units of meaning

The identification of translation units with the help of units of meaning was suggested by Danielsson (2001). This method is based on the assumption that a translation unit is a unit of meaning (Danielsson, 2001). According to Danielsson (2001), the unit of meaning is not necessarily a single word, it is often larger than that, such as *keep an eye on*. The single word *keep* has a 'complex, fuzzy, and ambiguous link to

67

content/meaning' (Danielsson, 2001), which does not allow *keep* to be labeled as a unit of meaning; hence, *keep* is not a translation unit. On the other hand, the larger unit *keep an eye on* is a unit of meaning, due to the meaning being unambiguous; hence, *keep an eye on* is a translation unit. Being a unit of meaning is necessary for being a translation unit. Therefore, the identification of translation units focuses on finding units of meaning such as *keep an eye on*.

One might wonder if this is similar to the monosemous principle (Teubert, 2004b). Apart from the fact that Teubert (2004b) uses the word 'monosemous' whilst Danielsson (2001) uses the term 'units of meaning', their basic definitions are the same: a translation unit is a unit which consists of one unambiguous meaning. Their methods and judgment of monosemous units (or units of meaning) differ, however. In Teubert's method (2004b), monosemous units are identified in parallel corpora; the examination of recurrent translations (i.e. how often a given item is rendered into the same translation) is the crucial criterion. On the other hand, in Danielsson's method (2001), units of meaning are identified in monolingual corpora; frequent collocates appearing five times or more are all regarded as units of meanings, which means that the criterion was frequency in a monolingual corpus only. In other words, what Teubert (2004b) wants to identify is monosemous units from the bilingual perspective; while, what Danielsson (2001) wants to identify is monosemous units (i.e. units of meaning) from the monolingual perspective.

Danielsson (2001) extracted translation units using three steps: (a) extracting units of meaning in a Swedish monolingual corpus; (b) examining how the units of meaning were rendered into English in the Swedish-English parallel corpus; and (c) matching the

Swedish units of meaning and their English equivalents. At the stage of (a), the 50 million-word Swedish corpus, Språkbanken, was used (Danielsson, 2001). One hundred of the frequent words occurring more than 200 times in the corpus was chosen as the sample words. For each sample word, their collocates (i.e. a 'recurrent word normally occurring within five words of a given keyword' (2001)) were examined, and the frequent collocates appearing five times or more were regarded as units of meaning. In total, 12,099 units of meaning were found, e.g. *gjorde ont* 'hurt'. All of them were regarded as candidates for translation unit status.

At the stage of (b), all the 12,099 candidates were examined in the 1 million-word fiction parallel corpus (the Swedish originals and their English translations) and their translation equivalents were identified. For example, the translation unit candidate *gjorde en rörelse* was found to be rendered into *made a movement* in the corpus. Not all of the 12,099 Swedish units were found in the parallel corpus; this is inevitable if one considers the size difference between the monolingual and parallel corpora; Språkbanken is about 100 times larger than the Swedish part of the parallel corpus. Only 51 candidates were found in the corpus, and their translation equivalents were recognised. At the stage of (c), among the 51 candidates, the units that occurred only twice or less in the parallel corpus were discarded, due to a lack of data. Seven candidates remained and were identified as translation units. Overall, seven pairs of translation units and their English translation equivalents were extracted.

One good thing about this method is that the judgment of translation units is unbiased and operatable. There are two criteria to meet: a given collocation is a translation unit if it occurs more than five times in a monolingual corpus (from the stage (a)), and occurs

three times or more in a parallel corpus (from the stage (b) and (c)). The occurrence

check in the two types of corpora is the only test that linguists have to carry out for the

identification of translation units. This means that translation units identified in this

method are simply frequent collocates in two corpora. However, this is not what most

linguists regard as translation units, as discussed in 2.3.1: a translation unit is an

inseparable unit for translation. It is not guaranteed that all frequent collocates are

deserving of being called translation units yet. Therefore, Danielsson's method of

identification is not used for this study; it does not identify what are generally

understood as translation units (see 2.3.1). Moreover, there is another fatal drawback for

using this method in the current study. All the target words are collocates in the method

used in Danielsson (2001); single words cannot be examined. Although translation units

are not necessarily single words, there is still the possibility that some translation units

are single words. Many theorists, indeed, posit the existence of single word translation

units (shown in Figure 2.2). Due to these two crucial issues, this method was not used in

this study.

Tognini-Bonelli (2001; 2002) is another theorist who has also identified translation

units with the help of units of meaning (Tognini-Bonelli uses the term 'functionally

complete units of meaning'). Unlike the study of Danielsson (2001) where units of

meaning are identified only by their frequencies, Tognini-Bonelli (2002) identifies units

of meaning from four aspects: collocations ('two or more words that co-occur more than

once' (Danielsson, 2001)); colligation ('the co-occurrence of grammatical choice'

(Sinclair, 1996)); semantic preference ('the co-occurrence of words with semantic

choices' (Sinclair, 2004b)); and, semantic prosody ('a sort of attitudinal or pragmatic

meaning' (Sinclair, 2004a)). Since the latter two are semantic properties, it is inevitable that this method for identifying units of meaning must be carried out manually, which could yield only a small amount of translation units. Indeed, Tognini-Bonelli (2001; 2002) only managed to identify a few pairs of translation units and their equivalents. As this study wanted to identify translation units quantitatively for its research questions, this method was not used.

3.3.5. Summary

The main research question of the current study is: to investigate how large or small translation units are, in English-Japanese newswire texts, rendered by experienced translators. It has been argued that none of the four existing methods are suitable for this research question. First, identification using the degree of cohesion method is rather subjective. This is because the term 'cohesion' is not well defined, and units of weak cohesion and no cohesion, in particular, are difficult to recognise in a systematic way. Seeing that the data-sets used in this study are much larger than the one Vinay and Darbelnet (1995) use, the judgment of translation units under this criterion is likely to be inconsistent. Thus, the degree of cohesion methodology was not used for this thesis.

Second, the no leftover principle is an unrealistic method to follow. The principle denies zero correspondences. However, in reality one can often see items which were simply omitted in translation, and have no counterpart in the translated texts. In such cases, analysis under this principle may cause incompatibility with the definition of a translation unit. Thus, this principle was not implemented in this study either. Third,

identification with the monosemous principle is an objective and straightforward method only in the case of one-to-one monosemy. If the case of one-to-many monosemy is examined, then, the identification tends to be biased. There is a risk that the judgments lose their consistency, especially in this study, as the data-sets are relatively large.

Lastly, identification using units of meaning is not suitable for the aims of this study. The drawbacks of this method are that: (a) it does not identify translation units as inseparable units for translation; and, (b) it does not allow one to examine single word translation units. The different definition of translation units (point (a)) and the impossibility of examining single word translation units (point (b)) are fatal, as the research aim cannot be fulfilled. Thus, identification using the units of meaning method was not used in this thesis research, either.

3.4.   An alternative method

In considering an alternative method, objectivity was the most important factor. This is because the data-sets used in this thesis are large enough to require automated or semi-automated processes, and therefore judgement relying on intuition only was not preferred due to the high chance of inconsistency. Among the four existing methods, only the identification methods using the monosemous principle and units of meaning principle are designed for quantitative corpus research. Therefore, I have attempted to modify the two methods for the aims of this thesis by overcoming their drawbacks.

First, identification with the units of meaning method takes a given item as a translation unit if the item is frequently recurrent in monolingual corpora and in parallel corpora. As discussed in 3.3.4, this method has the problematic issue of single word translation units; single words cannot be identified as translation units. This problem needs to be overcome for the current research. This, however, is not feasible. In the identification process, all the frequent collocations (i.e. ones occurring more than five times in monolingual corpus and occurring three times or more in parallel corpus) are regarded as translation units. How is it possible to apply this criterion to single words? A simple application only leads to an unrealistic result; all the single words meeting these conditions can be translation units. There would be easily thousands of such translation units identified. Is it appropriate to say that all the frequent single words can be translation units? Alternatively, one can set up a cut-off point, for example, 100 times instead of five times in monolingual and three times in parallel corpora, but again, the result would be that all the single words occurring more than 100 times are translation units. As long as the frequency of a given item is the criterion to determine if the item is a translation unit or not, this identification method cannot be applied to single words. Also, as discussed in 3.3.4, the identified translation units in such a method are not the translation units that this study wants to extract. These two drawbacks do not seem to be surmounted using this method.

Second, identification with the monosemous principle takes a given item as a translation unit, if (a) all examples of the item in the corpus were rendered into one single translation equivalent, i.e. one-to-one monosemy, or (b) all examples of the item in the corpus were rendered into several but synonymous translation equivalents, i.e.

one-to-many monosemy. As discussed in 3.3.3, this principle uses the thorny issue of synonymy, related to the case of one-to-many monosemy; the judgment of synonyms tends to be intuitive. One possible way to deal with this problem is to establish a systematic way of judging synonyms. Examining collocations, i.e. 'two or more words that co-occur more than once' (Danielsson, 2001), might be helpful, as semantic similarities and differences often become observable from a collocational perspective.

For example, Stubbs (1996) takes the examples of *couch* and *sofa*. Although they have the same denotation (i.e. 'literal meaning'), they occur in different collocations: *casting couch*, *couch-potato*, *psychiatrist's couch*, and *sofa-bed*. *Couch* fits better with some surroundings, whilst *sofa* fits better with other surroundings. This means that *couch* and *sofa* are not actually synonyms in a strict sense; they are not interchangeable due to their connotations, that is, their 'affective, associative, attitudinal and emotive meaning' (Stubbs, 1996). Such a collocational examination can be done in an objective and systematic way; however, it is very rare that two or more items share exactly the same collocations. After all, there is no synonymy, in a strict sense, in a language system, since 'there would seem to be no reason why a language should have words which mean exactly the same' (Stubbs, 1996), due to their differential connotations. Taking this 'no synonymy' position further, the case of one-to-many monosemy in the monosemous principle can be dismissed. There are no synonyms in language; therefore, if a given item has several translation equivalents in a corpus, the item should not be recognized as a translation unit. The monosemous method, therefore, is modified into a more objective and repeatable analytical system by dismissing the case of one-to-many monosemy.

3.4.1. One-equivalent principle

This thesis therefore takes a radical point of view on synonyms; no synonyms exist in a language system. Therefore, the one-to-many monosemy of the monosemous principle is abandoned. Now, if a given item is found to have more than one translation equivalent, then, the item is not a translation unit; if a given item is found to have only one translation equivalent, then the item is accepted as a translation unit. This criterion provides a straightforward judgment on how to determine translation units in an objective manner. For the sake of the convenience of discussion, this is called the 'one-equivalent principle' in this thesis.

3.4.2. Method

Translation units were identified in this study quantitatively. Taking one of my target words, *market*, as an example, I first randomly extracted 2,000 lines of *market* and their corresponding Japanese lines out of the ARC. Second, I identified the Japanese translations of *market* semi-automatically; I used ParaConc (Barlow, 1995) (see details in 3.4.6) for this identification, followed by thorough manual checking. Third, I discarded some noise examples (see details in 3.4.3). Finally, I selected the first 1,000 pairs of examples of *market* for analysis, which was the platform for the identification of translation units (see relevant discussion in 3.4.4).

The first task on the platform was to examine if the single word *market* was a translation unit or not. The criterion was a one-to-one relationship between a translation

unit and its equivalent under the one-equivalent principle. If all the 1,000 samples of *market* were rendered into one Japanese translation equivalent, it could be regarded as a translation unit. On the other hand, if *market* had more than one translation equivalents in the 1,000 lines, then the single word *market* was not considered as a translation unit, indicating that translation units were larger than the single word *market*.

The second task was to extract units larger than *market* and see whether they could be regarded as translation units or not. In order to retrieve larger units, collocational information was used. By using the term 'collocation', I refer to 'two or more words that co-occur more than once' (Danielsson, 2001). I extracted collocations of two/three/four/five/six words using the 'Cluster' function (minimum frequency was 3 within the R5-L5 window) in WordSmith (Scott, 1996) (see details in 3.4.6). There were 363 two-word, 169 three-word, 45 four-word, 7 five-word, and 1 six-word collocations of *market* in the 1,000 sample set. Not all the collocations were examined. I first removed collocations which did not contain *market* (the 'cluster' function in WordSmith can contain spurious hits that lack the search word). Second, I extracted the grammatical sequences, i.e. phrases (nominal, adverbial, prepositional, and verbal) and clauses (subject-verb, verb-object, and subject-verb-object), and discarded the other non-syntactic collocations. The 'collocate' list consists of any adjacent sequences regardless of their grammatical relationship so that some entries are ungrammatical units. For example, *market and* was extracted as a two-word collocation by WordSmith; however, this item is not what this study is interested in. *Market and* is an incomplete unit at the phrase level; it is not a nominal phrase, nor an adverbial or prepositional

phrase. Therefore, along with *market and*, items such as *market in*, *the market to*, *market in the*, *market the* were discarded.

There were some units which looked like valid items, but were not if the concordance lines were examined. For example, *market brokers* was extracted as a two-word collocation; it occurred five times in the sample set. The concordance lines (Lines 1-5, Example 3.1), however, showed that it should not be regarded as a two-word collocation; *market* and *brokers* were divided by a comma and they do not form a phrase.

Example 3.1. Invalid units

```
1   … support for the Shanghai stock market, brokers said.
2 … market today and helped boost the market, brokers said.
3   … interest rates also dented the market, brokers said.
4 … remained a dominant theme, of the market, brokers said.
5    … trade may pull down the whole market, brokers said.
```

There is one thing to note regarding the one-equivalent principle. I made one exception to the criteron: an item was regarded as a translation unit, even if the item had more than one translation equivalent, only if one of these equivalents was very dominant (85 percent of the examples or more). For example, *market sources* occurred in 14 lines; 13 of them (93 percent) had a Japanese translation equivalent and one of them (7 percent) had another Japanese translation equivalent. In this case, since *market sources* had a translation equivalent which took up more than 85 percent of the examples, it was regarded as a translation unit. This is beneficial for filtering out mistranslations and creative translations (see relevant discussion in 3.2.2). By saying creative translations, I refer to 'the creating of new words' and 'the novel collocation of existing words'

(Kenny, 2001). As discussed in 2.3, this study is concerned with expressions which are already established in 'a universe of discourse'. It is important that 'unusual' translations do not disturb the identification of such established translation units. Therefore, such a filter is necessarily. The setting of my cut-off level at 85 percent instead of 90 percent or 60 percent was arbitrary; I assumed that mistranslations and creative translations were unlikely to occur in more than 15 percent of the Reuter news articles. I discuss whether this level was appropriate or not in 7.5.2.

3.4.3.   Noise examples

The noise examples I removed out of the sample sets need to be clarified (briefly discussed in 3.4.2). There were three types: zero correspondences; proper noun usages; and, duplicated examples. An example of zero correspondences (i.e. examples in which *market* was not rendered into Japanese) was: *Stocks are not building up as fast as the market had expected*, which was rendered into *Zaiko no tsumimashi peesu wa, yosoo yori mo osoi* 'Stock building up is slower than expected'. Since *market* was not rendered, the example provides little information on translation units of *market*; therefore, zero correspondences are not examined in this study. An example of the proper noun usage (i.e. where *market* was used in proper nouns) of target words were: *Federal Open Market Committees*. It is rather obvious that these are inseparable units for translation; however, these are not the focus of this study. If *market* appeared capitalised in the middle of a sentence, it was removed from the sample set.

78

I also rejected duplicated examples (i.e. the same sentences in the source and/or target texts); a 'sentence' simply means a stretch between two full stops in the corpus. One example was: *British Telecommunications Plc has long harboured hopes of cracking the French market and unveiled on September 17 in Lyon, central France, its "Pleiade" service for a national service for long-distance calls for French companies*. This exact sentence occurred twice in the 2,000 pre-sample set of *market*. I rejected them as noise data. This was because the duplicates in the ARC were not suitable data for identifying translation units. It is often the case that 'real' translation did not occur where the duplicates were found in the ARC.

Example 3.2 is a good example for this claim. Lines 1-3 were found in the pre-sample set of *new*. The comparison between Lines 1-3 showed that each of them carried different information. There were only two sets of content words which appeared in common: *new technology/ies* and *financial services*. Despite such significant differences, the ARC told that Lines 1-3 were rendered into the exactly same Japanese translation: Line 4. This is dubious, however. It is very unlikely that Line 1 was actually rendered into Line 4; Line 1 had too little information as the source text of Line 4. Similarly, it is hard to believe that Line 2 was rendered into Line 4 either; Line 2 did not have enough information to be the original text of Line 4. Line 3 was the closest match with Line 4; the two shared almost the same information, except the subject usage: *Minehan* in Line 3 but *soosai* 'President' in Line 4. Therefore, Line 3 was likely to be the source text of Line 4. Lines 1-2 were not the 'real' source texts that Line 4 was rendered from. Since translation units required 'real' translation data for their identification, Lines 1-2 were not suitable.

79

Example 3.2. Duplicated examples of *new* (i)

```
1  The conference focused on the impact of new technology on banking and
   financial services.
2  The three-day conference focused on the challenges, risks and opportunities
   new technologies are creating in the financial services sector.
3  Minehan said the Fed "can and should play a central role" in assessing the
   risks associated with new technology while guaranteeing stability in the
   financial services sector.
4  Soosai wa, renpoo jyunbi riji kai (FRB) wa, kinyuu saabisu bunya no antee o h
   oshoo shi nagara mo, atarashii tekunorojii ni tomonau risuku he no taioo ni o
   ite, chuushin teki yakuwari o ninau koto ga kanoo de ari, mata soo su beki da
    to nobe ta
   'President said that FRB can and should play a central role in dealing with
   the risk associated with the new technology while guaranteeing stability in
   the financial services sector'
```

One might wonder why the cases like Example 3.2 happened in the ARC. There are two possibilities. One was that the alignment between the source and target sentences was not appropriate where duplicates occurred in the ARC. The other was that translations were re-used by translators when duplicates occurred. Dr Masao Utiyama, the senior researcher who is in charge of the ARC at the NIST in Japan, posited both possibilities (personal communication). This thesis did not investigate which one of these possibilities caused the case of Example 3.2 and other similar cases. However, an important implication from Example 3.2 was that the ARC did not necessarily tell the 'real' alignment between source and target sentences in the duplicated example sets. Thus, they were regarded as noise data in this study.

3.4.4. Preliminary work

All the findings in this study are based on 1,000 sample lines of each target word. The figure 1,000 was suggested by a series of pilot studies of the most frequent noun, adjective, and verb in the ARC: *market*; *new*; and, *said* respectively. The first pilot study

was carried out with 100 sample lines of *market*; this is a 100 sample set after the process of eliminating the noise examples. The single word *market* was first examined to identify if it was a translation unit or not. *Market* was rendered into seven different translation equivalents; the most dominant translation was *shijoo* (occurring in 78 out of 100 lines), followed by the second most dominant one *sooba* (occurring in 12 out of 100 lines). Under the one-equivalent principle, the single word *market* would not be identified as a translation unit. Second, larger units than the single word *market* were examined and the collocations of *market* were extracted. There were only 17 collocations (occurring three times or more), e.g. *the market* (42), *stock market* (11), *market was* (6), *the market is* (5), *on the market* (4); the numbers in parentheses show the frequencies in the sample.

Although WordSmith reported that *the market* occurred 42 times, if one looks at the concordance lines and sees if all the phrases are syntactic sequences or not, not all of them are valid for further analysis (shown in Example 3.3). For example, *the market* in Line 2 is an incomplete syntactic phrase, as *the market economy* is the whole phrase. Therefore, this example was discarded. Similarly, *the market* in Line 4 was removed for the same reason (see relevant examples in 4.1.2). After checking the concordance lines of the 17 collocations manually and screening out such examples, there were only seven grammatical collocations of *market* remaining: two nominal phrases, *the market* (36) and *stock market* (4); two prepositional phrases, *on the market* (4) and *in the market* (3); and three clauses, *market is* (5), *market was* (4), and *the market was* (4). Among them, no translation unit was identified under the one-equivalent principle. This was not satisfactory for the aim of the thesis.

Example 3.3. Pilot study: concordances of *the market*

```
1              Some traders said the market appeared to be biding its …
2  … not conform to the pattern of the market economy," it said.
3              "This shows that the market is very highly valued and a …
4      … interest rates lower than the market level, to help boost the …
```

The second pilot study was carried out with 500 sample lines of *market*; again the samples did not include duplicated examples, zero correspondences, and proper name usages. The first examination was to see if the single word *market* was regarded as a translation unit or not. It was rendered into fourteen different translation equivalents. The most dominant translation was again *shijoo*, (occurring in 394 out of 500 lines, 78 percent), followed by the second one *sooba* (occurring in 62 out of 500 lines, 12 percent). Therefore, the single word *market* was not identified as a translation unit as it is. Next, the collocations were examined. There were 121 of them identified. Their grammatical circumstances were checked in the concordance lines manually and the incomplete syntactic phrases were removed. After these processes, 47 of them remained: 26 noun phrases, e.g. *market sentiment* and *currency market*, 9 prepositional phrases, e.g. *for the market* and *out of the market*, and 9 clauses, e.g. *market sources said* and *pressured the market*. Under the one-equivalent principle, nine of them were identified as translation units, shown in Table 3.2

Table 3.2. Pilot study: translation units of *market*

stock market	*market talk*	*gold market*
market sources	*foreign exchange market*	*the copper market*
labor market	*B share market*	*market will reopen*

This is an improvement on the first pilot study. The 100 sample lines brought nil translation units. On the other hand, the 500 sample lines brought 9 translation units.

This is not ideal, but is a manageable amount of translation units for further discussion on translation unit size. In particular, the discussion should be illuminating when the other target words are examined, and their translation units and their sizes are compared.

The target words in this study (which will be discussed in 3.4.5) were not only nouns, but also adjectives and verbs. Therefore, I carried out two more pilot studies and checked if the figure of 500 samples was suitable for them as well. However, the study of the adjective *new* with 500 sample lines managed to identify only three translation units: *new company*, *new orders*, and *new low*. Similarly, the study of the verb *said* with 500 sample lines managed to identify only one translation units: *a trader said*. These studies indicated that a sample of 500 is not sufficient for the analyses of *new* and *said*. It would be possible to examine nouns in 500 sample lines and adjectives and verbs in larger sample lines. However, this would make interpretation harder. Therefore, a sample set of 500 lines was discarded for all the analyses of the three parts of speech. Instead, a sample set of 1,000 lines was used. The figure 1,000 is a manageably large set and reasonable for the size of the corpus (see 3.4.5).

3.4.5. Target words

Due to methodological reasons, the target words have to be frequent words that occur at least 1,000 times in the ARC. In fact, it is desirable that they occur 2,000 times or more in the corpus, since there are always examples which cannot be included in the sample sets: duplicated examples, zero correspondences, and proper name usages (discussed in

3.4.2). According to WordSmith (see details in 3.4.6), the ARC has 103 frequent words which appear 2,000 times or more. Table 3.3 shows the top ten.

Table 3.3. Frequency list of the ARC

No	Word	Freq
1	*the*	107,067
2	*to*	47,608
3	*of*	45,432
4	*in*	40,589
5	*said*	38,796
6	*a*	38,493
7	*and*	33,250
8	*on*	29,827
9	*for*	15,407
10	*that*	14,361

The most frequent words, as shown in Table 3.3, are dominantly function words, i.e. a word 'with little or no intrinsic semantic content which primarily serves some grammatical purpose' (Trask, 1993). Using Sinclair's term (1999), the function words in Table 3.3 are divided into two groups: words used as (a) 'reference', e.g. *the*, *a*, and *that*, and, (b) 'syntactic extension' with prepositions, e.g. *of*, *to*, *in*, *for*, and *on*, or with conjunctions, e.g. *and*. There is one word in Table 3.3, which does not belong to the category of function word: *said*. It is a content word, i.e. a 'word with predominantly lexical significance, such as a noun, a verb or an adjective' (Hartmann and James, 1998); it is also called a 'vocabulary word' (Sinclair, 1999). This scarcity of content words in the frequency list is not unique at all; '[a]mong the most frequent words there are only a very few nouns and verbs which can be said to have a meaning of their own' (Teubert and Čermáková, 2004). The reason for *said* being the most frequent content word in the ARC is probably because the corpus consists of news texts. The same picture can be seen in the frequency list of the Bank of English, where texts of media

language are dominant; *said* is the most frequent content word, followed by *new*, *people*, *year*, and *years* (Sinclair, 1999).

This study focused on content words, not function words. This is because 'meaning is the core issue of translation'; what links between the original and translated text is 'meaning alone' (Teubert, 2002). Therefore, the semantic transformation between languages, i.e. how the content words are rendered, is the primary aspect for studies of translation units. The grammatical transformation between languages, i.e. how the function words are rendered, is regarded as an extensive project, which is not included in this thesis due to the space limitation. Following the definition of Hartmann and James (1998), three word categories were regarded as content words: noun, verb, and adjective. Since the ARC is not a tagged corpus, the corpus software, Wmatrix (Rayson, 2003; 2007), was used for tagging the corpus and creating the frequency lists for each word category. There are 24 nouns, 5 adjectives, and 14 verbs, which are frequent enough to be target words for this study.

First, there are only five adjectives occurring 2,000 times or more in the ARC (Table 3.4, tagged and counted by Wmatrix). The most frequent three, i.e. *new*, *economic*, and *foreign*, were chosen to be target words. *Central* and *monetary* were not chosen. Considering the amount of manual labour involved in analysis, examining three words from each word category seemed reasonable.

Table 3.4. Frequency list of the ARC: adjectives

No	Word	Freq
1	*economic*	3,206
2	*new*	2,653
3	*foreign*	2,575
4	*central*	2,193
5	*monetary*	2,041

Second, there are 24 nouns occurring 2,000 times or more in the ARC. The most frequent ten are shown in Table 3.5 (tagged and counted by Wmatix). Some words are not suitable for this study. First, proper nouns, e.g. *U.S.*, were excluded. Since the translation units of proper names are beyond the interest of this thesis, *U.S.* was not chosen to be a target word. Second, there are four nouns for days of the week in Table 3.5, *Tuesday*, *Wednesday*, *Monday*, and *Thursday*; Wmatrix (Rayson, 2003; 2007) gave these the tag of 'weekday noun', distinguishing them from the general nouns. They were not considered suitable for this study either. Unlike the noun *market* which has 47 collocations in the sample 500 lines (discussed in 3.4.4), the weekday noun *Tuesday* has only one collocation in the first 500 lines out of the ARC: the prepositional phrase *on Tuesday*. The collocational variations are far less than for *market*. Since collocational information plays one of the key roles in the identification method in this study, *Tuesday* was not suitable. Assuming that the other weekday nouns have as few collocations as *Tuesday*, weekday nouns were not chosen for this study.

Table 3.5. Frequency list of the ARC: nouns

No	Word	Freq
1	*market*	5,125
2	*U.S.*	4,975
3	*Tuesday*	4,028
4	*bank*	3,921
5	*Wednesday*	3,762
6	*Monday*	3,669
7	*Thursday*	3,624
8	*government*	3,560
9	*prices*	3,335
10	*year*	3,269

The nouns remaining in Table 3.5 were, according to the tags of Wmatrix (Rayson, 2003; 2007), common nouns (e.g. *market*, *bank*, *government*, and *prices*) and a temporal noun (e.g. *year*). Among them, *market*, *government*, and *year* were chosen as the target words; while, *bank* and *prices* were discarded. Both *bank* and *prices* are financial words, as well as *market*. Due to the variety of target words, the finance-oriented word *market*, the politics-oriented word *government*, and the time-oriented word *year* were selected.

Lastly, there are fourteen verbs occurring 2,000 times or more in the ARC. The most frequent ones are shown in Table 3.6. Although Wmatrix (Rayson, 2003; 2007) includes modal auxiliaries, e.g. *would* and *will* in the frequency list of verbs, they were not content words (Hartmann and James, 1998), therefore, they were discarded. However, *would* and *will* are not the only words that are used as auxiliaries in Table 3.6. In fact, most of the words in Table 3.6 are often used as auxiliaries as well as main verbs. If they form the continuous tense or passive voice, *be*, *was*, *were*, *are*, and *is* act as auxiliaries. If it forms the past perfect tense, *has* acts as an auxiliary as well. Since the tags of Wmatrix do not distinguish *was* as a main verb from *was* as an auxiliary, the extraction of the former data for the analysis has to be performed manually, which is

87

rather time-consuming. For the same reasons, all the words belonging to *BE*, i.e. *be*, *was*, *were*, *being*, *am*, *been*, *are*, and *is*, *DO*, i.e. *do*, *did*, *doing*, *done*, and *does*, and *HAVE*, i.e. *have*, *had*, *having*, and *has*, were removed.

Table 3.6. Frequency list of the ARC: verbs

No	Word	Freq
1	*said*	38,340
2	*was*	11,382
3	*is*	10,165
4	*be*	8,422
5	*would*	7,313
6	*will*	7,169
7	*has*	5,288
8	*were*	5,234
9	*had*	5,015
10	*are*	4,878

There are only two main verbs that occur 2,000 times or more in the ARC (Table 3.7): *said* and *told*. Seeing that three adjectives and three nouns were chosen, having three verbs was seen as preferable. Therefore, I took the third frequent verb *expected*, which occurred only 1,612 times in the ARC, and examined if it was possible to have 1,000 sample lines of *expected* after removing all the noise examples. Due to the fact that there is no proper name usage and fewer zero correspondences, I managed to collect 1,000 sample lines of *expected*. Thus, *said*, *told* and *expected* were chosen for this study.

Table 3.7. Frequency list of the ARC: main verbs

No	Word	Freq
1	*said*	38,340
2	*told*	3,398
3	*expected*	1,612
4	*rose*	1,430
5	*closed*	1,224

To sum up, the target words in this study are frequent nouns (i.e. *market*, *government*, and *year*), frequent adjectives (i.e. *economic*, *new*, and *foreign*), and frequent verbs (i.e. *said*, *told*, and *expected*). I could not expand the number of the target words to more than nine, since the figure 1,000 was chosen for the sample lines (as discussed in 3.4.4), which means that target words should be frequent enough to appear in the corpus 2,000 times or more. Therefore, the figure of sample lines and the number of target words are closely connected. If one is happy with a small number of sample lines, there are plenty of target words from which to choose; however, this type of analysis is not guaranteed to extract many translation units (as demonstrated in 3.4.4). On the other hand, if one wants to have a large number of sample lines, there are not many choices for the target words; but, it is possible to identify a reasonable amount of translation units. This dilemma is only solved if a large parallel corpus is available. The ARC only consists of 1.8 million words for English texts and 2.5 million morphemes for Japanese texts (counted by ParaConc (Barlow, 1995)). If one has corpora larger than the ARC, it would be possible to examine more than nine target words, without changing the number of sample lines.

3.4.6. Analytical tools

This study used two software programmes to deal with the corpus data: ParaConc (Barlow, 1995) and WordSmith (Scott, 1996). First, ParaConc is a multilingual concordancer. The version used was ParaConc v2.0.369. Its major roles in this analysis were to display bilingual concordance lines during analyses and manipulate them with a series of functions, i.e. sorting the contexts, finding translations equivalents, and

extracting pairs of data. Example 3.4 shows the display screen of the ten pairs of the

English lines of *market* and their Japanese translations. Each example has line numbers

to show the correspondences between the English and Japanese sentences. The keyword

market is underlined and their translations are also underlined. The concordances in

Example 3.4 were sorted by the words in Japanese sentences, i.e. 1st right of the

underlined words.

Example 3.4. ParaConc display of *market*
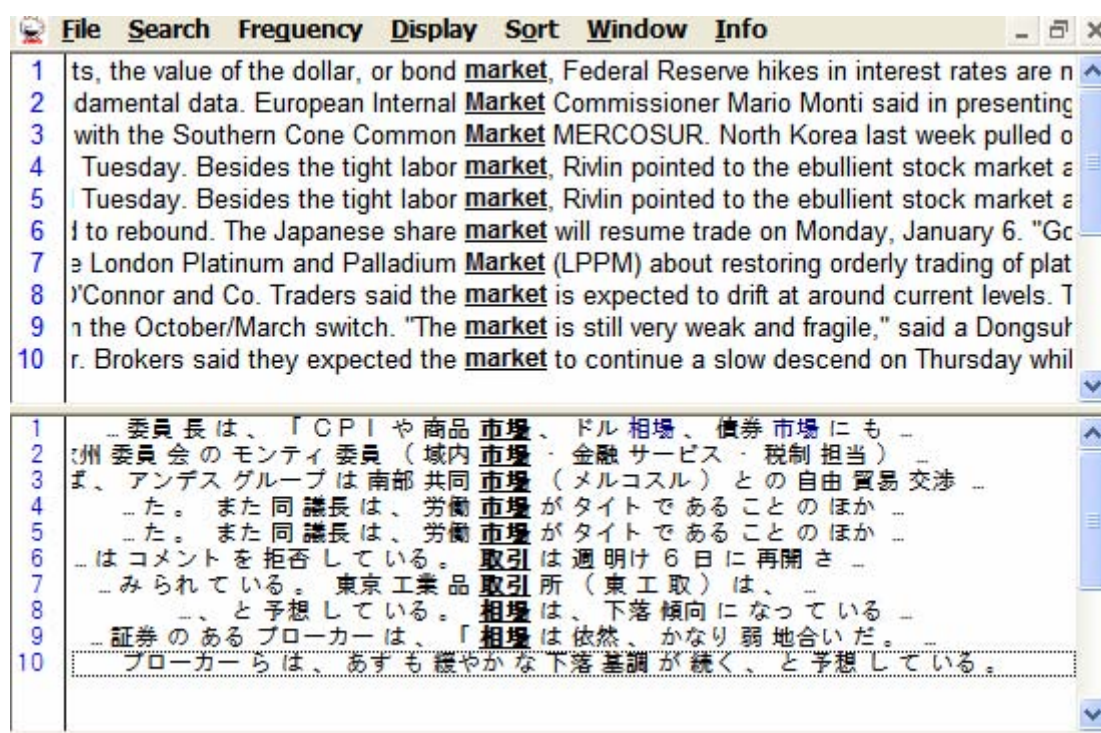
ParaConc was an indispensable tool in this study, especially for the process of selecting

the 1,000 sample lines. The process had four steps (see in 3.4.2) and all the data

manipulation was done in this concordance programme. First, taking *market* as an

example again, ParaConc found 7,853 examples of *market* and 2,000 of them were

randomly extracted along with their Japanese translations. Since ParaConc can only

extract examples of every Nth line, the extracted examples were not actually 2,000 lines; every 4th in the 7,853 examples of *market* was extracted, therefore, there were 1,963 lines retrieved. I could have used WordSmith instead, which allows one to extract the exact 2,000 lines randomly. However, WordSmith is a monolingual concordancer; therefore, it only extracts the English lines. Finding and retrieving their matching Japanese lines would be very time-consuming. On the other hand, ParaConc extracted around 2,000 lines of English and Japanese instantly, which was much preferred.

Second, the 'Hot Words' function of ParaConc suggested the possible translations of a given word, based on a statistical 'association rate' in the corpus data (P. Danielsson, personal communication). This function works better when data-sets are large. For example, in the case of 7,853 examples of *market* in the ARC, ParaConc suggested 10 translations: *shijoo* 'market', *maaketto* 'market', *FOMC* 'Federal Open Market Committee', *sooba* 'market', *kookai* 'exhibition', *shea* 'share', *jiai* 'market tone', *kansan* 'inactive', *burookaa* 'broker', and *kabushiki* 'stock' (the first definition from the Genius Japanese-English Dictionary (2003) are shown in the quotation marks, except one example *jiai*, whose definition is from the largest Japanese-English online dictionary, ALC (www.alc.co.jp)). Among the ten, only five of them were appropriate translations of *market*: *shijoo*, *maaketto*, *FOMC*, *sooba*, and *jiai*.

In the case of 1,963 examples of *market* in the ARC, ParaConc also suggested 10 translations; however, only one of them was a relevant translation: *maaketto* 'market'. Similarly, ParaConc managed to provide only one suitable translation, *shijoo* 'market', when the 10 examples of *market* (Example 3.4) were examined. Therefore, in order to

extract as many as possible appropriate translations by this function, I used all the data

for a given word in the ARC, e.g. 7,853 examples for *market*. For the analysis of *market*,

thus, the five translation equivalents *shijoo*, *maaketto*, *FOMC*, *sooba*, and *jiai* were

identified and underlined in the display screen of ParaConc (Example 3.4). Among the

sample sets of 1,963 lines of *market*, 83 percent of them (i.e. 1,604 lines) were rendered

into one of the five translations. The rest of the examples of *market* were manually

examined to investigate how *market* was rendered; and their translation equivalents

were also highlighted. Since the automatic identification might be wrong, the 1,604

examples of *market* were also checked manually. This manual check was done on the

screen on ParaConc as well.

Third, the dual display of ParaConc was useful, when noise examples were removed out

of the ~2,000 pre-sample sets: (a) examples of proper nouns, e.g. Lines 2-3 and 7 in

Example 3.4, (b) duplicated examples, e.g. Lines 4-5, (c) zero correspondences, e.g.

Line 10. When the English examples were removed out of the screen, the matching

Japanese lines were also deleted at the same time. Lastly, I sorted all the concordance

lines in the original order and chose the first 1,000 pairs of examples of *market* for

analysis on the screen. The pairs of sentences were saved as the sample set. In the

identification process, the one-equivalent principle was examined on the screen of

ParaConc as well. Therefore, ParaConc was the main and crucial tool of this thesis, due

to the functions mentioned above and the effective display for viewing the bilingual

concordance lines simultaneously and allowing linguists to manipulate data first-hand.

Another software programme used in this study was WordSmith (Scott, 1996). The version used was WordSmith v. 4.0.0.365. When it was found that target words were not single word translation units, their collocations were considered (see in 3.4.2). The major role of WordSmith in this study was to extract such collocations. The 'Cluster' function of WordSmith was used for this. The cluster settings were: (a) clusters consist of two to six words; (b) the minimum frequency is three times; and (c) a fixed-size window is 5L and 5R, i.e. 5 words before the given word and five words after the given word. The setting of (a) allowed WordSmith to make a list of word sequences which contain between two and six words. The reason I chose six word sequences as a maximum size was that there were no seven word sequences found, as long as the nine target words were only being analysed.

Example 3.5. Cluster list of *market*

No	Cluster	Freq
1	THE MARKET	415
2	IN THE	142
3	MARKET WAS	78
4	STOCK MARKET	73
5	ON THE	70
6	MARKET IS	66
7	OF THE	56
8	THE MARKET WAS	51
9	THE MARKET IS	38
10	SAID THE	37

The setting of (b) regarding the minimum frequency was set at three times after a series of pilot studies. For the case of *market*, using a minimum frequency of three, WordSmith found 585 word sequences; the most frequent ones are shown in Example 3.5. As introduced in 3.4.2, not all the word sequences in the list were appropriate for analysis. After removing the inappropriate sequences, e.g. *in the*, *on the*, *of the*, and *said*

the, only 66 collocations remained. If the minimum frequency setting was left with the default value, i.e. five, there would only have been only 28 collocations of *market* for the actual analysis. Similarly, the same default setting meant that WordSmith found only 7 collocations of *new* and only 14 collocations of *said*. These numbers of collocations are too few for the analyses, particularly in the cases of *new* and *said*. Therefore, the default value of the minimum frequency was replaced by three. The setting of (c) regarding the window size, on the other hand, remained at the default, i.e. 5L and 5R.

3.4.7. Advantages

One advantage of this method is its objectivity in identifying translation units. The one-equivalent principle was a straightforward criterion to follow. The judgment whether a given item was a translation unit or not was as unbiased as possible. This makes the research reproducible. If somebody else will do the same research with the same data, the result (i.e. the identified translation units) would be same. Because of this reproducibility, this current study would be comparable with others in which the same methodology will be used (see the relevant discussion in 8.1.2). In the systematic review of translation unit size, such method has not been found.

The other advantage of this identification method was that it allowed the examination of the target words at single word, phrase, clause, and sentence levels. First, the single word *market* was examined to see if it was a translation unit or not; if not, the phrases of *market* were examined; if not, the clauses and sentences of *market* were examined. Any frequent collocations of *market* at all levels were investigated thoroughly. Compared

with the methods of Danielsson (2001), who examined word sequences of two or more words only, and Wang (2006), who examined two word phrases only, the method presented here is unique, in the sense that all the levels between single word and sentence were examined. This identification method suits the research aim: how large and small are the translation units in the language pair of English and Japanese in newswire texts rendered by experienced translators?

The last advantage of this identification method was that this study examined words from three different word categories: nouns, adjectives, and verbs. This allowed the outcome to be examined with a balanced view on the translation unit size of frequent content words. Having three target words from each word category was made possible by the size of sample sets (i.e. 1,000 lines). As discussed in 3.4.5, the larger the sample sets are, the more translation units can be identified; however, only a few words are able to be target words. Considering the aim of this thesis, it is preferable to examine several target words from the different word groups and to identify several translation units for each. For this purpose, the figure 1,000 was suitably large. The target words were nine i.e. three each of nouns, adjectives, and verbs; and at the same time, the sample sets were expected to yield about a dozen translation units for each word (discussed in 3.4.4).

3.4.8. Limitations

The limitations of the methodology need to be clarified as well. First, the one-equivalent principle made the definition of translation units strict; a translation unit is a monosemous unit which has only one translation equivalent in the corpus. In other words, a translation unit is a unit which has a one-to-one relationship with its translation equivalent. This definition is narrower than those put forward by many theorists (see in 2.3). Therefore, it is inevitable that this principle could identify a smaller number of translation units than other theorists. For example, Danielsson (2001) identified seven translation units; however, if the one-equivalent principle was used in her research, only three translation units would have been identified. Similarly, Wang (2006) examined 30 frequent phrases and 25 of them were identified as translation units; however, if the one-equivalent principle had been imposed, only 20 of these would have been recognised as translation units.

Second, this study used only collocational information when larger units were extracted. For example, when *market* was examined to see if it was a translation unit above the word level, frequent collocations were regarded as candidates for translation units: e.g. *stock market*, *market sources*, and *market will reopen* (3.4.2). However, as Sinclair (1996) points out, collocation is only one type of word co-occurrence relationship. There are three more: colligation – 'the co-occurrence of grammatical choice' (1996); semantic preference – 'the co-occurrence of words with semantic choices' (2004b); and semantic prosody – 'a sort of attitudinal or pragmatic meaning' (2004a). It was possible to investigate whether, (i) colligations of *market* such as '*market for* N', (ii) semantic

preferences of *market* such as '*market* associated with stock exchange', or (iii) semantic prosody of *market* such as '*market* in negative sense', can also be translation units or not. These questions, however, are left for future study, due to space limitations.

Lastly, this method made it possible to identify translation units at the levels of single word, phrase, and clause. However, translation units at the morpheme level were not investigated (3.4.2). An investigation at the morpheme level would have required creating another set of sample lines, which was unrealistic due to the time constraints. Generally speaking, '[t]ranslation on the level of the morpheme is even more rare than translation on the level of phoneme', according to Barkhudarov (1993). Unless the two languages share morphological similarities, translation units at the morpheme level do not occur (2.4.2). Taking the example of some grammatical morphemes, the morphological systems between English and Japanese are quite different.

For example, English has the plural marker *–s*; while Japanese has no equivalent for it. When *fingers* and *tattoos* were rendered into Japanese, the plural markers were lost in translation (Baker, 1992); *It concerns fingers, or rather the lack of them. And tattoos*, which was rendered into *Sore wa yubi to, tsumari yubi no kesson to kanren suru. Sarani irezumi tomo* 'this concerns finger, or rather the lack of finger. And tattoo' (Baker, 1992)). There is no plural marker for *finger* and *tattoo* in translation, since '[t]he form of a noun in these languages [= Japanese, Chinese, and Vietnamese] does not normally indicate whether it is singular or plural' (Baker, 1992). Japanese does not have the same plurality system at the morpheme level as English. Other grammatical morphemes such as past tense markers differ between English and Japanese as well. The suffix *–ta*

attaches to verbs and adjectives in Japanese (Takahashi et al., 2005); while the English equivalent *–(e)d* only attaches to verbs in English. Considering such differences, it is unlikely that translation happens at the level of morpheme when it comes to my target words *market*, *government*, *year*, *economic*, *new*, *foreign*, *said*, *told*, and *expected*. Thus, I discarded the possibility of morphemes as translation units regarding the nine target words selected.

3.4.9. Summary

In order to fulfill the aim of this thesis, the one-equivalent principle was used for identifying translation units. This criterion is beneficial for its unbiased reproducibility. There was a dilemma when the number of sample lines and the number of target words were set up (3.4.5). However, after a series of pilot studies, the number of sample lines was set to 1,000, which allowed this study to have nine target words from different word groups: most frequent nouns (i.e. *market*, *government*, and *year*), most frequent adjectives (*economic*, *new*, and *foreign*), and most frequent verbs (i.e. *said*, *told*, and *expected*) in the ARC. Also, the number 1,000 was expected to be enough to identify at least a manageable amount of translation units for each word. This study allowed the investigation of translation units at the level of the single word, phrase, clause, and sentence, in English-Japanese Reuter newswire texts, rendered by experienced translators. In particular, the method was aimed to be appropriate for my research interest, namely, which of the sizes were dominant among single word, phrase, clause, and sentence?

98

3.5. Hypotheses

Which sizes are likely to be translation units in the ARC? Hypotheses were formulated from the four perspectives discussed in 2.4. First, whether a translation unit is defined as a lexical or cognitive unit affects how small or large the unit is; lexical translation units tends to be smaller and cognitive ones tend to be larger (see details in 2.4.1). Since the translation units in this study are lexical, the sizes of outcome are expected to be at the lower levels, i.e. at the word, phrase, clause, and sentence level, according to Figure 2.2.

Second, whether the two languages are closely related or unrelated affects how small or large the translation units are; units between closely related languages tend to be smaller, while units between unrelated languages tend to be larger (see details in 2.4.2). The language pair of English and Japanese is likely to belong to the latter case. In addition to the morphological differences (argued in 3.4.8), the two languages differ from the language typology point of view as well. English is in the Indo-European language family (Ross, 2006) and Japanese is in the Altaic language family (Miller, 1980; Masaomi Kondo and Wakabayashi, 1998; Shibatani, 2006). Matsubara et al. (2000) summarise the differences of the core grammatical systems between English and Japanese (Table 3.8). As shown clearly in Table 3.8, the language systems do not overlap at all.

Table 3.8. Comparison between Japanese and English (Matsubara et al., 2000)

	Japanese	English
1) Position of the predicate	around the end of a sentence	after the subject
2) Flexibility of word order	flexible	strict
3) Interrogative expression	end of a sentence	head of a sentence
4) Negative expression	end of a sentence	after the subject
5) Subjective case ellipses	very frequent	very few

However, this does not help very much when it comes to predicting specific translation unit sizes. As discussed in 2.4.2, there has been little empirical study on language pairs outside the Indo-European languages. Toury's research on the English-Hebrew comparison (1986) was the only one of such studies; he identified translation units at the levels of phrase and clause. Since both English-Hebrew and Japanese-English comparisons were the language pairs of English and a language outside the Indo-European family, Toury's finding (1986) was taken as translation unit sizes of unrelated language pairs. Needless to say, this assumption was not as reliable as the ones from the other variables, in which several literatures were considered (reflected by weaker shading in Figure 3.1).

Third, whether the given corpus consists of technical documents, news texts, academic articles, or literature texts affects the size of translation units; translation units in news texts are likely to be larger than those in technical documents, but smaller than those in literature (see details in 2.4.3). Presuming that newswire texts belong to the general news texts category, the translation unit sizes in the ARC are likely to be at the level of single words and phrases (Figure 2.7). Finally, whether language data are rendered by inexperienced or experienced translators affects the size as well; inexperienced

translators tend to use smaller translation units and experienced translators tend to use larger ones (see details in 2.4.4). The ARC consists of texts translated by experienced translators; therefore, from this perspective, the sizes of outcome are likely to be phrases, clauses, and sentences (Figure 2.9). Figure 3.1 presents a summary of all the hypotheses from these four aspects.

Figure 3.1. Hypotheses on translation unit size

		mo.	w.	ph.	cl.	sen.	par.	tex.
Definitions:	lexical							
Language pairs:	unrelated							
Text types:	news text							
Translators:	experienced							
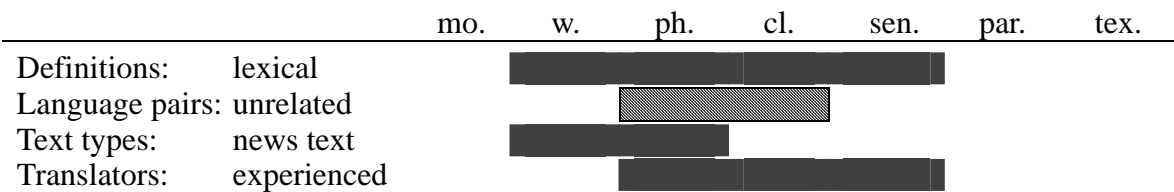
Therefore, as far as these four aspects were concerned, the translation units are likely to be single words, phrases, clauses, and sentences. The other sizes such as morphemes, paragraphs, and texts were not expected to be translation units in this study. Among the expected sizes, there are slight differences. For example, translation units at the single word level were supported by only two aspects: definitions and text types. The other two variables regarding language pair and translator were associated with translation units larger than single words. Therefore, single word translation units might not be identified very frequently; the same argument applies to the sentence level, it is only supported by two of the aspects determining translation unit size. Translation units at the clause level were supported by a larger set of variables: the definition of the translation unit; the language pair; and, type of translator; however, newswire texts were empirically found to be associated with translation units smaller than the clause (Teubert, 2001; Alves and Gonçalves, 2003). On the other hand, translation units at the

phrase level were supported by all the four variables: definitions, language pair, text type, and translator. Thus, the phrase level was expected to be the translation unit size identified most dominantly in this study, while, the single word level and sentence were expected to be the translation unit sizes identified the least dominantly. Clausal translation units were likely to be identified more often than single word and sentence ones, but less often than phrasal translation units.

It was methodologically impossible, however, to identify translation unit at the sentence level in this study. This was because of the removal of duplicated sentences out of the sample sets. For example, the sentence, *The GM deal would be the first under a new approach, with the USOC and NBC working jointly to enlist corporations for both Olympic sponsorship and TV advertising*, appeared three times in the pre-sample set of *new*, i.e. 2,512 lines of *new*. However, all of them were removed, since, as discussed in 3.4.3, the real translation did not seem to occur where the duplicates were found in the ARC. Therefore, all the sentences in the sample set could not reach the minimum frequency; no sentences could be translation unit candidates.

Therefore, hypothesis I formulated in Figure 3.1 needed to be modified. Consulting with the relevant literature and taking into account the research conditions, translation units in this study were expected to be single words, phrases, and clauses. Among the three, the most likely translation units were phrases; while, the least likely translation units were single words in this piece of research.

3.6. Conclusion

This chapter has discussed the methodology that I implemented in this thesis research. First, the corpus data used in this study, the ARC, was clarified with its advantages and disadvantages. The ARC is, at present, the largest parallel corpus in the public domain for the study of English translation units from the Japanese perspective. In addition to the size, the ARC has the linguistic, mechanical and diachronic advantages discussed in 3.2. Second, the existing methodologies for identifying translation units were discussed in 3.3. There were four approaches: ones using the degree of cohesion (Vinay and Darbelnet, 1995); the no leftover principle (Toury, 1986; 1995); the units of meaning principle (Danielsson, 2001); and, the monosemous principle (Teubert, 2004b). None of these were used in this thesis; the reasons for this were clarified. Put very generally, objectivity was lacking in these methods. Therefore, the most systematic method, the monosemous principle, was altered and an alternative method was introduced: the 'one-equivalent principle' in 3.4. Given these research conditions, I formulated hypotheses concerning translation unit sizes from four perspectives: definition of translation unit; language pair; text type; and, translator's quality. The outcomes of the analyses (i.e. identified translation units) were expected to be at the three levels: single word; phrase; and, clause. The phrase-level was likely to be the most dominant; the single word-level was likely to be the least dominant. This hypothesis will be re-visited in 7.1, following the case studies on the nouns (*market*, *government*, and *year*) in Chapter 4, the adjectives (*economic*, *new*, and *foreign*) in Chapter 5, and the verbs (*said*, *told*, and *expected*) in Chapter 6.

4. Analyses of nouns

This chapter will present the identification of translation units of the frequent nouns, *market*, *year*, and *government*, by the one-equivalent principle. Whether their translation units are identifiable at the level of single word, phrase, or clause is the main question to be solved.

4.1. Translation units

4.1.1. Word level

The first task was to examine if translation units could be at the level of the single word. The criterion used was the one-equivalent principle, i.e. a one-to-one relationship between a translation unit and its equivalent in the sample set (discussed in 3.4.1). My first target word, *market*, had 19 Japanese translations in the 1,000 pairs of translations. Table 4.1 shows the top five of these. Each translation is listed with: (i) the first two definitions from the Genius Japanese-English Dictionary (2003) to provide a guiding definition; (ii) parts of speech based on the comprehensive Japanese dictionary, Daijirin containing 250,000 words (http://dictionary.goo.ne.jp/) and the usage in the given context; and, (iii) raw frequencies and the percentage occurrence out of the 1,000 instances.

Table 4.1. *Market*

	Translations		P.O.S.	Freq.	(%)
1	*shijoo*	(market/marketplace)	noun	789	(79%)
2	*sooba*	(market/price)	noun	125	(13%)
3	*torihiki*	(business/dealing)	noun	37	(4%)
4	*maaketto*	(market/supermarket)	noun	23	(2%)
5	*akinai*	(business/trade)	noun	5	(1%)
			Total	979	(98%)

Market was mainly translated into either *shijoo* or *sooba* (Table 4.1). The most frequent translation, *shijoo*, was very dominant, appearing in 789 out of the 1,000 lines; while the second, *sooba*, was rather minor, appearing only in 125 lines. The rest of the translations each occurred in less than 5 percent of the total examples. Since *market* has more than one translation equivalent, and the most dominant translation equivalent *shijoo* does not achieve a high coverage of 85 percent, the single word *market* was not regarded as a translation unit.

Similarly, the second target word, *year*, was not recognised as a single word translation unit either. The 1,000 sample lines showed that there were 25 translations of *year*. The top five are shown in Table 4.2. None of the five Japanese translations appeared to be clearly dominant. The most frequent translation *nen* only appeared in 302 lines out of the total 1,000 examples; if one compares with *market*, the most frequent translation of which appeared in 789 out of 1,000 lines (Table 4.1), this translation *nen* is far less frequent. Since *year* has more than one translation, and the most dominant one did not reach the required coverage of 85 percent, the one-equivalent principle could not identify the single word *year* as a translation unit.

Table 4.2. *Year*

	Translations			Freq.	(%)
1	*nen*	(year)	noun	302	(30%)
2	*kotoshi*	(this year)	noun	196	(20%)
3	*zennen*	(the previous year/ the year before)	noun	182	(18%)
4	*rainen*	(next year)	noun	67	(7%)
5	*sakunen*	(last year)	noun	58	(6%)
			Total	805	(80%)

The third target word, *government*, however, was different. It had 14 Japanese translations in the 1,000 sample lines. Table 4.3 presents the most frequent five translations; each translation is listed again with the first two definitions from the Genius Japanese-English Dictionary (2003) to provide a rough guide to meaning; however, for the fourth translation *seechoo*, Jim Breen's WWWJDIC Japanese Dictionary (http://www.csse.monash.edu.au/~jwb/wwwjdic.html) was used. Unlike *market* and *year*, the most frequent translation of *government*, *seefu*, covered 85 percent, as shown in Table 4.3. The rest of the translations each took up only 8 percent of the total examples or less. Based on the one-equivalent principle, the single word *government* was regarded as a translation unit. Among the three most frequent nouns, *government* was the only translation unit identified at the single word level.

Table 4.3. *Government*

	Translations			Freq.	(%)
1	*seefu*	(government/administration)	noun	849	(85%)
2	*seeken*	(power/regime)	noun	82	(8%)
3	*tookyoku*	(authority)	noun	23	(2%)
4	*seechoo*	(government office)	noun	19	(2%)
5	*naikaku*	(cabinet/ministry)	noun	8	(1%)
			Total	981	(98%)

4.1.2. Above word level

The next task was to extract larger units of *market* and *year* and to identify their translation units at the levels of phrase, and clause. For *market*, there were 66 collocations found in the sample set: 45 nominal phrases; 12 prepositional phrases; and, 9 clauses. Some from each category are listed in Example 4.1, with their raw frequencies in parentheses.

Example 4.1. Examined collocations of *market* (selected)

Nominal phrases	Prepositional phrases	Clauses
the market (371)	*on the market* (33)	*market will reopen* (8)
stock market (20)	*in the market* (30)	*market was quiet* (8)
market sentiment (17)	*of the market* (17)	*market sources said* (7)
market sources (14)	*for the market* (12)	*market reacted* (3)
foreign exchange market (10)	*to the market* (12)	*support the market* (3)

The examined nominal phrases included: determiner-noun sequences, e.g. *the market*; noun-noun sequences, e.g. *stock market* (some appeared with an extra adjective, e.g. *foreign exchange market*, or with an extra noun, e.g. *market interest rates*); and adjective-noun sequences, e.g. *domestic market* (some occurred with an extra noun, e.g. *domestic copper market*). The prepositional phrases examined in this analysis occurred with *on*, *in*, *of*, *to*, *for*, *into*, *as*, *by*, *out of*, or *over*. The examined clauses included: subject-verb sequences, e.g. *market will reopen*; and verb-object sequences, e.g. *support the market*, *supported the market*, and *to support the market*. Based on the one-equivalent principle, 10 of them could be regarded as translation units. Table 4.4 presents all the translation units of *market*. The translation units and their corresponding translation equivalents are listed with the raw frequencies in parentheses, and the

corresponding items of *market* in bold. Word segmentation in Japanese was made using the software ChaSen (Matsumoto et al., 2002) in the ARC (Utiyama and Isahara, 2003).

Table 4.4. Translation units of *market*

	Translation units	Translation equivalents
a.	*market sources* (12)	***shijoo*** *suji* (12)
b.	*market rumours* (4)	***shijoo*** *no uwasa* (4)
c.	*market economy* (3)	***shijoo*** *keezai* (3)
d.	*stock market* (20)	*kabushiki* ***shijoo*** (20)
e.	*gold market* (5)	*kin* ***shijoo*** (5)
f.	*copper market* (4)	*doo* ***shijoo*** (4)
g.	*market interest rates* (3)	***shijoo*** *kinri* (3)
h.	*Seoul stock market* (5)	*sooru kabushiki* ***shijoo*** (5)
i.	*domestic gold market* (3)	*kokunai kin* ***shijoo*** (3)
j.	*market will reopen* (8)	***torihiki*** *wa saikai sa reru* (7)

The translation units of *market* were predominantly at the phrase level; nine of the ten were phrasal translation units (Table 4.4 *a-i*). They were all nominal phrases (noun-noun and adjective-noun sequences); no prepositional phrases were identified as translation units. There was only one translation unit identified at the clause level, a subject-verb sequence (Table 4.4 *j*). The proportion of clausal translation units was very low. It was noteworthy that *market* in the clausal translation unit *market will reopen* was translated into *torihiki*. This was the only one translation in which *market* corresponded to this translation; *market* in all the other translation units was rendered into *shijoo*.

Similarly, phrasal translation units were dominant in the case of *year* as well. There were 48 collocations (Example 4.2): 17 nominal phrases (determiner-noun, noun-noun, adjective-noun, and noun-adverb sequences); 29 prepositional phrases (with *of*, *per*, *for*, *in*, *from*, *for*, *than*, *over*, *on*, *with*, *from*, *since*, *over*, and *about*); and, 2 clauses (verb-object sequences).

108

Example 4.2. Examined collocations of *year* (selected)

Nominal phrases	Prepositional phrases	Clauses
this year (330)	*of the year* (62)	*compared with last year* (3)
last year (93)	*of this year* (58)	*compared with the previous year* (3)
year ago (42)	*per year* (15)	
previous year (28)	*for the year* (15)	
marketing year (8)	*of next year* (13)	

Among all the 48 collocations, only two of them were regarded as translation units of *year* (Table 4.5). These were both at the level of the phrase. To be more specific, they were nominal phrases (noun-noun and noun-adverb sequences); no prepositional phrases were seen as translation units. Also, no clauses were regarded as translation units.

Table 4.5. Translation units of *year*

	Translation units	Translation equivalents
a.	*a year ago* (42)	***zennen*** (36)
b.	*marketing year* (7)	*shijoo **nendo*** (6)

It is worth mentioning that the concordance lines were carefully selected for the identification of translation units. For example, when the noun-noun phrase *gold market* was examined to see if it could be a translation unit or not, Lines 1-5 were examined (Example 4.3). The other occurrences of *gold market,* Lines 6-10, were discarded, due to the extra elements used to form phrases: for example, *the world's biggest* in Line 6; *a domestic* in Lines 7-9; and *the Hong Kong* in Line 10. The point was that this study wanted to investigate whether the phrase *gold market* could be a translation unit or not; if *gold market* is only a part of the phrase (Lines 6-10), it is hard to see how the phrase *gold market* alone was translated. This left five lines, Lines 1-5, for the investigation;

the bilingual concordance lines showed that *gold market* in these examples were all

rendered to *kin shijoo*; hence, it was recognised as a translation unit.

Example 4.3. Concordances of *gold market*

```
 1 … huge profits," WGC manager of  gold market  analyst George Milling-Stanley …
 2   … has taken steps to create a   gold market  by freeing the prices at …
 3  … slow process of setting up a   gold market  .
 4                            "The   gold market  was showing little reaction to …
 5 … the way for now, and that the   gold market  had managed to absorb both the …
 6  In India, the world's biggest    gold market  , analysts forecast lower prices …
 7      … and to create a domestic   gold market  .
 8      … up to create a domestic    gold market  via measures that would make it …
 9   … to liberalise the domestic    gold market  to channel commercial bank funds …
10                 The Hong Kong     gold market  is closed on Friday and Monday …
```

One might say that *gold market* in Lines 2-5 (Example 4.3) should also have been

removed from the examined set, as well as Lines 6-10. The nominal phrase in Line 2,

for example, is *a gold market*; *gold market* alone is only a part of the phrase. Similarly,

Line 3 has *a* and Lines 4-5 have *the*; *gold market* alone is again a part of the phrase.

However, this view was not adopted in this study; the article was not taken into account

when this study distinguished valid examples from invalid ones for the investigation.

The reasons were two; one methodological and the other empirical.

First, Table 4.6 shows all the translation units and the numbers of their examined

concordance lines in parentheses. Most of the identified translation units had only five

valid examples or less (*g-m* Table 4.6). If I had taken the stricter view of article

existence, three translation units would have been disqualified before the analyses due

to a lack of data: (*g*) *gold market*; (*k*) *market economy*; and (*m*) *domestic gold market*.

There would have been only 10 identified translation units, which would have decreased

the size of the set and increased the difficulty of interpreting the findings from it.

Moreover, translation units are identified mostly in the cases of the nouns (argued in

3.4.4); the numbers of translation units identified in the cases of the adjectives and verbs

would have been less than 10 if I had taken the stricter view of articles This would not

have been satisfactory for the aim of this thesis, which is to investigate frequent

translation unit sizes.

Table 4.6. Translation units of nouns

	Translation units	
a.	*government*	(849)
b.	*a year ago*	(42)
c.	*stock market*	(20)
d.	*market sources*	(12)
e.	*market will reopen*	(8)
f.	*marketing year*	(7)
g.	~~*gold market*~~	(5)
h.	*Seoul stock market*	(5)
i.	*market rumours*	(4)
j.	*copper market*	(4)
k.	~~*market economy*~~	(3)
l.	*market interest rates*	(3)
m.	~~*domestic gold market*~~	(3)

The other reason came from the observation of the data. I examined 114 collocations of

the nouns (66 of *market* and 48 of *year*) to see if they could be translation units or not.

The examination of these collocations made it clear that taking the stricter view of

article existence was not a matter of great importance. The collocations showed that *the*

was not rendered into Japanese. For example, *property market* was used with the

definite article, shown in Lines 1-5, Example 4.4 (the words in parentheses are the

translations of the underlined items; words in bold correspond to *market*). *The property*

market was rendered into *fudoosan shijoo* 'property market' in Lines 1-2; *fudoosan*

shikyoo 'property market situation' in Line 3; *fudoosan kabu* 'property stock' in Line 4; and *shijoo* 'market' in Line 5. The meaning of *the* was not found in any translation; *property market* alone was rendered. By examining many similar cases to *property market*, taking the stricter view of articles revealed itself as of limited relevance. Articles are usually lost in translation; therefore, article variations such as *a gold market* and *the gold market* do not need to be eliminated from the examined set of *gold market* (further relevant discussion will be made in 4.3.1).

Example 4.4. Concordances of *property market*

```
1. capital gains, the property market may collapse, causing …   (fudoosan shijoo)
2    … is hitting the property market.                          (fudoosan shijoo)
3                 The property market, after years of …         (fudoosan shikyoo)
4               " The property market has a powerful voice …     (fudoosan kabu)
5    … prices in the property market.                           (shijoo)
```

So far, 13 translation units were identified in total in the case studies of the frequent nouns. In terms of their sizes, the analyses showed that:

1. Translation units were found at the level of the single word, phrase, and clause.
2. The identified translation units were predominantly at the phrase level.

4.2.   Translation equivalents

4.2.1. Sizes

The previous section (4.1) has shown that the most identified translation units were at the level of the phrase; 11 out of the 13 translation units were at the phrase level. Does this mean that their translation equivalents are predominantly at the level of the phrase

as well? Table 4.7 shows all the translation units and their equivalents; the grammatical

structures of the translation equivalents are listed for the sake of the discussion; the

target words are highlighted in bold and their corresponding translations, and their parts

of speech, are shown in bold as well. According to Table 4.7, ten translation equivalents

were indeed at the level of phrase (*c-l*). Therefore, the translation equivalents were

predominantly at the level of phrase, which was the same result as the translation unit

sizes. Thus, the translation unit sizes were kept through translations in most cases.

Table 4.7. Translation equivalents and their grammatical structures

	Translation units	Translation equivalents	
a.	***government***	***seefu***	**N**
b.	a ***year*** ago	***zennen***	**N**
c.	***market*** sources	***shijoo*** suji	**N** + N
d.	***market*** rumours	***shijoo*** no uwasa	**N** + N
e.	***market*** economy	***shijoo*** keezai	**N** + N
f.	stock ***market***	kabushiki ***shijoo***	N + **N**
g.	gold ***market***	kin ***shijoo***	N + **N**
h.	copper ***market***	doo ***shijoo***	N + **N**
i.	marketing ***year***	shijoo ***nendo***	N + **N**
j.	***market*** interest rates	***shijoo*** kinri	**N** + N
k.	Seoul stock ***market***	sooru kabushiki ***shijoo***	N + N+ **N**
l.	domestic gold ***market***	kokunai kin ***shijoo***	N + N+ **N**
m.	***market*** will reopen	***torihiki*** wa saikai sa reru	**N** + V

However, some translation pairs underwent a size change. Table 4.8 shows the numbers

of translation units and equivalents at each level. If one takes a closer look at the levels

of single word and phrase, one can spot some small value differences. For example,

there was only one single word translation unit identified, while there were two

translation equivalents found at the same level. Similarly, 11 phrasal translation units

were recognised; on the other hand, 10 translation equivalents were identified at the

phrasal level. Such incompatibility is not unusual in translation pairs. The relevant

literature often suggests that a translation unit and its equivalent are not necessarily at the same level: 'a collocation can become a clause; a whole clause can be reduced to a single word' (Teubert, 2002). These incompatible transformations are called 'shifts' ['changes which occur or may occur in the process of translating' (Bakker et al., 1998)]. Three types of such shifts were examined in this study: 'unit-shifts' (4.2.2); 'structure-shifts' (4.2.3); and 'class-shifts' (4.2.4). These investigations clarify our understanding of what was happening to the translation units in the newswire texts translated from English to Japanese.

Table 4.8. Sizes of translation units and equivalents

	Translation units	Translation equivalents
Single word	1	2
Phrase	11	10
Clause	1	1
Sentence	0	0
Total	13	13

4.2.2. Unit-shifts

The first shift to be examined was the 'unit-shift', which refers to a translation which involves changing the levels of the translation units ['the translation equivalent of a unit at one rank in the SL is a unit at a different rank in the TL'; for Catford (1965), the 'rank' refers to any one of the five units: sentence; clause; group; word; and morpheme]. Such a shift was seen in only one translation pair: *a year ago* and *zennen*. The phrasal translation unit *a year ago* was rendered into its single word translation equivalent *zennen* (Table 4.9 *a*); therefore, the level was changed in this translation pair.

Table 4.9. Unit-shifts (i)

	Translation units		Translation equivalents	
a.	*a **year** ago*	**N** + Adv	***zennen***	**N**

On the other hand, Table 4.10 shows that none of the other translation pairs underwent unit-shifts. The phrasal translation units were rendered into phrasal translation equivalents (Table 4.10 *b-k*). Similarly, the clausal translation unit was rendered into clausal equivalents (Table 4.10 *l*); the single word translation units corresponded to single words (Table 4.10 *m*). All the transformations were carried out at the same levels. Since a unit-shift was seen only in one translation pair, it could not be said to be a major characteristic of the translation equivalence of the frequent nouns.

Table 4.10. Unit-shifts (ii)

	Translation units		Translation equivalents	
b.	***market** sources*	**N** + N	***shijoo** suji*	**N** + N
c.	***market** rumours*	**N** + N	***shijoo** no uwasa*	**N** + N
d.	***market** economy*	**N** + N	***shijoo** keezai*	**N** + N
e.	*stock **market***	N + **N**	*kabushiki **shijoo***	N + **N**
f.	*gold **market***	N + **N**	*kin **shijoo***	N + **N**
g.	*copper **market***	N + **N**	*doo **shijoo***	N + **N**
h.	*marketing **year***	N + **N**	*shijoo **nendo***	N + **N**
i.	***market** interest rates*	**N** + N+ N	***shijoo** kinri*	**N** + N
j.	*Seoul stock **market***	N + N+ **N**	*sooru kabushiki **shijoo***	N + N+ **N**
k.	*domestic gold **market***	Adj + N + **N**	*kokunai kin **shijoo***	N + N+ **N**
l.	***market** will reopen*	**N** + Aux + V	***torihiki** wa saikai sa reru*	**N** + V
m.	***government***	**N**	***seefu***	**N**

4.2.3. Structure-shifts

Another shift, often seen in translation pairs in general (Catford, 1965), is the 'structure-shift', i.e. 'a change in grammatical structure between ST [= source text] and

TT [= target text]' (Shuttleworth and Cowie, 1997). In my studies of nouns, structure-shifts were found in four pairs (Table 4.11).

Table 4.11. Structure-shifts (i)

	Translation units		Translation equivalents	
a.	***market** interest rates*	**N + N+ N**	***shijoo** kinri*	**N + N**
b.	*a **year** ago*	**N + Adv**	***zennen***	**N**
c.	*domestic gold **market***	Adj + N + **N**	*kokunai kin **shijoo***	**N + N+ N**
d.	***market** will reopen*	**N + Aux + V**	***torihiki** wa saikai sa reru*	**N + V**

First, the translation unit *market interest rates* was composed by three nouns; however, it was rendered into an equivalent which had only two nouns (Table 4.11 *a*). This was because the two-word item *interest rates* corresponded to a single word in Japanese; the phrasal structure inevitably changed. Similarly, due to the existence of one single word for *a year ago* in Japanese, the translation pair, *a year ago* and *zennen* (Table 4.11 *b*), went through a structure-shift. Third, *domestic gold market* was composed by one adjective and two nouns in English; however, its equivalent was composed by three nouns, because of the transformation from the adjective *domestic* to the noun *kokunai* (Table 4.11 *c*). Lastly, the clausal translation unit *market will reopen* was composed by a noun followed by a future tense auxiliary and a verb; on the other hand, its Japanese equivalent lost the future tense (Table 4.11 *d*). As Kindaichi (1988) points out, Japanese future tense is not used as often as the English one, this is the most likely reason behind the structure-shift of this pair.

On the other hand, the other translation pairs kept their structures through translation (Table 4.12). The translation units composed by two nouns were rendered into their

equivalents composed by two nouns (Table 4.12 *e-k*). Similarly, three-noun translation units corresponded to three-noun equivalents (Table 4.12 *l*), and the same for single-noun translation units (Table 4.12 *m*). It is noteworthy that among the total 13 translation units, all the units composed by two nouns, i.e. N + N, belonged to this no-structure-shift group. This indicated that N + N kept their structures through translation better than other types of translation units such as N + N + N and clauses.

Table 4.12. Structure-shifts (ii)

	Translation units		Translation equivalents	
e.	*market* sources	**N** + N	*shijoo* suji	**N** + N
f.	*market* rumours	**N** + N	*shijoo* no uwasa	**N** + N
g.	*market* economy	**N** + N	*shijoo* keezai	**N** + N
h.	stock *market*	N + **N**	kabushiki *shijoo*	N + **N**
i.	gold *market*	N + **N**	kin *shijoo*	N + **N**
j.	copper *market*	N + **N**	doo *shijoo*	N + **N**
k.	marketing *year*	N + **N**	shijoo *nendo*	N + **N**
l.	Seoul stock *market*	N + N+ **N**	sooru kabushiki *shijoo*	N + N+ **N**
m.	*government*	**N**	*seefu*	**N**

4.2.4. Class-shifts

The other type of possible shift is a 'class-shift', i.e. 'the translation equivalent of a SL item is a member of a different class from the original item' (Catford, 1965). Catford's example (1965) was for the English word *medical* and the French word *médecine*; the adjective *medical* goes through a class-shift when the phrase *a medical student* is rendered into *un étudiant en médecine*. In French, *médecine* is a noun, which means that the adjectival class of *medical* in the English phrase was not retained. Among the 13 translation units of the nouns, such class-shifting was not seen at all. The single word translation unit *government* is a noun, and its equivalent is a noun as well (Table 4.13 *a*).

Similarly, all the phrasal translation units were nominal; their equivalents were nominal phrases (Table 4.13 *b-k*) or nouns (Table 4.13 *l*). No class-shifts occurred.

Table 4.13. Class-shifts

	Translation units		Translation equivalents	
a.	***government***	**N**	***seefu***	**N**
b.	***market*** *sources*	**N** + N	***shijoo*** *suji*	**N** + N
c.	***market*** *rumours*	**N** + N	***shijoo*** *no uwasa*	**N** + N
d.	***market*** *economy*	**N** + N	***shijoo*** *keezai*	**N** + N
e.	*stock* ***market***	N + **N**	*kabushiki* ***shijoo***	N + **N**
f.	*gold* ***market***	N + **N**	*kin* ***shijoo***	N + **N**
g.	*copper* ***market***	N + **N**	*doo* ***shijoo***	N + **N**
h.	*marketing* ***year***	N + **N**	*shijoo* ***nendo***	N + **N**
i.	*Seoul stock* ***market***	N + N+ **N**	*sooru kabushiki* ***shijoo***	N + N+ **N**
j.	***market*** *interest rates*	**N** + N+ N	***shijoo*** *kinri*	**N** + N
k.	*domestic gold* ***market***	Adj + N + **N**	*kokunai kin* ***shijoo***	N + N+ **N**
l.	*a* ***year ago***	**N** + Adv	***zennen***	**N**

Needless to say, examination of class-shift was not possible when a given translation unit was clausal; for example, *market will reopen*. A clause does not belong to any class, such as nouns, adjectives, etc., in the same way that a word or a phrase belongs to the nouns (nominal groups) or the adjectives (adjectival groups). Therefore, it was concluded that none of the translation pairs exhibited class-shift in the case of the frequent nouns.

4.2.5. Conclusion

As discussed in the previous section (4.1), the translation unit size of the frequent nouns was predominantly the phrase. Their translation equivalents were examined in this section, and their sizes were also predominantly phrases. This means that translation unit size was often kept the same through translation. Careful investigation of each

translation pair confirmed that most translation was indeed carried out at the same levels, and maintained the same structures and classes. As shown in Table 4.14, among all the 13 translation pairs, 10 of them did not go through any shifts: *a-f*; *h*; and *l-m* (shown in bold in Table 4.14). On the other hand, four of them required one, or more than one, shift in translation: *g* and *i-k*.

Table 4.14. Unit-shifts, structure-shifts, and class-shifts (i)

	Translation units	Unit	Structure	Class
a.	**market sources**	not shifted	not shifted	not shifted
b.	**market rumours**	not shifted	not shifted	not shifted
c.	**market economy**	not shifted	not shifted	not shifted
d.	**stock market**	not shifted	not shifted	not shifted
e.	**gold market**	not shifted	not shifted	not shifted
f.	**copper market**	not shifted	not shifted	not shifted
g.	*market interest rates*	not shifted	shifted	not shifted
h.	**Seoul stock market**	not shifted	not shifted	not shifted
i.	*domestic gold market*	not shifted	shifted	not shifted
j.	*market will reopen*	not shifted	shifted	-
k.	*a year ago*	shifted	shifted	not shifted
l.	**marketing year**	not shifted	not shifted	not shifted
m.	**government**	not shifted	not shifted	not shifted

Why did some translation units go through shifts and some not? Table 4.14 suggests that the occurrence of a shift is related to the grammatical structure of the translation unit. If one looks at all the translation units of N + N structures (Table 4.14 *a-f* and *l*), they were always rendered without shifts. Similarly, the shifts were not seen in the single word translation unit (Table 4.14 *m*) either. On the other hand, the translation units of three-word phrases (Table 4.14 *g-i*) tended to go through structure shifts. *Seoul stock market* (Table 4.14 *h*) was exceptional; a shift was not seen in this pair. This could be because the translation unit contained the proper name *Seoul*; however, this would require further research to prove. Even so, the important finding of the investigation of shifts is that the grammatical structure of the translation unit is closely related to the

occurrence of shifts. If a given translation unit is a N + N sequence or a single word translation unit, the unit is not likely to require any shifts in translation. If a given translation unit has any other grammatical structure, the unit is likely to require shifts.

4.3.   Variations of translation units

In the previous section (4.2), translation equivalence was examined from the perspective of shifts. Which types of translation units are likely to change their sizes, grammatical structures, and classes through translation was investigated. In this section, the focus will be on the variations of translation units. Variations of the translation units can occur by: (a) different article usages, e.g. *market economy*, *the market economy*, and *a market economy*; (b) singular and plural forms, e.g. *government* and *governments*; and, (c) modifier insertions, e.g. *gold market* and *gold futures market* (the definition of modifier is clarified in 4.3.3). The question arising here is how these variations affect translation equivalence, i.e. whether these variations caused different translations or not. The variations (a), (b), and (c) will be discussed in 4.3.1, 4.3.2, and 4.3.3-4.3.4, respectively.

4.3.1.   Articles

Definite and indefinite articles are not usually translated into Japanese; this most comprehensive bilingual dictionaries state (**English-Japanese Dictionary For The General Reader**, 1999; **Genius English-Japanese Dictionary**, 2001; **Lighthouse English-Japanese Dictionary**, 2007). The data of *gold market* agrees with this assertion (Example 4.5; the words in parentheses are the translations of the underlined

120

items; words in bold correspond to target words). *A gold market* in Lines 1-2 was

rendered into *kin shijoo* 'gold market'; *the gold market* in Line 3-4 was also rendered

into *kin shijoo*, so was *gold market* alone in Line 5. Despite such a variety of article

usage, all of them were rendered into *kin shijoo*. Article variations did not cause

different translations.

Example 4.5. Article variations: *gold market*

```
1  … taken steps to create a gold market by freeing the prices …    (kin shijoo)
2  … process of setting up a gold market .                          (kin shijoo)
3                    "The gold market was showing little …          (kin shijoo)
4    … for now, and that the gold market had managed to absorb …    (kin shijoo)
5 … profits," WGC manager of gold market analyst George Milling-…   (kin shijoo)
```

There were four more translation units which gave similar results to *gold market*

(Example 4.6 and Example 4.7). First, *market rumours* occurred without articles

(Example 4.6 Lines 1-3) and with the definite article (Line 4); both *market rumours*

alone and *the market rumours* were rendered into *shijoo no uwasa*. No item denoting

the meaning of *the* was found in the translation; the translation of Line 4 was *shijoo no*

uwasa ya shinbun hoodoo wa jujitsu de wa nai 'market rumour and newspaper report

were not true', although the original clause was *the market rumours and the*

newspaper's reports were untrue. The definite article was not rendered into Japanese in

the case of *market rumours* either.

Example 4.6. Article variations: *market rumours, market economy,* and *domestic gold market*

```
 1   … Ministry, amid market rumours of an impending …      (shijoo no uwasa)
 2   … Thursday denied market rumours that the authorities … (shijoo no uwasa)
 3 … no foundation to market rumours the bank would …       (shijoo no uwasa)
 4 … Thailand that the market rumours and the newspapers …  (shijoo no uwasa)
 5   … to introduce a market economy, including currency …  (shijoo keezai)
 6   … pattern of the market economy," it said.             (shijoo keezai)
 7   … unsuited to the market economy, he said.             (shijoo keezai)
 8   … and to create a domestic gold market.                (kokunai kin shijoo)
 9   … up to create a domestic gold market via measures …   (kokunai kin shijoo)
10 … to liberalise the domestic gold market to channel …   (kokunai kin shijoo)
```

Similarly, the translation unit *market economy* occurred with the indefinite article (Example 4.6 Line 5) and with the definite article (Lines 6-7); despite these different article usages, both *a market economy* and *the market economy* were rendered into *shijoo keezai*. For example, the translated clause of Line 7 was *keeee mo shijoo keezai ni soguwanai* 'management is not suited to market economy either', compared with the original clause, *their management was unsuited to the market economy*; the meaning of *the* was missing in translation. The same could be seen in Lines 8-10. *Domestic gold market* appeared with the indefinite article in Lines 8-9 and with the definite article in Line 10; both *a domestic gold market* and *the domestic gold market* were rendered into *kokunai kin shijoo*. The article variations did not cause different translations. No matter whether these translation units appeared with: (a) definite articles; (b) indefinite articles; or (c) no article; their translation equivalents were the same. This happened even in the case of the single word translation unit *government* as well. *Government* occurred with an indefinite article (Line 1, Example 4.7), with a definite article (Lines 2-4), and without articles (Lines 5-7). All of these were rendered into *seefu*. The translation equivalence between *government* and *seefu* was not affected by article variation.

122

Example 4.7. Article variations: *government*

```
1          "Obviously a government which sees there is a …       (seefu)
2      … Kyong-shik said the government has no plans to intervene …   (seefu)
3      … as it's a signal the government wants the market to be …     (seefu)
4 … see the likelihood of the government issuing (baht) bonds just yet. (seefu)
5  Instead, he suggested that government make more efficient use …    (seefu)
6      … the realisation that government was working on the accounts …  (seefu)
7         … showed signs that government would like the market …       (seefu)
```

The other nine translation units had only one article variation each in the sample set (Example 4.8). *Market sources* and *market interest rates* always occurred without articles (Line 1-2). There were no examples of other article usages. Similarly, *stock market*, *copper market*, *Seoul stock market*, *market will reopen*, and *marketing year* always appeared with the definite article *the* (Lines 3-7); while, the translation unit *a year ago* only occurred with the indefinite article *a* (Line 8). Do these cases also indicate that definite and indefinite articles are not rendered into Japanese? The concordance lines were examined closely to answer this question.

Example 4.8. Article variations: others

```
1                           Market sources have anticipated lower prices …
2 … such as in the case where market interest rates were pushed higher by…
3                        "The stock market is up and that's probably …
4 … he continues to watch the copper market, noting that trading during …
5   … would be listed on the Seoul stock market in the first half of 1997.
6                       The market will reopen on Monday.
7                       The marketing year will end Aug 30.
8      … 21, up 67 percent on a year ago", the State Statistics Committee …
```

All the translation units which occurred only with the definite article showed that *the* was not rendered into Japanese at all. For example, the original clause of Line 3 (Example 4.8) was *The stock market is up*, which was translated into *kabushiki shijoo ga jyooshoo* 'stock market is up'; the meaning of the definite article was lost in

123

translation. The original clause of Line 4, *to watch the copper market*, corresponded to the nominal phrase *doo shijoo no kanshi* 'monitor of copper market'; again, *the* was not seen in translation. Similarly, the original phrase of Line 5, *on the Seoul stock market*, was rendered into *sooru kabushiki shijoo ni* 'on Seoul stock market'; the original clause of Line 6, *The market will reopen on Monday*, was translated into *torihiki wa 16 nichi ni saikai sareru* 'market is reopen on 16th'; the original sentence of Line 7, *The marketing year will end Aug 30*, was translated into *shijoo nendo wa 8 gatsu 30 nichi made* 'marketing year until 30th August'. Such zero translation of definite articles was seen in all the concordance lines of these five translation units: *stock market*, *copper mark*, *Seoul stock market*, *market will reopen*, and *marketing year*.

On the other hand, the translation unit which occurred only with indefinite articles, *a year ago* (Lines 8, Example 4.8), gave different findings. The original phrase in Line 8, *up 67 percent on a year ago*, was rendered into *zennen wo 67% uwamawa tte iru* 'is exceeding 67% on previous year'. The indefinite article was rendered implicitly in *zennen* 'previous year'; *zennen* refers to one single year which has just passed. This was the only translation unit which indicated that articles can be rendered into Japanese.

4.3.2.   Singular and plural forms

The second variation I investigated was the singular and plural variations of translation units. Plurality is normally expressed by *–s* in English; therefore, recognising which translation units are singular or plural is rather obvious. Among all the 13 identified translation units (Table 4.15), 10 translation units were singular forms and three were

plural forms. On the other hand, if one looks at the translation equivalents in Table 4.15, one cannot find any lexical or morphological markers of plurality and singularity; none of the translation equivalents express clearly whether they are singular or plural. This, of course, is due to the different plurality systems between the two languages.

Table 4.15. Translation units of nouns

	Translation units	Translation equivalents
a.	**market** sources	**shijoo** suji
b.	**market** rumours	**shijoo** no uwasa
c.	**market** economy	**shijoo** keezai
d.	Stock **market**	kabushiki **shijoo**
e.	gold **market**	kin **shijoo**
f.	copper **market**	doo **shijoo**
g.	**market** interest rates	**shijoo** kinri
h.	Seoul stock **market**	sooru kabushiki **shijoo**
i.	domestic gold **market**	kokunai kin **shijoo**
j.	**market** will reopen	**torihiki** wa saikai sa reru
k.	a **year ago**	**zennen**
l.	marketing **year**	**shijoo nendo**
m.	**government**	**seefu**

Japanese nouns, according to Baker (1992), do 'not normally indicate whether [they are] singular or plural'. Bunt (2003) also agrees with this point, maintaining that 'Japanese does not generally have a plural form for nouns'. Therefore, both *market* and *markets* can be *shijoo* in Japanese. Only when plurality is necessary for communication is it possible to express it lexically by adding words such as *ooku no* 'many', e.g. *ooku no shijoo*; but this is rare. Taking the first 100 examples of *markets* in the ARC, none of them had such lexical markers to express plurality. The 'apparent lack of interest in the difference between one and more than one' in Japanese nouns (Kindaichi, 1988; Baker, 1992) triggered this investigation. Does this mean that a singular translation unit (e.g. *market economy*) and its plural form (e.g. *market economies*) share the same translation

equivalent *shijoo keezai*? I looked up all the translation units in the ARC to investigate if this was the case or not.

It was possible to examine six pairs (Table 4.15 *a-d, j-k*); while, the rest of the translation pairs (Table 4.15 *e-i*, *l-m*) did not have singular and plural variations in the ARC. As far as the former translation pairs were concerned, singular and plural variations did not cause different translations in most of the cases. For example, the translation unit *market economy* was a singular form and always rendered into *shijoo keezai* in the sample set (in Lines 1-3, Example 4.9). The plural form *market economies* was found in the ARC (in Line 4, Example 4.9); it was also rendered into *shijoo keezai*. This indicated that, despite the differences of singular and plural forms, *market economy* and *market economies* share the same translation equivalent, *shijoo keezai*. The plural variations did not cause different translations of *market economy*.

Example 4.9. Singular and plural variations: *market economy*

```
1 … plans to introduce a market economy, including currency …   (shijoo keezai)
2… to the pattern of the market economy," it said.              (shijoo keezai)
3  … was unsuited to the market economy, he said.               (shijoo keezai)
4   … property rights of market economies and legal certainties… (shijoo keezai)
```

Similarly, as shown in Example 4.10 and Example 4.11, the translation units *market will reopen* and *market rumours* showed that singular and plural variations did not cause different translations either. *Market will reopen* and its plural form *markets will reopen* were rendered into the same translation equivalent *torihiki wa saikai sa reru* (Example 4.10); the translation unit *market rumours* and its singular form *market rumour* shared the same translation equivalent *shijoo no uwasa* as well.

Example 4.10. Singular and plural variations: *market will reopen*

```
1 The market will reopen for trade on Friday.          (torihiki wa saikai sa reru)
2 The market will reopen on February 11.               (torihiki wa saikai sa reru)
3 The market will reopen on Friday.                    (torihiki wa saikai sa reru)
4 The market will reopen on Monday.                    (torihiki wa saikai sa reru)
5 The market will reopen with normal trading hours on …(torihiki wa saikai sa reru)
6 The Markets will reopen on Tuesday April 1 at …      (torihiki wa saikai sa reru)
7 The Markets will reopen on Wednesday.                (torihiki wa saikai sa reru)
```

Example 4.11. Singular and plural variations: *market rumours*

```
1    … no foundation to market rumours the bank would …    (shijoo no uwasa)
2    … Thailand that the market rumours and the newspapers … (shijoo no uwasa)
3    … firmly denied a market rumour that Germany was about … (shijoo no uwasa)
4    … Thursday denied a market rumour that Lionel Jospin …  (shijoo no uwasa)
```

The translation unit *stock market* and its plural form *stock markets* did not disagree with the above findings. As shown in Lines 1-5, both *stock market* and *stock markets* were translated into *kabushiki shijoo* (Example 4.12). The plural variation *stock markets* in fact occurred rather often in the ARC: 36 lines. Most of them were translated into *kabushiki shijoo* (Lines 3-5), appearing in 28 out of the 36 lines (78 percent); this supported the finding that the plural variation did not cause different translations. It is fair to mention that there were, however, some other translations of *stock markets* in the corpus, e.g. *kabushiki sooba* in Line 6, *shijoo* in Line 7, and *kanetsu sooba* in Line 8.

127

Did they demonstrate any counter-examples? The answer was no. This was because none of these minor translations expressed the plurality of *stock markets*: *kabushiki sooba* 'stock price' in Line 6, *shijoo* 'market' in Line 7, and *kanetsu sooba* 'red-hot price' in Line 8. The finding from the analysis of *stock market* and its plural form *stock markets*, therefore, was that the plurality was not rendered into Japanese; the translation unit *stock market* and its plural form *stock markets* usually shared the same translation.

Example 4.12. Singular and plural variations: *stock market*

```
1              "The stock market is up and that's …       (kabushiki shijoo)
2   … more money from the stock market, it said.         (kabushiki shijoo)
3     … seemed limited to stock markets only.            (kabushiki shijoo)
4    … illegally into the stock markets, seriously damaging …  (kabushiki shijoo)
5     … government on the stock markets, traders said.   (kabushiki shijoo)
6      … to cool down the stock markets, has greatly weakened … (kabushiki sooba)
7   … money supply on the stock markets, in particular the …   (shijoo)
8          … to curb the stock markets, which resulted in …    (kanetsu sooba)
```

On the other hand, there were three translation units which demonstrated 'authentic' counter-examples for this investigation. First, the translation unit *a year ago* and its plural form *years ago* did not share the same translation equivalent (Example 4.13). *A year ago* was rendered into the single word *zennen* in all of the 1,000 samples (Lines 1-3). The plural variation *years ago*, on the other hand, occurred in 54 lines in the ARC and was predominantly rendered into *nen mae* (Lines 4-6). Plurality clearly made a difference for translation in the case of *a year ago*. This is reasonable, if one thinks that *zennen* refers to one year which has just passed (discussed in 4.3.2). *Zennen* denotes a singularity; therefore, it could not be shared by the plural variation *years ago*.

128

Example 4.13. Singular and plural variations: *a year ago*

```
1    … but behind 37 percent  a year ago and also behind 33 …      (zennen)
2  … substantially lower than  a year ago.                         (zennen)
3    … from 19,288 short tons  a year ago.                         (zennen)
4     … about 10 percent five  years ago, Ma said.                 (nen mae)
5      … trees surfaced three  years ago, but a revolt in the …    (nen mae)
6   … more prospect than two   years ago, but it is clearly …      (nen mae)
```

Also, the translation unit *market sources* and its singular form *market source* did not share the same translation equivalent. This time, it was singularity which made the difference. *Market sources* occurred 12 times in the sample set; these were all rendered into *shijoo suji* (Lines 1-5, Example 4.14). For the singular form *market source*, there were two relevant examples in the ARC (Lines 6-7). In both examples, *market source* occurred with the indefinite article; *a market source* was rendered into *aru shijoo suji* 'a certain market source'. *Market sources* and *market source* did not share the same translation equivalent. The translation equivalence between *market sources* and *shijoo suji* was not resistant to a change in number.

Example 4.14. Singular and plural variations: *market sources*

```
1                         Market sources expect continued …   (shijoo suji)
2                         Market sources also said AIG …        (shijoo suji)
3                         Market sources said early gains …    (shijoo suji)
4         … unclear but   market sources said an apparent …    (shijoo suji)
5   … Gateway 2000 Inc,   market sources said.                 (shijoo suji)
6  operational funds, a   market source said.                  (aru shijoo suji)
7 … not run properly, a   market source said.                  (aru shijoo suji)
```

The translations of indefinite articles in the case of *a market source* are interesting if one compares them with the ones of *a market rumour* (Lines 7-8, Example 4.11). The singular form *market rumour* always appeared with indefinite articles, but *a* was not rendered into Japanese. On the contrary, when the singular form *market source* appeared with the indefinite article, *a* was lexically rendered into *aru* in Japanese. Why were

these indefinite articles translated differently? I looked up how often indefinite articles were actually translated into Japanese and how often they were omitted. In order to obtain a small sample, I used the 1,000 lines of *market*; there were 505 examples of the indefinite article *a* found (the head nouns were not necessarily *market*, e.g. *a cotton marketing specialist*; *a weekly market comment*; *a local broker*). Taking the first 100 examples of *a*, three types of translation of *a* were recognised (Table 4.16).

Table 4.16. *Market* and the indefinite article, *a*

	Translations	Freq.
1	zero correspondences	86
2	*Aru*	9
3	other translations	5
	total	100

The most dominant translation was zero correspondence: *a* was not rendered into Japanese. This occurred 86 out of 100 times. This agrees with the assertion that *a* is hardly ever translated, as the dictionaries (**English-Japanese Dictionary For The General Reader**, 1999; **Genius English-Japanese Dictionary**, 2001; **Lighthouse English-Japanese Dictionary**, 2007) suggest. The examples of *a market rumour* (Line 3-4, Example 4.11) belong to this group. The second most dominant translation was *aru*, appearing in only 9 examples. The examples of *a market source* (Line 6-7, Example 4.14) belong to this group. The other translations entail the examples in which indefinite articles were used in idiomatic phrases, such as *a certain degree of*, *a few times*, *a little bit of*, and *500 tonnes a year*. In order to clarify the cases of *a market source* and *a market rumour*, the dominant two types of indefinite article translation were used for a closer look at the concordance lines (Example 4.15).

Example 4.15. *Market* **and indefinite article** *a*

```
 1  … of checking rates," said a dealer with a European bank.          (aru)
 2    … death in stride," said a dealer with a local brokerage house.  (aru)
 3   … promising for futures," a European commission house dealer …    (aru)
 4     … of fresh incentives," a local broker said.                    (aru)
 5   … market is so depressed," a local rubber trader said.            (aru)
 6  … no news to support gold," a senior dealer at a local …           (aru)
 7 … reached its bottom," said a Seoul Securities broker.              (aru)
 8        … like these prices," a trader for a commercial firm said.   (aru)
 9  … the futures higher," said a trader with a Japanese institution.  (aru)
10                                  A black market in cotton has …     (-)
11  … might also regard this as a buying opportunity too and so the    (-)
12    … the office would serve a critical market as Japan is a  …      (-)
13  … domestic consumption and a fall in production in India.          (-)
14  … sugar market experienced a large move over 10 years and in …     (-)
15 …  Wheat Associates said in a market letter.                        (-)
16     … & Wireless on "taking a new direction to enter the German …   (-)
17  … base prices, suggesting a property market recovery.              (-)
18       … since 1990 following a rise in domestic consumption and …   (-)
19    … A&M University, said in a weekly market comment.                (-)
```

Lines 1-9 are all the nine examples in which the indefinite article was rendered into *aru* in the 100 lines. If one looks at the head nouns after the indefinite articles, they always refer to people, such as *dealer*, *broker*, and *trader*. On the other hand, the nouns of the zero correspondences are different, as shown in the sub-sample of Lines 10-19: no nouns denote people. The same pattern was observed in all of the 86 examples. This suggests why the case of *a market rumour* belongs to the zero correspondence translation, and the case of *a market source* belongs to the *aru* translation. The former has *rumour* as a head noun, which does not denote people. The latter has *source* as a noun, which here denotes people; 'the **source** of something is the person, place, or thing which you get it from [original emphasis]' (**Collins Cobuild Advanced Learner's English Dictionary** 2003). Due to the noun *source* denoting people, the indefinite article in the unit *a market source* was lexically rendered into *aru*.

131

Lastly, the case of the single word translation unit *government* showed a mixture of evidence. There were 28 lines of its plural form *governments* in the ARC; there were two translations of *governments* identified, which were equally dominant. One was *seefu*, appearing in 11 lines (Lines 1-5, Examples 4.16). As the translation unit *government* always corresponded to *seefu*, these examples indicate that *government* and *governments* shared the same translation equivalent.

Examples 4.16. Singular and plural variations: *government*

```
 1      … cooperation between governments, banks, brokerage …    (seefu)
 2  … countries since 1990 as governments have opened mining …   (seefu)
 3           … it will be the governments and political powers … (seefu)
 4     … that the capacity of governments to influence …        (seefu)
 5 … deficit, but warned that governments must continue to …     (seefu)
 6     … in the short term as governments cut costs to meet …   (kakukoku seefu)
 7    … will be important for governments to give directly to … (kakukoku seefu)
 8        … by participating governments using them to reduce … (kakukoku seefu)
 9       … appealed again to governments to give food aid …     (kakukoku seefu)
10      … version, which said governments should continue to …  (kakukoku seefu)
```

However, the other dominant translation of *governments* showed a contradictory finding (Lines 6-10, Examples 4.16). The plurality of *governments* was rendered lexically into Japanese: *kakukoku seefu*. This translation occurred in 11 lines. *Kakukoku* literally means "each country"; therefore, by translating *governments* into *kakukoku seefu* 'each country's government', the translator achieved the plurality expressed in translation. In these cases, the plurality was kept in the target text; therefore, it caused different translations from the one of the singular form *government*. The single-word translation unit *government* and its plural variation *governments* did not share the same translation equivalent in these cases. Due to the space limitation, the investigation of the mysteries

132

of *kakukoku* – why *kakukoku* appeared in Lines 6-10 and why it did not in Lines 1-5 – was not pursued further in this study.

In most cases singular and plural variations did not cause different translations. The exceptions were the three translation units: *market sources*, *a year ago*, and *government*. The singular variation *a market source* was translated differently from its plural unit *market sources*; this was caused by the head noun *source* denoting people. The plural variation *years ago* was translated differently from the singular unit *a year ago*. The translation equivalent of *a year ago* was *zennen*, which indicates singularity explicitly; therefore, it was not possible for *zennen* to be shared by both *years ago* and *a year ago*. Lastly, the plural variation *governments* was sometimes translated differently from the singular unit *government*. This happened because the plurality of *governments* was lexically rendered in some cases: *kakukoku seefu* 'government of each country' instead of just *seefu* 'government'.

4.3.3. Modifiers (i)

The third variation I investigated was of modifier variations of translation units. A modifier, as discussed here, refers to 'a linguistic element in a sentence which is grammatically linked to a second element and adds information about that second element'; for example, *girl* has two modifiers, *little* and *in the pond*, in the sentence, *The little girl in the pond was shrieking delightedly* (Trask, 1998). Only modifiers appearing in the immediate context of a given translation unit are considered; modifiers appearing out of the immediate context to a given translation unit are not considered, e.g. the

133

modifier *giggling* of the head noun *women* in Trask's example sentence *The women who had volunteered for the wet T-shirt contest climbed giggling onto the stage*. This is because the main focus of this study is how the immediate context affects the rendering of a given translation.

The translation unit *market economy*, for example, was regarded as a translation unit and its translation equivalent was *shijoo keezai*. So, what happens if *market economy* occurs in phrases with other words, e.g. *free market economy* and *market sector economy*? How are the translations of *market economy* affected by having such modifiers? First, I examined the modifier variations in which an intervening word appeared in the middle of a translation unit. The whole of the ARC was consulted to find such variations, however, not many were found. Only two pairs were extracted: (a) *gold market* and its modifier variation *gold futures market* shown in Example 4.17; and (b) *market economy* and its variation *market sector economy* shown in Example 4.18.

Example 4.17. Modifier variations: *gold market*

```
1  … , and that the gold market had managed to …      (kin shijoo)
2      … creating a gold futures market but is …      (kin no sakimono torihiki)
3  …is to launch a gold futures market on Friday, …   (kin sakimono)
```

First, the translation unit *gold market* had *kin shijoo* as its translation equivalent (Line 1 Example 4.17). The modifier variation *gold futures market* occurred in the ARC twice. In one example, it was translated into *kin no sakimono torihiki* 'future trading of gold', as shown in Line 2. By having the modifier *futures*, the items *gold* and *market* were not rendered into *kin* and *shijoo* any longer; *gold* was rendered into *kin no* 'of gold' and

market into *torihiki* 'trading'. This indicates that one intervening modifier caused the different translations of *gold* and *market*. Line 2 showed that the translation pair of *gold market* and *kin shijoo* was not resistant to this modifier variation.

In Line 3, *gold futures market* was rendered differently: *kin sakimono* 'gold futures'. The equivalent of the word *market* was omitted in translation; there was no element denoting the meaning of *market*. Was this omission triggered by the modifier *futures*? The wider context suggested that this was not the case. The original sentence was: *The Istanbul gold exchange is to launch a gold futures market on Friday* (rendered into *Toruko no isutanbuuru kin torihiki joo wa, 15 nichi ni kin sakimono o joojoo suru*). Due to the item *gold exchange* at the beginning of the sentence, the translation was clear even though *market* in *a gold futures market* was omitted. Since this could be the main reason for the zero translation of *market*, this example was not suitable for the investigation of modifiers. Line 3 would give little evidence, whether the existence of the modifier *futures* affected the translations of *gold* and *market* or not.

Example 4.18. Modifier variations: *market economy* (i)

```
1 … unsuited to the market economy, he said.        (shijoo keezai)
2     … is that the market sector economy is too …   (shijoo keezai)
```

Second, the other translation unit *market economy* had *shijoo keezai* as its translation equivalent (Line 4 Example 4.18); the modifier variation *market sector economy* was also rendered into *shijoo keezai* (Line 5). The meaning of *sector* was omitted in this translation; therefore, the intervening word *sector* did not cause any change to the translation of *market* and *economy*. The possible reason behind this was the unusual

three-word phrase, *market sector economy*. If one searches this item in Google, there were only nine hits, as shown in Example 4.19 (duplications were deleted). Among them, there were two lines in which *market sector economy* was used as a unit (Lines 1-2). In the rest of the examples (Lines 3-6), the three words were separated by commas, forward slashes, and periods. This suggests that *market sector economy* is an unusual collocation which is hardly ever used. This infrequent usage may mean that *market sector economy* is a new or one-off item. It is likely that the translation of such an infrequent item is not yet established in the community. According to Baker (1992), a word of 'non-equivalence' is commonly rendered into 'a more general word', to which the case of *market sector economy* can be said to belong exactly: *market sector economy* was translated into the general word *shijoo keezai* 'market economy'.

Example 4.19. Concordances of *market sector economy* from Google

```
1 .. non-traded components of the market-sector economy are used. The market
2      … , who control the local market-sector economy, dealing in
3        , studying the company, market, sector, economy.
4 level(s) that a certain large market sector / economy can command / sustain
5   … approach (household/firm, market, sector/economy) to provide insights into
6 … blazed by the commercial or market sector. Economy and efficiency were not
```

Thus, the two translation units were examined. The first translation unit *gold market* and its modifier variation *gold futures market* showed that an intervening modifier mattered in translation. By having the modifier *futures*, the translation units were rendered differently. However, the second translation unit *market economy* and its modifier variation *market sector economy* indicated that the intervening modifier *sector* did not cause the different translation. This translation unit *market economy* was resistant to the modifier insertion.

136

4.3.4. Modifiers (ii)

So far, the investigation has focused on modifier variations in which a modifier appeared in the middle of a translation unit. However, it is possible for modifiers to appear before the translation unit. The ARC had five pairs of such modifier variations: (a) *market economy* and *free market economy*; (b) *stock market* and *red-hot stock market*; (c) *market interest rates* and *capital market interest rates*; (d) *copper market* and *world copper market*; and, (e) *market rumours* and *stock market rumours*. Infrequent modifiers occurring only once were not examined, due to the space limitations of the present work. Also, the most dominant modifier variations for each translation unit were investigated.

First, the pair of *market economy* and *free market economy* indicated that a modifier appearing before the translation unit did not cause different translations of *market* and *economy* (Example 4.20). As shown in Line 1, the translation unit *market economy* was always rendered into its translation equivalent *shijoo keezai*. The modifier variation *free market economy* occurred in the ARC twice (Lines 2-3). In both examples, *free market economy* was rendered into the same translation equivalent: *jiyuu shijoo keezai*. Even though the modifier *free* appeared, *market* and *economy* were still translated into *shijoo* and *keezai*. Therefore, the translation pair of *market economy* and *shijoo keezai* was resistant to the modifier *free*. One might wonder if this modifier variation *free market economy* can be a translation unit or not. As it was always rendered into *jiyuu shijoo keezai*, it could have been regarded as a translation unit under the one-equivalent principle only if it occurred more than three times in the sample.

137

Example 4.20. Modifier variations: *market economy* (ii)

```
1  …to introduce a market economy, including currency …   (shijoo keezai)
2  …… , said a free market economy expanded production …  (jiyuu shijoo keezai)
3      … in a free market economy, Schnull said.          (jiyuu shijoo keezai)
```

Second, the translation unit *stock market* and its modifier variation *red-hot stock market* also showed that a modifier before the translation unit did not cause a different translation (Example 4.21). *Stock market* was recognised as a translation unit and its translation equivalent was *kabushiki shijoo* in the sample set (Line 1). The modifier variation *red-hot stock market* was found in three lines in the ARC (Lines 2-4). Lines 2-3 indicate that *stock* and *market* were still rendered into *kabushiki* and *shijoo*, although the modifier *red-hot* appeared. Line 4 was not valid for the investigation as it was an example of zero correspondence. The whole clause including *red-hot stock market* was omitted (the original sentence was: *Late profit-taking on electronics eroded early gains and sent Taiwan share prices to a lower close on Thursday amid caution after the government's persistent moves to cool the red-hot stock market, brokers said*; the whole dependent clause beginning with *after* was not translated into Japanese).

Example 4.21 Modifier variations: *stock market*

```
1  … money from the stock market, it said.              (kabushiki shijoo)
2  … down the red-hot stock market, though analysts …   (kabushiki shijoo no kanetsu)
3  … cool the red-hot stock market," said HSBC James …  (kabushiki shijoo no kanetsu)
4  … cool the red-hot stock market, brokers said.       (-)
```

Similarly, the third pair *market interest rates* and *capital market interest rates* agrees with the findings above; the modifier variations did not cause different translations. As shown in Example 4.22, the translation unit *market interest rates* corresponded to its translation equivalent *shijoo kinri* (Line 1). The modifier variation *capital market*

138

interest rates appeared three times in the ARC (Lines 2-4). In all the examples, *market interest rates* was always rendered into *shijoo kinri*, although the translations of the modifier *capital* differed between the three examples. *Capital* was rendered into *shihon* 'capital' in Line 2, into *chooki* 'long-term' in Line 3, and into zero correspondence in Line 4. No matter how the extra modifier *capital* was rendered, the translations of *market interest rates* remained the same. Therefore, the modifier *capital* did not cause different translations of *market interest rates*.

Example 4.22. Modifier variations: *market interest rates*

```
1                        Market interest rates are likely … (shijoo kinri)
2 … Monday that capital  market interest rates could ease … (shihon shijoo kinri)
3 … importantly, capital market interest rates could remain …(chooki shijoo kinri)
4    … lows in capital   market interest rates had raised … (shijoo kinri)
```

In the case of the single word translation unit *government*, it was not necessary to consult the modifier variations in the whole data-set of the ARC. This was because the 1,000 sample lines of *government* contained enough data. In the sample set, *government* appeared with various modifiers. When *government* was used with the indefinite article, adjectives such as *safe*, *secure*, and *exile* appeared before *government*. As shown in Example 4.23, no matter which adjective appeared, *government* was always rendered into *seefu*. These adjectival modifiers did not affect how the translation unit *government* was rendered.

Example 4.23. Modifier variations: *a government*

```
1          "Obviously a government which sees there is a      (seefu)
2    cotton and wheat since a government move to clamp down on  (seefu)
3  liquid market with a safe government -- safe in the sense    (seefu)
4         people for a secure government, economic stability and (seefu)
5    Ku, who headed an exile government that fought for Korean   (seefu)
```

Similarly, when *government* was used with the definite article, many different modifiers appeared before *government*. Most of them were adjectives denoting countries such as *German*, *Philippine*, *Russian*, and *Vietnamese* shown in Lines 4-7, Example 4.24. There were some other words such as *central*, *federal*, *latest*, *military*, and *present* presented in Lines 8-12. Regardless of whatever appeared before *government*, *government* was rendered into *seefu* in all lines, as shown in Example 4.24. Therefore, again, the translation pair of *government* and *seefu* was resistant to modifier variations.

Example 4.24. Modifier variations: *the government*

```
1       … Kyong-shik said the government has no plans to intervene …   (seefu)
2    … list' to help focus the government on key areas where …         (seefu)
3      … as it's a signal the government wants the market to be …       (seefu)
4                 The German government lowered the yield on several … (seefu)
5  … reporters the Philippine government had taken steps to assure …    (seefu)
6                The Russian government may sign next week a decree …   (seefu)
7    … from the Vietnamese government to form a joint venture …         (seefu)
8       … from the central government to set up a fully-owned …        (seefu)
9     …(PTU) and the federal government over the future of …           (seefu)
10  According to the latest government figures released at the …        (seefu)
11  … has urged the military government to release funds to …           (seefu)
12  … attempts by the present government to lower local interest ……     (seefu)
```

So far, all the pairs demonstrated that they were resistant to modifier variation; the translation equivalence of each pair was not broken down by having a modifier before the translation units. However, the next pair differed: the translation unit *copper market* and its modifier variation *world copper market* (Example 4.25). *Copper market* was

always rendered into its translation equivalent *doo shijoo* (Line 1). The modifier *world*

occurred before the translation unit twice in the ARC. In one example (Line 2), *world*

copper market was rendered into *sekai doo shijoo*; *copper* and *market* corresponded to

doo and *shijoo* respectively. The modifier in Line 2 did not cause different translations,

which echoes the findings above. In Line 3, on the other hand, *world copper market* was

rendered differently: *sekai no doo torihiki*. *Market* was not rendered into *shijoo* any

longer; instead, it was translated into *torihiki*. The modifier *world* in Line 3, therefore,

caused a different translation of *market*.

Example 4.25. Modifier variations: *copper market*

```
1         … problems in the copper market, but also …    (doo shijoo)
2   … manipulated the world copper market.               (sekai doo shijoo)
3   … (CFTC) about the world copper market even earlier … (sekai no doo torihiki)
```

The last pair *market rumours* and *stock market rumours* also showed a mixture of

evidence for modifier variation (Example 4.26). The translation unit *market rumours*

was always rendered into *shijoo no uwasa* (Line 1). The ARC found the modifier

variation *stock market rumours* in four lines (Lines 2-5). In Lines 2-3, *market rumours*

was still rendered into the same translation equivalent *shijoo no uwasa*. This meant that

the modifier *stock* did not have an impact on the translation of *market* and *rumours*. In

Line 4, however, it was rendered into a different translation: *shijoo de no uwasa*. The

difference between *shijoo no uwasa* (Lines 1-3) and *shijoo de no uwasa* (Line 4) comes

from the different usage of the function words. The former has *no* 'of'; therefore, it

literally means 'rumour of market'. On the other hand, the latter has *de no* 'in'; it,

therefore, literally means 'rumour in market'. The modifier *stock* in Line 4 made the translation of *market rumours* subtly different.

Example 4.26. Modifier variations: *market rumours*

```
1 … Thursday denied market rumours that the …   (shijoo no uwasa)
2 … has denied stock market rumours that it …   (kabushiki shijoo no uwasa)
3       … to stock market rumours that the …    (kabushiki shijoo no uwasa)
4    … that stock market rumours that some …     (kabushiki shijoo de no uwasa)
5           Stock market rumours of a merger … (kabushiki shijoo de wa … uwasa ga)
```

In Line 5, the translation of *stock market rumours* was very different from the others. *Stock market rumours* in Line 5 was rendered into two separate items: *kabushiki shijoo de wa* 'at stock market' and *uwasa ga* 'rumour'. The original sentence was: <u>*Stock market rumours* of a merger between Suria, which is controlled by the eastern Sabah state government, and MBf Capital in the last two months, have boosted the stock of both companies</u>; it was translated into *Mareeshia <u>kabushiki shijoo de wa</u>, saba shuu seefu sanka no suria to, MBf no gappee no <u>uwasa ga</u> 2 kagetsu mae kara hiroga tte i ta*, 'At the Malaysian <u>stock market</u>, <u>rumour</u> of merger between Suria which is controlled by the eastern Sabah state government and MBF was spread in the last two months'. By having this modifier *stock*, *market* and *rumours* were translated into two items which were located separately in the translation. The translation pair of *market rumours* and *shijoo no uwasa* was not resistant to the modifier insertion in Line 5.

4.3.5. Conclusion

The variations of translation units related to nouns were examined in this section (4.3): the article, singular/plural, and modifier variations were all covered. Most of the

translation units had some variations which did not cause different translations. Table 4.17 is a summary of such variations for each translation unit; the numbers of the variations are shown in parentheses; '~' is used where one modifier is inserted.

Table 4.17. Variations of translation units

	Translation units	Variations	
a.	*government*	*government, a government, the government, governments, a ~ government, the ~ government*	(6)
b.	*market economy*	*a market economy, the market economy, market economies, market ~ economy, a ~ market economy*	(5)
c.	*stock market*	*the stock market, stock markets, the stock markets, the ~ stock market*	(4)
d.	*gold market*	*gold market, a gold market, the gold market, gold ~ market*	(4)
e.	*market rumours*	*the market rumours, a market rumour, the ~ market rumours*	(3)
f.	*copper market*	*the copper market, the ~ copper market*	(2)
g.	*market will reopen*	*the market will reopen, the markets will reopen*	(2)
h.	*market interest rates*	*market interest rates, ~ market interest rates*	(2)
i.	*domestic gold market*	*a domestic gold market, the domestic gold market*	(2)
j.	*Seoul stock market*	*the Seoul stock market*	(1)
k.	*marketing year*	*the marketing year*	(1)
l.	*market sources*	*market sources*	(1)
m.	*a year ago*	*a year ago*	(1)

The single word translation unit *government* had the most variations: the article variations, e.g. *government, a government, the government*, the plural variation, e.g. *governments*, and the modifier ones, e.g. *a* + modifier + *government* and *the* + modifier + *government*. All the six variations shared one translation equivalent: *seefu*. Translation equivalence was established between six translation units and one translation equivalent (Figure 4.1). On the other hand, the translation unit *market sources* had only one variation: *market sources*. The translation equivalence was established between one translation unit and one equivalent (Figure 4.1). Since the singular variation *a market source* caused different translations, this variation could not join the translation pair.

143

Such one-to-one translation equivalence was seen in the other three translation units: *Seoul stock market*, *marketing year*, and *a year ago*.

Figure 4.1. Translation equivalence: *government* and *market sources*
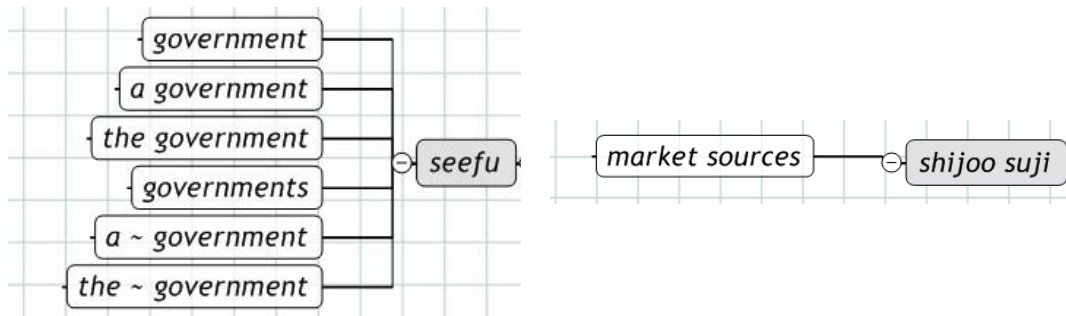
Figure 4.2 shows the three types of translation units according to the translation equivalence; the numbers in parentheses mean the numbers of variations of translation units. The translation units of the one-to-one translation equivalence are located at the left end: *Seoul stock market*, *marketing year*, *market sources*, and *a year ago*. The translation units of the many-to-one translation equivalence are located at the right end: *government*, *market economy*, *stock market*, and *gold market*. The other five translation units are listed in the middle. Their translation equivalence is established between two or three translation units and one equivalent: *market rumours*, *copper market*, *market will reopen*, *market interest rates*, and *domestic gold market*.

Figure 4.2. Translation equivalence: from one-to-one to many-to-one (i)

One-to-one   Many-to-one

⟵──⟶

Seoul stock market (1)	*market rumours* (3)	*government* (6)
marketing year (1)	*copper market* (2)	*market economy* (5)
market sources (1)	*market will reopen* (2)	*stock market* (4)
a year ago (1)	*market interest rates* (2)	*gold market* (4)
	domestic gold market (2)	

It is hard to discover the distinctive characteristics of each group. For example, all groups have two-word translation units: e.g. *marketing year* and *market sources* having the one-to-one translation equivalence; *market rumours* and *copper market* holding the two/three-to-one translation equivalence; and, *market economy*, *stock market*, and *gold market* establishing four/five/six-to-one translation equivalence. Similarly, all the three-word translation units spread out into two groups: the one-to-one and two/three-to-one translation equivalences. Both clausal and single word translation units occurred only once; therefore, one cannot claim from Figure 4.2 that (a) clausal translation units tend to have two/three-to-one translation equivalences, or that (b) single word translation units tend to have many-to-one translation equivalence. This indicates that the discussion will be more productive when all the translation units of the nine target words are identified.

4.4. Chapter summary

This chapter has presented the analyses of the three frequent nouns: *market*, *year*, and *government*. First, the one-equivalence principle identified 10 translation units of *market*, two translation units of *year*, and one translation unit of *government*. The translation units were not restricted to be at one size; rather, the sizes were scattered.

The most dominant translation unit size was the phrase; the least dominant one was the clause. There were some translation units at the level of the single word and clause; however, these were a minority. Second, these findings of translation unit sizes were compared with the translation equivalent sizes from the perspectives of shifts such as unit-shift, structure-shift, and class-shift. The investigation found out which translation units are likely to go through shifts in translation and which ones are not. Lastly, the four variations of translation units were examined. Some translation pairs were found to have translation equivalence between many translation units and one equivalent (i.e. many-to-one translation equivalence); some were identified to have translation equivalence between one translation unit and one equivalent (i.e. one-to-one translation equivalence).

5. Analyses of adjectives

Another set of translation units were identified: the translation units of the three frequent adjectives, *economic*, *new*, and *foreign*. The analyses will be demonstrated in detail in this chapter. First, the translation units and their sizes will be presented in 5.1. Second, the translation unit sizes will be compared with their equivalent sizes, followed by discussion of shifts between translation units and equivalents in 5.2. Lastly, the resistance of these translation units to variations will be discussed in 5.3.

5.1. Translation units

5.1.1. Word level

Among the three frequent adjectives, *economic* was the only word which was regarded as a single word translation unit under the one-equivalent principle. In the 1,000 sample set, *economic* was rendered into 9 Japanese translations. The most frequent five of them are listed in Table 5.1.

Table 5.1. *Economic*

	Translations		P.O.S.	Freq.	(%)
1	*keezai*	(economy/economic)	noun	855	(86%)
2	*keeki*	(business/boom)	noun	97	(10%)
3	*keezai teki*	(economical)	adjective	37	(4%)
4	*keezai teki ni*	(economically)	adverb	5	(1%)
5	*kin'yuu*	(finance/market)	noun	1	(0%)
			Total	995	(100%)

The most frequent translation was *keezai* and appeared in 855 out of the 1,000 examples of *economic*, which was obviously very dominant. The second most frequent translation was *keeki* (occurring in 97 lines), the third was *keezai teki* (occurring in 37 lines). The rest of the translations were very minor, occurring only five times or less in the sample set. Based on the one-equivalent principle, the single word *economic* was, thus, regarded as a translation unit, since it had one dominant translation which took up 85 percent of the total examples.

However, the other two adjectives *new* and *foreign* were not regarded as single word translation units. First, *new* was rendered into 35 Japanese translations in the sample data. As shown in Table 5.2, the most frequent translation was the Japanese prefix, *shin*; it, however, occurred only in 291 lines of the total 1,000 examples. Compared with *economic*, the most frequent translation of which appeared in 855 out of 1,000 lines (Table 5.1), *shin* is far less frequent. The one-equivalent principle could not take *new* as a single word translation unit.

Table 5.2. *New*

	Translations		P.O.S.	Freq.	(%)
1	*shin*	(new/least)	prefix	291	(29%)
2	*arata na*	(new/fresh)	adjective	231	(23%)
3	*shinki*	(fresh/new)	noun	149	(15%)
4	*atarashii*	(new/brand-new)	adjective	141	(14%)
5	*arata ni*	(newly/anew)	adverb	47	(5%)
			Total	859	(86%)

Second, *foreign* was rendered into 18 different translations in the sample set. The most frequent translation *gaikoku* occurred only in 541 out of the 1,000 examples (Table 5.3),

which was not frequent enough to be recognised as a single word translation unit. Thus, the identification of translation units at the word level yielded only one: *economic*.

Table 5.3. *Foreign*

	Translations		P.O.S.	Freq.	(%)
1	*gaikoku*	(foreign country/overseas)	noun	541	(54%)
2	*kaigai*	(foreign country/overseas)	noun	111	(11%)
3	*kawase*	(exchange/order)	noun	77	(8%)
4	*gaika*	(foreign currency)	noun	63	(6%)
5	*gaitame*	(foreign exchange)	noun	60	(6%)
			Total	852	(85%)

5.1.2. Above word level

The two adjectives, *new* and *foreign*, required further investigation to identify their translation units, whether at the level of phrase, and clause. First, the collocations of *new* were extracted; there were 53 in total. Most of them were nominal phrases (49 out of 53). All the extracted nominal phrases were adjective-noun sequences; some occurred with one or more extra adjectives, e.g. *new British government*, or with an extra noun, e.g. *new crop months*. There were small numbers of prepositional phrases and clauses: 3 prepositional phrases (*in* or *of* used) and 1 clause (verb-object sequence). Example 5.1 shows the frequent collocations for each category. The prepositional phrases of *new* were actually unusually few, compared with the other target words; for *market*, 12 prepositional phrases were extracted; for *year*, 29 prepositional phrases were identified, as discussed in 4.1.2.

Example 5.1. Examined collocations of *new* (selected)

Nominal phrases	Prepositional phrases	Clause
new rules (16)	*in new orders* (4)	*hold new positions* (5)
new company (16)	*in the new company* (3)	
new government (15)	*of new model cars* (3)	
new products (13)		
new orders (12)		

Among these 53 collocations of *new*, seven were recognised as translation units (Table 5.4). All the translation units of *new* were at the level of the phrase. To be more specific, they were all nominal phrases; no prepositional phrases were identified as translation units. Also, the only extracted collocation at the clausal level, *hold new positions*, was not regarded as a translation unit; therefore, there was no translation unit at the clause level. This indicated that the translation unit size of *new* was restricted to one size only, the phrase.

Table 5.4. Translation units of *new*

	Translation units	Translation equivalents
a.	*new company* (16)	**shin** *gaisha* (14)
b.	*new orders* (12)	**shinki** *juchuu* (12)
c.	*new investments* (4)	**shinki** *tooshi* (4)
d.	*new low* (4)	**shin** *yasune* (4)
e.	*new evidence* (3)	**aratana** *shooko* (3)
f.	*new prime minister* (5)	**shin** *shushoo* (5)
g.	*new crop months* (4)	**shin**koku *gengetsu* (4)

Similarly, the translation units of *foreign* were restricted to one size as well. I examined all the 38 extracted collocations of *foreign*: 25 nominal phrases; 10 prepositional phrases; 2 adjectival phrases; and 1 clause (Example 5.2). All the extracted nominal phrases were adjective-noun sequences, e.g. *foreign investors*; some of them occurred with one extra adjective, e.g. *foreign direct investment*, or with an extra noun, e.g.

foreign exchange market. In the prepositional phrases, *of*, *by*, *to*, *on*, *with*, and *in*, were used. There was only one clause extracted in the analysis; it had a verb-object structure. The unique collocations of *foreign* were the adjectival phrases, which were not extracted in the cases of the other target words. Whether they would be translation units or not was investigated by looking at their translation equivalents in the sample set.

Example 5.2. Examined collocations of *foreign* (selected)

Nominal phrases	Prepositional phrases
foreign investors (85)	*by foreign investors* (12)
foreign funds (32)	*of foreign exchange* (9)
foreign investment (30)	*on foreign investment* (9)
foreign exchange market (28)	*of foreign funds* (7)
foreign firms (23)	*in foreign exchange markets* (7)

Adjectival phrases	Clause
more foreign (9)	*told foreign journalists* (3)
domestic and foreign (5)	

Among the 38 collocations, two were regarded as translation units of *foreign* (Table 5.5). Again, all the translation units were at the level of the phrase; in particular, they were nominal phrases. None of the prepositional and adjectival phrases were identified as translation units. The one-equivalent principle could not recognise the only clause extracted, *told foreign journalists*, as a translation unit of *foreign* either. Therefore, the translation unit size was restricted to be at the level of the phrase in the case of *foreign* as well.

Table 5.5. Translation units of *foreign*

	Translation units	Translation equivalents
a.	*foreign ministry* (4)	**gai***mushoo* (4)
b.	*foreign exchange reserves* (5)	**gai***ka junbi* (5)

One of the translation units actually should not be regarded as a translation unit in this study: *foreign ministry*. If one looks up *foreign ministry* in the whole of ARC, it appears in 171 lines; the capitalised *Foreign Ministry* appears in 111 lines (65 percent); lower-cased *foreign ministry* appears in 60 lines (35 percent). A close look at the concordance lines of both usages (Example 5.3) suggested that whether *foreign ministry* should be capitalised or lower-cased is not yet clearly established in the newswire world. However, the high coverage of capitalised *Foreign Ministry* in the ARC indicated that this is the majority trend in Reuter newswire texts. Since capitalisation is regarded as the sign of a proper noun in this study (3.4.2) and this thesis is not interested in proper nouns, *foreign ministry* was discarded from this study.

Example 5.3. Concordance lines of *foreign ministry*

```
1      Beijing on April 9, a Foreign Ministry spokesman in Hanoi said on
2   the Korean peninsula," a Foreign Ministry spokesman said.
3     of other countries," a Foreign Ministry spokesman said.
4                          A foreign ministry spokesman declined comment on
5                          A foreign ministry spokesman said China's decision
6     up political asylum, a foreign ministry spokesman said.
```

All the other translation units were never used with capitals. For example, the other translation unit *foreign exchange reserves* occurred in 24 lines in the ARC; however, it always appeared in lower case. One might claim that *new prime minister* should be disqualified, since it may be a proper noun as well. If one takes *new* out of the unit, *prime minister* alone actually was likely to be a proper noun. The phrase *prime minister*

152

occurred 889 times in the whole ARC; as shown in Example 5.4, the capitalised usage appeared 770 times (87 percent) and the lower-cased usage appeared 119 times (13 percent). However, the all-capitalised *New Prime Minister* never occurred; therefore, the whole phrase *new prime minister* was not regarded as a proper noun. Thus, *new prime minister* remained in this study as a translation unit.

Example 5.4. Concordance lines of *prime minister*

```
1       day gain since former  Prime Minister Banharn Silpa-archa dissolved
2       strained under former  Prime Minister John Major, to improve.
3       mission led by former  prime minister Chatichai Choonhavan, said recent
4   again today that former   prime minister Chatichai Choonhavan, leader of
5                   Former    prime minister Lee Kwan Yew said at the weekend
```

So far, nine translation units were identified in total in the case studies of the frequent adjectives. In terms of their sizes, the analyses showed that:

1. Translation units were found at the level of the single word and phrase.
2. The identified translation units were predominantly at the phrase level.
3. No translation unit was identified at the level of clause, sentence or larger unit.

5.2.   Translation equivalents

5.2.1. Sizes

As seen in 4.2.1, translation unit sizes were not necessarily identical to their translation equivalent sizes. This conclusion was supported by the case studies of frequent adjectives as well. Table 5.6 is the summary of the sizes of the identified translation units and their equivalents. As discussed in the previous section (5.1), the identified translation units had one dominant size, the phrase. Eight out of nine translation units

were at the level of the phrase. On the other hand, the translation equivalents had two dominant sizes: single word and phrase. They were almost equally dominant; four translation equivalents were at the level of single word; the other five were at the level of phrase. This indicates that some translation pairs underwent a size change. The questions were: which pairs, and why did they change their sizes in translation? The three types of shift were investigated in order to answer these questions: these being, unit-shift (5.2.2), structure-shift (5.2.3), and class-shift (5.2.4).

Table 5.6. Sizes of translation units and equivalents

	Translation units	Translation equivalents
Single word	1	4
Phrase	8	5
Clause	0	0
Sentence	0	0
Total	9	9

5.2.2. Unit-shifts

There were three translation pairs which went through a size change (Table 5.7); using Catford's terms (1965), this phenomenon is called a unit-shift (see the definition in 4.2.2). All the translation units in Table 5.7 were nominal phrases composed by one head noun and one or two adjectives. However, their translation equivalents were at the level of word, not phrase. This was because the adjective *new* was rendered into the prefix *shin*, which was attached to a noun and formed a single word. For example, *shin gaisha* is a single word composed by the prefix *shin* 'new' and the head noun *gaisha* 'company'.

154

Table 5.7. Unit-shifts (i)

	Translation units		Translation equivalents	
a.	***new*** *company*	**Adj** + N	***shin*** *gaisha*	**Pref** + N
b.	***new*** *low*	**Adj** + N	***shin*** *yasune*	**Pref** + N
c.	***new*** *prime minister*	**Adj** + Adj + N	***shin*** *shushoo*	**Pref** + N

The other six translation pairs did not go through a unit-shift (Table 5.8). All the phrasal translation units had their translation equivalents at the phrase level (Table 5.8 *d-h*). Similarly, one single word translation unit had its translation equivalent at the single word level (Table 5.8 *i*). This is because all the English adjectives were rendered either into adjectives or nouns in Japanese, which composed phrases with a head noun. There was one pair in which *new* was rendered into a prefix *shin* (Table 5.8 *g*); however, it did not cause a unit-shift. The translation of *new crop months* was carried out at the phrasal level.

Table 5.8. Unit-shifts (ii)

	Translation units		Translation equivalents	
d.	***new*** *evidence*	**Adj** + N	***aratana*** *shooko*	**Adj** + N
e.	***new*** *investments*	**Adj** + N	***shinki*** *tooshi*	**N** + N
f.	***new*** *orders*	**Adj** + N	***shinki*** *juchuu*	**N** + N
g.	***new*** *crop months*	**Adj** + N + N	***shin****koku gengetsu*	**Pref** + N + N
h.	***foreign*** *exchange reserves*	**Adj** + N + N	***gaika*** *junbi*	**N** + N
i.	***economic***	**Adj**	***keezai***	**N**

5.2.3. Structure-shifts

Structure-shifts (see the definition in 4.2.3) were seen in most of the translation pairs (Table 5.9). All the target words shown in bold in Table 5.9 were adjectives; however, they were rendered into either nouns (*a-d*) or prefixes (*e-h*) in Japanese, which

inevitably caused structural changes. There was only one translation pair which did not go through such a shift: (*i*) *new evidence* and *aratana shooko*. The adjective *new* was rendered into the adjectival equivalent *aratana* and the noun *evidence* was rendered into its nominal equivalent *syooko*.

Table 5.9. Structure-shifts (i)

	Translation units		Translation equivalents	
a.	***new*** *investments*	**Adj** + N	***shinki*** *tooshi*	**N** + N
b.	***new*** *orders*	**Adj** + N	***shinki*** *juchuu*	**N** + N
c.	***foreign*** *exchange reserves*	**Adj** + N + N	***gaika*** *junbi*	**N** + N
d.	***economic***	**Adj**	***keezai***	**N**
e.	***new*** *company*	**Adj** + N	***shin*** *gaisha*	**Pref** + N
f.	***new*** *low*	**Adj** + N	***shin*** *yasune*	**Pref** + N
g.	***new*** *prime minister*	**Adj** + Adj + N	***shin*** *shushoo*	**Pref** + N
h.	***new*** *crop months*	**Adj** + N + N	***shin****koku gengetsu*	**Pref** + N + N
i.	***new*** *evidence*	**Adj** + N	***aratana*** *syooko*	**Adj** + N

Compared with the translation units of nouns (4.2.3), this finding is outstanding. The majority of the translation units of nouns kept their grammatical structures through translations. On the other hand, the majority of the translation units of adjectives could not keep their grammatical structures in translation. This shows a peculiar attribute of adjective translations from English into Japanese.

5.2.4. Class-shifts

A class-shift (see the definition in 4.2.4), on the other hand, was seen in only one pair (Table 5.10 *a*). The transformation from the single word translation unit *economic* to its single word equivalent *keezai* involved a class change from adjective to noun. The other pairs, however, did not go through a class-shift. All the translation units (Table 5.10 *b-i*)

were nominal phrases; their equivalents were either nominal phrases (Table 5.10 *b-f*) or nouns (Table 5.10 *g-k*). Class was kept through translation.

Table 5.10. Class-shifts

	Translation units		Translation equivalents	
a.	***economic***	**Adj**	***keezai***	**N**
b.	***new*** *evidence*	**Adj** + N	***aratana*** *shooko*	**Adj** + N
c.	***new*** *investments*	**Adj** + N	***shinki*** *tooshi*	**N** + N
d.	***new*** *orders*	**Adj** + N	***shinki*** *juchuu*	**N** + N
e.	***new*** *crop months*	**Adj** + N + N	***shin****koku gengetsu*	**Pref** + N + N
f.	***foreign*** *exchange reserves*	**Adj** + N + N	*gaika junbi*	**N** + N
g.	***new*** *company*	**Adj** + N	***shin*** *gaisha*	**Pref** + N
h.	***new*** *low*	**Adj** + N	***shin*** *yasune*	**Pref** + N
i.	***new*** *prime minister*	**Adj** + Adj + N	***shin*** *shushoo*	**Pref** + N

5.2.5. Conclusion

As discussed in the previous section (5.1), the translation unit size of the frequent adjectives was predominantly at the level of the phrase. The translation equivalent sizes were examined in this section and it was found that their dominant sizes were actually two: phrases and single words. This meant that the identified translation pairs were likely to involve some shifts. Careful investigation of each translation pair indicated that most translations actually went through one or, more than one, shift: *a-e* and *g-i* (Table 5.11). There was only one pair whose translation was carried out without any shifts: *f* (shown in bold in Table 5.11).

Table 5.11. Unit-shifts, structure-shifts, and class-shifts (ii)

	Translation units	Unit	Structure	Class
a.	*economic*	not shifted	shifted	shifted
b.	*new company*	shifted	shifted	not shifted
c.	*new orders*	not shifted	shifted	not shifted
d.	*new investments*	not shifted	shifted	not shifted
e.	*new low*	shifted	shifted	not shifted
f.	***new evidence***	not shifted	not shifted	not shifted
g.	*new prime minister*	shifted	shifted	not shifted
h.	*new crop months*	not shifted	shifted	not shifted
i.	*foreign exchange reserves*	not shifted	shifted	not shifted

Comparing this finding with the one from the translation units of nouns, the difference is striking. Most noun translation units were rendered at the same level and kept the same structure and class, while most adjective translation units were rendered with more than one shift. The reason behind this is that the given nouns were all rendered into nouns (Table 4.13), but the given adjectives were rendered into adjectives in only one case (Table 5.10): *new evidence*. The adjectives were often translated into prefixes or nouns, which caused more than one shift in translation. Shift occurrence is therefore likely to be related to whether a given translation unit contains adjectives or not. If a given unit has an adjective, it is likely to go through shifts in its translation. If a given unit does not have an adjective, it is unlikely to go through such shifts.

5.3.   Variations of translation units

The variations of translation units were examined in the case of the frequent adjectives as well. The variations examined were: article, singular/plural, and modifier variations. Whether such variations cause any different translations or not was the main point of investigation.

5.3.1.   Articles

Among the nine translation units of the frequent adjectives, five demonstrated that their article variations did not cause different translations: *economic*, *new company*, *new orders*, *new evidence*, and *new prime minister*. First, the translation units *economic* and *new prime minister* had the most variations of article usage in the sample set (Example 5.5). They appeared with the indefinite article (Line 1, 4); with the definite article (Line 2, 5); and without any article (Line 3, 6). Despite the different article usages, *new prime minister* was rendered into *shin shushoo* all the time, as was *economic* into *keezai*. No word denoting *a*, *an*, and *the* was seen in the Japanese translations.

Example 5.5. Article variations: *new prime minister* and *economic*

```
1      … on the choice of a new prime minister to succeed him, …   (shin shushoo)
2   … long to nominate the new prime minister.                     (shin shushoo)
3      … would be appointed new prime minister," said Bumnan …     (shin shushoo)
4      … their hopes for an economic turnaround on the back of …   (keezai)
5                    "The economic situation and the …             (keezai)
6                      " Economic expansion has been …             (keezai)
```

159

New evidence and *new orders* had fewer variations in their article usages (Example 5.6). They appeared only with the definite article (Lines 1, 4) or without articles (Lines 2-3, 5-7). The indefinite article was not seen in the sample set. The concordance lines in Example 5.6 show that *the new evidence* and *new evidence* shared the same translation, and that *the new orders* and *new orders* shared the same translation. Therefore, article variation did not cause different translations in the case of *new evidence* and *new orders* either.

Example 5.6. Article variations: *new evidence*

```
1     … , but they said the new evidence was serious.          (arata na shooko)
2     … revised in light of new evidence.                       (arata na shooko)
3      … believes there is new evidence to support Jewish …    (arata na shooko)
4              "Clearly the new orders, the backlog, the …      (shinki juchuu)
5                          New orders, supplier deliveries …    (shinki juchuu)
6   … and further growth in new orders.                         (shinki juchuu)
7 … competitiveness and win new orders.                         (shinki juchuu)
```

Similarly, another translation unit, *new company*, agreed with the findings above. It occurred only with indefinite and definite articles. As shown in Example 5.7, both *a new company* and *the new company* were rendered into *shin gaisha*. The article did not make a difference to how this translation unit was rendered.

Example 5.7. Article variations: *new company*

```
1    … supply arm into a new company, Centrica Plc.             (shin gaisha)
2  … they were forming a new company , BBI Healthcare …        (shin gaisha)
3… split the SNCF with a new company taking control of the …   (shin gaisha)
4                    The new company, which would be based …    (shin gaisha)
5 … percent stake in the new company and 44 percent to Texaco. (shin gaisha)
6  … co-chairman of the new company, known as Concert, …       (shin gaisha)
```

160

So far, all the five translation units indicate that article variations do not cause different translations. The rest of the translation units, on the other hand, did not have any variations in their article usages in the sample lines (Example 5.8). *New investments*, *new crop months*, and *foreign exchange reserves* always occurred without articles (Lines 1-4); whilst *new low* always occurred with one type of article (Lines 4-7). In the former cases it was not possible to examine how the definite and indefinite articles were rendered. The latter case, however, supported the central finding; articles were not rendered into Japanese. As shown in Lines 4-7, *a new low* corresponded to *shin yasune*; neither one of *shin* and *yasune* indicates the meaning of the indefinite article. The indefinite article, therefore, was not rendered into Japanese.

Example 5.8. Article variations: others

```
1                        " New investments, Bank of France …   (shinki tooshi)
2                          New crop months gained on ideas …   (shinkoku gengetsu)
3       … of its gold and foreign exchange reserves as at …    (gaika junbi)
4           "We've made a new low on the movement and the …    (shin yasune)
5            F&N struck a new low at S$10, down 30 cents, …     (shin yasune)
6     … rupiah fell to a new low against the dollar on …        (shin yasune)
7       … tin will try a new low this afternoon," …             (shin yasune)
```

Apart from the ones which could not be examined due to a lack of data (*new investments*, *new crop months*, and *foreign exchange reserves*), all of the translation units demonstrated that their indefinite and definite articles were not translated into Japanese. The article variations, therefore, did not cause any different translations. This finding echoes the finding from the analysis of the translation units of nouns (4.3.1). The majority of the noun translation units agreed with the finding of resistance to article variation.

161

5.3.2. Singular and plural forms

Among the nine translation units, only three of them had singular or plural variations in the ARC: *new investments*, *new company*, and *new low*. Do they also show that variations in number do not cause different translations, as shown in the studies of the translation units of nouns? First, Example 5.9 lists some concordance lines of *new investments* (Lines 1-3) and its singular form *new investment* (Lines 4-9). Despite the differences of singular and plural forms, both *new investments* and *new investment* were rendered into *shinki tooshi*. The translation unit *new investments* and its singular form *new investment* shared the same translation equivalent. The singular variations did not create different translations.

Example 5.9. Singular and plural variations: *new investments*

```
1 … than $200 million in new investments, salaries and jobs …   (shinki tooshi)
2    … implementation of new investments.                       (shinki tooshi)
3   … before deciding on new investments, Bank of France …      (shinki tooshi)
4   … , perhaps aimed at new investment by small businesses, …  (shinki tooshi)
5 … effort to encourage new investment in the industry, the …   (shinki tooshi)
6   … it would encourage new investment and lift exports.        (shinki tooshi)
7  … to impose a ban on new investment by Americans in Burma … (shinki tooshi)
8        … 1 billion on new investment in Australia by 2005 …  (shinki tooshi)
9 … elements that lured new investment into Mexico's stock …    (shinki tooshi)
```


Second, the translation unit *new company* and its plural variation *new companies* were different (Example 5.10). They did not share the same translation equivalent. The translation unit *new company* corresponded to *shin gaisha* (Lines 1-2); while, the plural form *new companies* corresponded to *shinki sannyuu suru kakusha* (each company newly joined) in Line 3 and to *2 kigyoo* 'two firm' in Line 4. The plurality was lexically translated in both lines: *kakusha* 'each company' in Line 3 and *2 kigyoo* 'two firm' in

162

Line 4. Therefore, *new companies* and *new company* did not share the same translation equivalent *shin gaisha*. Plurality made a difference to the translation.

Example 5.10. Singular and plural variations: *new company*

```
1 … of stock in the new company for each share of …   (shin gaisha)
2        It said the new company, Netrust, would …     (shin gaisha)
3                 The new companies can now compete …  (shinki sannyuu suru kakusha)
4                 The new companies are expected to be (2 kigyoo)
```

Lastly, the final translation pair, *new low* and *shin yasune*, was different from the pairs mentioned above. The concordance lines of *new low* and its plural variation *new lows* are listed in Example 5.11. As shown in Example 5.11, *new low* and *new lows* had different translations. The striking finding was that *low* and *lows* were actually always rendered into *yasune* in all lines. What made the translations in Lines 4-6 different from the ones in Lines 1-3 was *new*. *New* was rendered into the prefix *shin* in Lines 1-3; while, it was rendered into *tsugi no* 'next of' in Line 4, *ichi dai no* 'new contrast of' in Line 5, and *ichi dai shin* 'new contrast' in Line 6. The plural variation *new lows* had different translations of *new*; therefore, plurality made a difference to the rendering of this translation unit.

Example 5.11. Singular and plural variations: *new low*

```
1        "We've made a new low on the movement and the … (shin yasune)
2     … rupiah fell to a new low against the dollar on … (shin yasune)
3 … think tin will try a new low this afternoon," another … (shin yasune)
4  … the index will test new lows."                       (tsugi no yasune)
5   July and August set new lows.                         (ichi dai no yasune)
6   … at midsession and new lows were set on …            (ichi dai shin yasune)
```

Three translation pairs were investigated for singular and plural variations. One pair indicated that singular variations do not cause different translations; while, two pairs showed that plural variations do cause different translations. This finding is relatively similar to the finding from the noun translation units (4.3.2). However, more examples were examined in the case of the noun translation units; therefore there were more lines to support both findings. One difference between the findings from the nouns and adjectives was the case of *new low* and *new lows*. The plurality in *new lows* was not expressed lexically, but different translations occurred. Such a phenomenon was not seen in the noun translation units, in which the different translations occurred because the singularity and plurality were lexically rendered into Japanese.

5.3.3. Modifiers (i)

The last variation I investigated was that of modifier variations of translation units. There were four translation pairs for which the ARC had modifier variations: (a) *new company* and *new holding company*; (b) *new prime minister* and *new first prime minister*; (c) *new orders* and *new export orders*; and (d) *new low* and *new contract low*. The first two pairs indicated that modifier variations caused different translations; while, pair (c) indicated that the variations did not cause different translations; pair (d), on the other hand, showed a mixture of evidence.

First, the concordance lines of the translation unit *new company* and its modifier variation *new holding company* are shown in Example 5.12 (Lines 1-4). *New company* was rendered into *shin gaisha* (Lines 1-2). However, in Lines 3-4 where the intervening

164

word *holding* appeared between *new* and *company*, *new* was not rendered into the prefix *shin* any longer: it was rendered into the adjective *atarashii* 'new' in Line 3 and the adverb *arata ni* 'newly'. Similarly, the translation unit *new prime minister* and its variation *new first prime minister* showed that their translations were affected by the occurrence of modifiers (Example 5.12 Lines 5-8). The translation unit n*ew prime ministe*r had its equivalent *shin shushoo*; however, the intervening modifier did not allow *new* to translate to *shin* in Lines 7-8. In Line 7, *new* was rendered into the adjective *atarashii*; in Line 8, *new* was not rendered into Japanese at all. These examples indicate that adjacency matters; an intervening modifier caused different translations.

Example 5.12. Modifier variations: *new company and new prime minister*

```
1       … arm into a new company, Centrica Plc.      (shin gaisha)
2 … the SNCF with a new company taking control of …  (shin gaisha)
3     … will form a new holding company with Saban … (atarashii mochikabu gaisha)
4     …, creating a new holding company Norwich …     (arata ni mochikabu gaisha)
5 … would become the new prime minister," said Virat …(shin shushoo)
6 … the choice of a new prime minister to succeed …   (shin shushoo)
7     … to choose a new first prime minister.         (atarashii dai ichi shushoo)
8   … approval of a new first prime minister.         (dai ichi shushoo)
```

On the other hand, the translation pair of *new orders* and *shinki juchuu* was different (Example 5.13). Although another modifier *export* occurred in Lines 3-4, *new* and *orders* were still rendered into *shinki* and *juchuu*. These examples meant that this translation pair was resistant to modifier variation. The intervening modifier did not affect the translation of the units at all.

Example 5.13. Modifier variations: *new orders*

```
1                      New orders in the first quarter …   (shinki juchuu)
2 … whole in terms of new orders and creating jobs …      (shinki juchuu)
3  …, employment and new export orders indices also …     (shinki yushutsu juchuu)
4                      New export orders for Britain's …   (shinki yushutsu juchuu)
```

The last translation pair to be examined was *new low* and *shin yasune* (Example 5.14).

The modifier variation *new contract low* gave contradictory evidence. In Lines 3-5, the

intervening word *contract* did not affect the translations of *new* and *low*; both *new* and

low were rendered into *shin* and *yasune*, which was the same as in Lines 1-2. These

examples indicated, therefore, that the intervening word *contract* did not create different

translations. However, in Lines 6-8, *contract* affected the translations of *new* and *low*.

New was not translated into Japanese any longer in Lines 6-8; only *contract* and *low*

were rendered. For these examples, the variation caused different translations. The

translation pair of *new low* and *shin yasune* was not resistant to modifier variation.

Example 5.14. Modifier variations: *new low*

```
1        "We've made a new low on the movement and …      (shin yasune)
2 … think tin will try a new low this afternoon," another … (shin yasune)
3   … wheat dropped to a new contract low of $3.52-1/2 …   (ichi dai shin yasune)
4… ounce, after seeing a new contract low at $339.10.      (ichi dai shin yasune)
5… lower after setting a new contract low, with the rest … (ichi dai shin yasune)
6 … led by March, with a new contract low of $3.54-1/2 …   (ichi dai yasune)
7   … final minutes to a new contract low.                 (ichi dai no yasune)
8    … at 72.05 cents, a new contract low.                 (yakutei sai yasune)
```

Among the four translation pairs, some demonstrated that an intervening modifier

caused different translations and some did not. This finding agreed with those from the

study of noun translation units above (4.3.3).

5.3.4. Modifiers (ii)

Modifiers can appear before translation units. Such variations were seen in only one pair: *new orders* and *shinki juchuu* (Example 5.15). The translation unit *new orders* and its modifier variations *total new orders* were both rendered into the same translation equivalent *shinki juchuu*. The modifier *total* in Lines 3-4 was not translated into Japanese at all; therefore, the modifier did not break up the translation equivalence between *new orders* and *shinki juchuu*. Comparing this with the noun translation units, the findings were similar: a modifier before the translation unit did not cause different translations in most cases (4.3.4).

Example 5.15. Modifier variations: *new orders*

```
1… and further growth in new orders.                        (shinki juchuu)
2… expect an increase in new orders over the next few …     (shinki juchuu)
3  … to rise, both total new orders and output increased …  (shinki juchuu)
4… on Tuesday that total new orders for UK manufacturing …  (shinki juchuu)
```

5.3.5. Conclusion

The variations of translation units were investigated: specifically, article, singular, plural, and modifier variations. Table 5.12 lists all the variations which did not cause different translations.

Table 5.12. Variations of translation units

	Translation units	Variations	
a.	new orders	new orders, the new orders, new ~ orders, ~ new orders	(4)
b.	economic	economic, an economic, the economic	(3)
c.	new company	a new company, the new company, a new ~ company	(3)
d.	new prime minister	new prime minister, a new prime minister, the new prime minister	(3)
e.	new investments	new investments, new investment	(2)
f.	new low	a new low, a new ~ low	(2)
g.	new evidence	new evidence, the new evidence	(2)
h.	new crop months	new crop months	(1)
i.	foreign exchange reserves	foreign exchange reserves	(1)

The translation unit *new orders* had the most variations: *new orders*; *the new orders*; *new* + modifier + *orders*; and, modifier + *new orders*. Despite the differences, they were all rendered into the same translation equivalent: *shinki jyuchuu*. Therefore, the translation equivalence was established between many translation units and one equivalent (Figure 5.1). On the other hand, the three-word translation unit *foreign exchange reserves* had only one variation: *foreign exchange reserves*. No variation was found in the sample set. Therefore, their translation equivalence was identified to have a one-to-one relationship: one translation unit and one equivalent (Figure 5.1). Similarly, another three-word translation unit *new crop months* was the same; it established one-to-one translation equivalence.

Figure 5.1. Translation equivalence: *new orders* and *foreign exchange reserves*
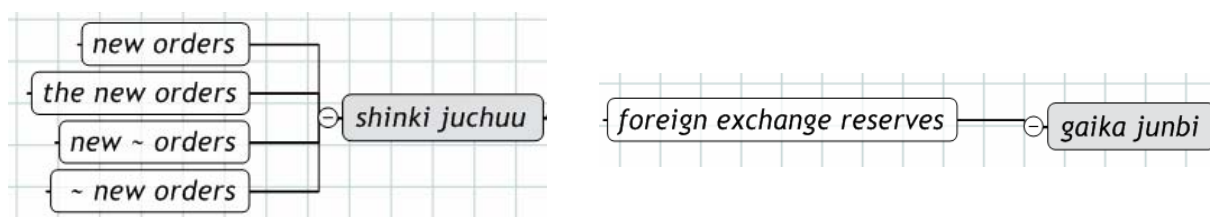
According to the types of translation equivalence, all the translation units can be located on the spectrum from the one-to-one translation equivalence on the left to the many-to-one translation equivalence on the right (Figure 5.2). There was only one translation unit which belongs to the group of four/five/six-to-one translation equivalence. Compared it with the noun translation units where there were four items belonging to this group (Figure 4.2), this was relatively infrequent. Also, one type of translation equivalence was significantly dominant: the two/three-to-one. Seeing that there was no dominant type in the noun translation units, this is a distinctive feature of the translation units of the adjectives.

Figure 5.2. Translation equivalence: from one-to-one to many-to-one (ii)

One-to-one   Many-to-one

⟵———————————————————————————————⟶

new crop months (1)   *economic* (3)   *new orders* (4)
foreign exchange reserves (1)   *new company* (3)
                                   new prime minister (3)
                                   new investments (2)
                                   new low (2)
                                   new evidence (2)
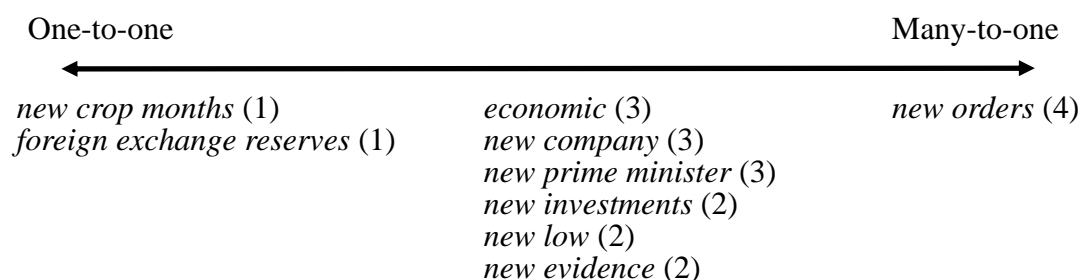
5.4.   Chapter summary

This chapter has presented the analyses of the three frequent adjectives: *economic*, *new*, and *foreign*. First, the one-equivalent principle identified nine translation units (one translation unit of *economic*, seven translation units of *new*, one translation unit of *foreign*). Their sizes ranged from the single word to the phrase; no translation units were found at the levels of clause and sentence. Among the two translation unit sizes, the phrase is the dominant one; eight out of the nine were identified at the phrase level.

Second, the translation equivalent sizes were compared. Their sizes were the single word and the phrase; however, both sizes were almost equally dominant. A closer look at translation equivalence revealed that most of the translation pairs went through one or more shifts. This was different from the noun translation units. Lastly, the variations of the translation units were examined. Most translation pairs had two or three variations which did not cause different translations.

6. Analyses of verbs

The frequent verbs, *said*, *told*, and *expected*, were examined and their translation units were identified. The translation units of these verbs were expected to be different from the identified translation units of the nouns and adjectives. As shown in Table 6.1, most of the translation units of the nouns and adjectives were nominal phrases; among the 21 translation units, 19 of them were nominal phrases. However, such nominal phrases were not expected to be identified in the case of the frequent verbs; verbs do not form nominal phrases. Then, what do the translation units of verbs look like? I examined the sample data sets of the three frequent verbs and carried out the investigation. In 6.1, the findings of their translation unit sizes will be presented; in 6.2, their translation equivalent sizes and shift occurrences will be discussed; and in 6.3, the resistance of translation units to some variations will be investigated.

Table 6.1. Translation units of nouns and adjectives

Single word	Phrase	Clause
government *economic*	*market* sources *market* rumours *market* economy stock *market* gold *market* copper *market* *market* interest rates Seoul stock *market* domestic gold *market* a *year* ago marketing *year* *new* company *new* orders *new* investments *new* low *new* evidence *new* prime minister *new* crop months *foreign* exchange reserves	*market* will reopen

6.1. Translation units

6.1.1. Word level

All the three verbs were investigated in 1,000 sample sets each. However, no translation units were identified at the word level. For example, the sample set of the first verb *said* showed that there were 52 ways to render *said*. The most dominant translation of *said* was *nobe ta* 'stated', appearing only in 259 lines (Table 6.2). This was not dominant enough to make *said* a translation unit.

Table 6.2. *Said* **(i)**

	Translations		P.O.S.	Freq.	(%)
1	*nobe ta*	(stated/expressed)	verb	259	(26%)
2	*yoru*	(depend/base)	verb	191	(19%)
3	*kata tta*	(talked/spoke)	verb	94	(9%)
4	*akiraka ni shi ta*	(demonstrated/declared)	verb	73	(7%)
5	*iu*	(say/tell)	verb	70	(7%)
			Total	687	(69%)

The 1,000 sample lines of the second verb *told* indicated that there were 47 different translations of *told*. As shown in Table 6.3, the most dominant translation was again *nobe ta* 'stated', occurring in 483 lines. The occurrence of *nobe ta* was much more frequent in the case of *told* than the ones in the case of *said*. However, it did not cover 85 percent of the total data of *told*; therefore, the one-equivalent principle could not recognise *told* as a single word translation unit.

Table 6.3. *Told* **(i)**

	Translations		P.O.S.	Freq.	(%)
1	*nobe ta*	(stated/expressed)	verb	483	(48%)
2	*katta tta*	(talked/spoke)	verb	329	(33%)
3	*shimeshi ta*	(showed/revealed)	verb	37	(4%)
4	*akiraka ni shi ta*	(demonstrated/declared)	verb	23	(2%)
5	*shiteki shi ta*	(pointed/indicated)	verb	18	(2%)
			Total	890	(89%)

The third verb *expected* had 62 translations in the data set. Among the most frequent five translations listed in Table 6.4, some were verbs, e.g. *yosoo shite iru* 'is anticipating' and *mi rare te iru* 'is being looked', and some were nouns, e.g. *mitooshi* 'visibility', *yosoo* 'anticipation', and *mikomi* 'chance'. The most dominant translation was the noun *mitooshi*, appearing in 315 lines. This, again, was not frequent enough; therefore, *expected* could not be regarded as a single word translation unit either. Thus,

173

there was no translation unit identified at the level of single word in the case of the frequent verbs.

Table 6.4. *Expected* **(i)**

	Translations		P.O.S.	Freq.	(%)
1	*mitooshi*	(visibility/prospect)	noun	315	(32%)
2	*yosoo*	(expectation/anticipation)	noun	84	(8%)
3	*yosoo shi te iru*	(is expecting/anticipating)	verb	78	(8%)
4	*mi rare te iru*	(is being looked/seen)	verb	62	(6%)
5	*mikomi*	(expectation/chance)	noun	53	(5%)
			Total	592	(59%)

One might suggest that the finding might be different if these analyses had been carried out with lemmas instead of word-forms (the lemma and word-form are defined by Hunston (2002) that '*eat*, *eats*, *eating*, *ate* and *eaten* are word-forms belonging to the lemma *EAT*'). However, the finding would still be the same; translation units of the verbs would not be identified at the single word level. *Said*, for example, was rendered into 22 lemmas. Table 6.5 shows the most frequent ones. The most dominant lemma *NOBE RU* 'state' includes *nobe ta* 'stated', *nobe te i ru* 'is stating', *nobe te i ta* 'was stating', *nobe ru* 'state', etc. The total occurrence of *NOBE RU* was still 278 lines out of 1,000 samples; it was not dominant enough to make the single word *said* a translation unit.

174

Table 6.5. *Said* (ii)

	Translations		P.O.S.	Freq.	(%)
1	*NOBE RU*	(state/express)	verb	278	(28%)
2	*YORU*	(depend/base)	verb	191	(19%)
3	*KATA RU*	(talk/speak)	verb	94	(9%)
4	*SHIMESU*	(show/reveal)	verb	77	(8%)
5	*AKIRAKA NI SURU*	(demonstrate/declare)	verb	76	(8%)
			Total	716	(72%)

Similarly, the most dominant lemma used as the translation of *told* was also *NOBE RU*; it occurred in 507 lines (Table 6.6). The main word-forms seen in the sample data were *nobe ta* 'stated', *nobe te iru* 'is stating', and *no be te i ta* 'was stating'. Even though all these word-forms were included, the lemma *NOBE RU* could not cover 85 percent of the total data of *told*.

Table 6.6. *Told* (ii)

	Translations		P.O.S.	Freq.	(%)
1	*NOBE RU*	(state/express)	verb	507	(51%)
2	*KATA RU*	(talk/speak)	verb	330	(33%)
3	*SHIMESU*	(show/reveal)	verb	41	(4%)
4	*SHITEKI SURU*	(point/indicate)	verb	24	(2%)
5	*AKIRAKA NI SURU*	(demonstrate/declare)	verb	24	(2%)
			Total	927	(93%)

Lastly, the most dominant lemma used as the translation of *expected* was the noun *MITOOSHI* 'prospect', seen in 288 lines. The word-form seen in the data was only one: *mitooshi*. Since the occurrence of *MITOOSHI* was not frequent enough, the finding remained the same; *expected* could not be regarded as a single word translation unit.

Table 6.7. *Expected* **(ii)**

	Translations		P.O.S.	Freq.	(%)
1	*MITOOSHI*	(visibility/prospect)	noun	288	(29%)
2	*YOSOO SURU*	(expect/anticipate)	verb	260	(26%)
3	*MI RU*	(look/see)	verb	118	(12%)
4	*YOSOO*	(expectation/anticipation)	noun	84	(8%)
5	*MIKOMU*	(expect/take into account)	verb	72	(7%)
			Total	822	(82%)

Thus, no matter whether lemmas or word-forms are concerned, these frequent verbs could not be regarded as translation units at the single word level. Among the three word classes (noun, adjective, and verb), the investigation of the frequent verbs is the only case in which single word translation units were not identified.

6.1.2. Above word level

All the three verbs required further investigations at the levels of phrase, and clause. First, collocations of *said* were extracted: one verbal phrase and 24 clauses were found. The verbal phrase formed the present perfect tense: *have said*. The clauses were all subject-verb sequences, e.g. *he said*, *it said*, and *brokers said*. Among the 25 collocations, only one of them was regarded as a translation unit: *Shanghai Securities News said* (Table 6.8).

Table 6.8. Translation units of *said*

	Translation units	Translation equivalents
a.	*Shanghai Securities News said* (3)	*shanhai shooken hoo ni* **yoru** *to* (3)

Similarly, larger units of *told* were identified: 26 collocations. They were: two verbal phrases, e.g. *had told*; five verb-object clauses, e.g. *told reporters*; and 19 subject-verb

clauses, e.g. *he told*. However, none of them was regarded as a translation unit. All the items failed to have one dominant translation in the sample set. For example, the verbal group *had told* appeared in 7 out of the 1,000 lines of *told* (Example 6.1). All translations were different: *kata tta* 'talked' in Line 1; *hatsugen o okona tta* 'performed a speech' in Line 2; *nobe te i ta* 'was telling' in Line 3; *tsutae ta* 'informed' in Line 4; *shiji shi ta* 'ordered' in Line 5; *shimeshi ta* 'showed' in Line 6; and *hookoku shi te iru* 'is reporting' in Line 7. By having such a variety of translations, *had told* could not be recognised as a translation unit. Similarly, all the other collocations of *told* had such diversity in their translations; hence, no translation units were identified in the case of *told*.

Example 6.1. Concordances of *had told*

```
1 … Monday that Hwang had told an official from the U.S …   (kata tta)
2    … Jacques Chirac had told him remarks about the …      (hatsugen o okona tta)
3        … authorities had told him they had arrested up …  (nobe te i ta)
4              Yeltsin had told Kohl about efforts to …     (tsutae ta)
5        … commission had told local futures brokerages …   (shiji shi ta)
6 … that Strauss-Kahn had told Socialist members of …       (shimeshi ta)
7    Sanyo Universal had told the stock exchange that …      (hookoku shi te iru)
```

For the third verb *expected*, 83 collocations were extracted for the identification of translation units: 48 verbal phrases, e.g. *is expected to be*; 10 verb-object clauses, e.g. *expected prices*; 21 subject-verb clauses, e.g. *analysts expected*; and 4 subject-verb-object clauses, e.g. *they expected the market*. Only one of them was regarded as a translation unit: *dealers expected*, as shown in Table 6.9.

177

Table 6.9. Translation units of *expected*

Translation units	Translation equivalents
a. dealers expected (5)	*diiraa ra wa … **yosoo shi te iru*** (5)

So far, two translation units were identified in the case of the frequent verbs: *Shanghai Securities News said* and *dealers expected*. Both were at the level of the clause. This finding is rather unique, if one considers the cases of the frequent nouns and adjectives. First, as shown in Table 6.1, the translation units of the nouns and adjectives were dominantly phrases, while no phrasal translation unit was identified in the case of the frequent verbs. Second, the translation unit sizes were more diverse in the cases of the nouns and adjectives. The translation units of the nouns were at the levels of single word, phrase, and clause; the translation units of the adjectives were at the levels of single word and phrase. On the other hand, the translation units of the verbs were restricted to a single level: the clause. Thus, the translation units of the verbs were distinctive from the ones of the nouns and adjectives, as discussed at the beginning of this chapter. However, neither of the translation units were verb-object clauses, as Newmark (1988b) suggests; rather, the translation units were all subject-verb clauses (the relevant discussion will be held in 7.2.3).

6.2.   Translation equivalents

6.2.1.   Sizes

As discussed in the previous section (6.1), the translation unit size was restricted to be at one size in the case of the frequent verbs: the clause. What about the sizes of their translation equivalents? The corpus data proved that they were also restricted to be at

one size, which was the clause. The translation unit, *Shanghai Securities News said*, corresponded to its translation equivalent, *shanhai shooken hoo ni yoru to*. This translation contained the verb *yoru* 'base', the object *shanhai shooken hoo* 'Shanghai Securities News' indicated by the object marker *ni* (Takahashi et al., 2005), and the conjunction *to* 'if' (Umesao et al., 1989); therefore, this formed a dependent clause which meant 'according to Shanghai Securities News'. Also, the other translation unit, *dealers expected*, had its translation equivalent, *diiraa ra wa ... yosoo shi te iru* 'dealers … are anticipating'. Consisting of the subject *diiraa ra* 'dealers' indicated by the subject and topic marker *wa* (Takahashi et al., 2005) and the verb *yosoo shi te iru* 'is anticipating', this translation equivalent formed an independent clause. Thus, both sizes of the translation units and equivalents were at the level of the clause; the translation unit sizes were kept through translations in the case of the frequent verbs.

6.2.2. Unit-shifts

As discussed in 6.2.1, the translation unit sizes were not changed in translations of the frequent verbs. The clausal translation units were rendered into their clausal translation equivalents. All the transformation was carried out at the same level. Therefore, none of the translation pairs underwent a unit-shift (see the definition in 4.2.2).

6.2.3. Structure-shifts

Structure-shifts (see the definition in 4.2.3), on the other hand, were seen in all the translation pairs. The first translation unit *Shanghai Securities News said*, for example, had a subject-verb sequence; while, its translation equivalent *shanhai shooken hoo ni yoru to* did not have such a sequence. Subjects in Japanese are usually indicated by subject markers *wa*, *ga*, or no case marker (Takahashi et al., 2005). *Shanhai shooken hoo ni yoru to*, however, had an object marker *ni* (Takahashi et al., 2005), which indicated that the word preceding, *shanhai shooken hoo*, is an object. Seeing that *shanhai shooken hoo ni yoru to* had the object *shanhai shooken hoo ni* (Shanghai Securities News), the verb *yoru* 'base', and the conjunction *to* 'if', this translation was a verb-object sequence with a conjunction. The translation unit structure was, therefore, changed through translation.

The second translation unit *dealers expected* also had a subject-verb sequence, while its translation equivalent *diiraa ra wa … yosoo shi te iru* had a slightly different structure. The subject *dealers* was rendered into the corresponding subject *diiraa ra* 'dealers' with its subject and topic marker *wa* (Takahashi et al., 2005); the verb *expected* was rendered into the corresponding verb *yosoo shi te iru* 'is anticipating'. However, *dealers* and *expected* were adjacent in English; while, *diiraa ra* and *yosoo shi te iru* were located remotely in the Japanese sentence. This happened because Japanese verbs come at the end of sentence (discussed in 3.5). For example, the original sentence, *Dealers expected the market to stay range-bound*, was rendered into the Japanese sentence, *diiraa ra wa, shijoo wa kongo renji torihiki ga tsuzuku, to yosoo shi te iru* 'dealers are anticipating

180

that market continues trade (within the) range in the future'. The subject *diiraa ra* appeared at the beginning and the verb *yosoo shi te iru* appeared at the end of the sentence. Therefore, the structure was not kept in translation in this translation pair either.

6.2.4. Class-shifts

Class-shifts (see the definition in 4.2.4) were not examined in the translation pairs of the verbs. As discussed in 4.2.4, clauses do not belong to any classes unlike words and phrases which can belong to nouns (nominal groups) or adjectives (adjectival groups). Since all the translation units were at the level of clause, the investigation of class-shifts was not possible.

6.2.5. Conclusion

Table 6.10 is a summary of the translation pairs and their shift occurrences. Unit-shifts were not seen in any of the pairs; while, structure-shifts were seen in all the pairs. Unlike the translation units of the frequent nouns and adjectives, all translation units required shifts in translations. This was because all the translation units were clauses; due to the different language systems, clausal translation units are more likely to require structural changes through translation than phrasal and single word translation units. Seeing that the other clausal translation unit, *market will reopen* (Table 4.14), also went through a structure change, this interpretation is reasonable.

Table 6.10. Unit-shifts, structure-shifts, and class-shifts (iii)

	Translation units	Unit	Structure	Class
a.	*Shanghai Securities News said*	not shifted	shifted	-
b.	*dealers expected*	not shifted	shifted	-

6.3. Variations of translation units

The variations of the two translation units were examined in the analyses of the frequent verbs as well. The variations considered were ones occurring through: (a) different article usages, e.g. *Shanghai Securities News said* and *the Shanghai Securities News said*; (b) singular and plural forms, e.g. *dealers expected* and *a dealer expected*; and (c) modifier insertions, e.g. *dealers expected* and *most dealers expected*. The main interest was whether such variations had any impact on the translation pairs or not.

6.3.1. Articles

The first translation unit *Shanghai Securities News said* occurred with the definite article (Lines 1-3) and without any articles (Line 4) in the sample set (Example 6.2). The definite articles in Lines 1-3 were not rendered into Japanese; therefore, *the Shanghai Securities News said* and *Shanghai Securities News said* shared the same translation. The article variation *the Shanghai Securities News said* did not cause a different translation from *Shanghai Securities News*.

182

Example 6.2. Article variations: *Shanghai Securities News said*

```
1     … firms, the Shanghai Securities News said on … (shanhai shooken hoo ni yoru to)
2   … markets, the Shanghai Securities News said on … (shanhai shooken hoo ni yoru to)
3        … , the Shanghai Securities News said on … (shanhai shooken hoo ni yoru to)
4      … second, Shanghai Securities News said on … (shanhai shooken hoo ni yoru to)
```

On the other hand, the other translation unit, *dealers expected*, did not have any article variations in the sample data. It always occurred without articles (Example 6.3); therefore, the investigation was not possible.

Example 6.3. Article variations: *dealers expected*

```
1                  Dealers expected the London market … (diiraa ra wa … yosoo shi te iru)
2                  Dealers expected the market to …     (diiraa ra wa … yosoo shi te iru)
3                  Dealers expected the Nikkei to …     (diiraa ra wa …yosoo shi te iru)
4                  Dealers expected gold to trade …     (diiraa ra wa …yosoo shi te iru)
5 … shadow, and dealers expected gold to test the …    (diiraa ra wa …yosoo shi te iru)
```

Such a finding was actually identical to the ones in the cases of the frequent nouns and adjectives. Definite and indefinite articles were not usually rendered into Japanese; therefore, article variations of a given translation unit did not cause different translations.

6.3.2.   Singular and plural forms

The singular and plural variations were investigated in the cases of the verbs. First, the translation unit *Shanghai Securities News said* contained the noun *Shanghai Securities News*. However, *Shanghai Securities News* was a proper name; there was no singular or plural form of this proper name existing in the ARC. Second, the translation unit *dealers*

expected was not able to be examined either. The singular variations such as *a dealer expected* or *dealer expected* were not found in the corpus.

6.3.3. Modifiers

On the other hand, both translation units had modifier variations in the ARC. As shown in Example 6.4, *Shanghai Securities News said* occurred with *the official* in Line 1, *Thursday's* in Line 2, *Tuesday's* in Line 3, and *Wednesday's* in Line 4. No matter what appeared before *Shanghai Securities News said*, it was always rendered into *shanhai shooken hoo ni yoru to*; the extra items were not expressed in Japanese. The modifiers did not cause different translations.

Example 6.4. Modifier variations: *Shanghai Securities News said*

```
1   … in the official  Shanghai Securities News said. (shanhai shooken hoo ni yoru to)
2        in Thursday's  Shanghai Securities News said. (shanhai shooken hoo ni yoru to)
3         in Tuesday's  Shanghai Securities News said. (shanhai shooken hoo ni yoru to)
4       in Wednesday's  Shanghai Securities News said. (shanhai shooken hoo ni yoru to)
```

Also, the ARC had one modifier variation of the translation unit *dealers expected* (Example 6.5): *most dealers expected* in Line 3. If one compares Line 3 with Lines 1-2, in which *dealers expected* appeared without modifiers, one realises that the modifier *most* did not cause a different translation of *dealers* and *expected*. *Dealers* in Line 3 still corresponded to *diiraa ra wa* 'dealers', which was the same as Lines 1-2. *Expected* in Line 3 still corresponded to *yosoo shi te iru* 'is anticipating', which was again the same as Lines 1-2.

Example 6.5. Modifier variations: *dealers expected*

```
1          Dealers expected the Nikkei to …        (diiraa ra wa …yosoo shi te iru)
2          Dealers expected the market to …        (diiraa ra wa …yosoo shi te iru)
3    Most  dealers expected the STI index …   (taihan no diiraa ra wa …yosoo shi te iru)
```

The modifier variation in which a modifier appeared in the middle of a given translation

unit was also examined as well. In the analyses of the frequent nouns and adjectives, the

variations examined were all nominal phrases: (a) the translation unit *gold market* and

its variation *gold futures market*; (b) the translation unit *market economy* and its

variation *market sector economy*; (c) the translation unit *new company* and *new holding*

company; (d) the translation unit *new prime minister* and *new first prime minister*; (e)

the translation unit *new orders* and *new export order*; and (f) the translation unit *new*

low and *new contract low*. Such modifier variations were not found in the cases of the

frequent verbs. However, there was a different variation found; an intervening word

appeared in the verb phrase, as shown in Lines 3-4 (Example 6.6).

Example 6.6. Modifier variations: *dealers expected*

```
1          Dealers expected the London market …    (diiraa ra wa …yosoo shi te iru)
2          Dealers expected the market to …        (diiraa ra wa …yosoo shi te iru)
3          Dealers had expected Singapore to fall … (diiraa ra wa …yosoo shi te ita)
4          Dealers had expected technical support … (diiraa ra ni yoru to)
```

Example 6.6 shows that this intervening word *had* caused a different translation. Lines

1-2 present how the translation unit *dealers expected* was rendered without *had*: *diiraa*

ra wa … yosoo shi te iru 'dealers…are anticipating'. The translation of *dealers had*

expected in Lines 3-4 differed from Lines 1-2: *diiraa ra wa … yosoo shi te ita*

'dealers…were anticipating' in Line 3 and *diiraa ra ni yoru to* 'according to dealers' in

Line 4. Therefore, this intervening word *had* caused different translations of *dealers* and *expected.*

6.3.4. Conclusion

Both translation units had some variations which did not cause different translations. The translation unit *Shanghai Securities News said* had four: *Shanghai Securities News said*; *the Shanghai Securities News said*; *the* + modifier + *Shanghai Securities News said*; and, modifier + *Shanghai Securities News said*. They were all rendered into the same translation equivalent: *shanhai shooken hoo ni yoru to* (Figure 6.1). This means that this translation pair established a many-to-one translation equivalence.

Figure 6.1. Translation equivalence: *Shanghai Securities News said*
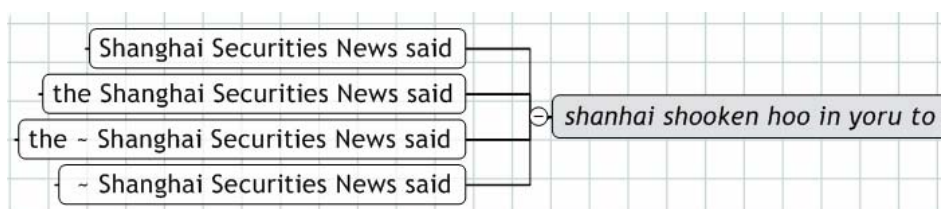
The other translation unit *dealers expected* had fewer variations; *dealers expected*; modifier + *dealers expected*; and, *dealers* + modifier + *expected*. These three items shared the same translation equivalent: *diiraa ra wa … yosoo shi te iru* (Figure 6.2). Therefore, again, a many-to-one relationship was recognised in this pair as well. There was no translation unit which had one-to-one translation equivalence in the case of the verbs.

Figure 6.2. Translation equivalence: *dealers expected*
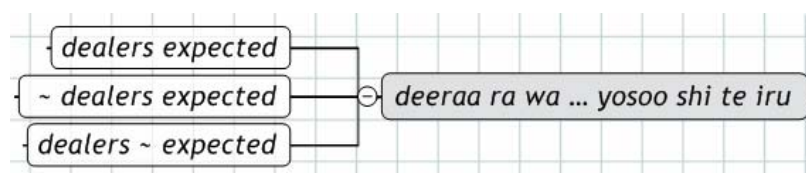
6.4.   Chapter summary

This chapter has presented the analyses of the three frequent verbs: *said*, *told*, and *expected*. In total, there were only two translation units identified. One of the interesting findings was that their sizes were restricted to one level: the clause. Moreover, their translation equivalent sizes were also at a single level: the clause again. Such restriction on translation unit and equivalent sizes was not seen in the cases of the nouns and adjectives. Another finding was that shifts always occurred in rendering the translation units. This was because all the translation units were clauses; clauses required structural shifts in translation due to the different language systems. The investigation of the translation units and their variations revealed that both translation units displayed many-to-one translation equivalence.

7. Rethinking translation unit size

The previous chapters examined all of the nine target words and identified their translation units. The discussions of the translation unit size, however, were made separately by word groups. This chapter, therefore, will gather all the findings from each word group and interpret them. It is appropriate that this chapter will start by revisiting the research questions and their related hypotheses. The main focus in 7.1 is on whether the analyses of the nine words could give relevant answers to the questions or not; and, whether the results were similar to the hypotheses in 3.5. The focus in 7.2 will be on the detailed implications of the identified translation unit sizes, followed by the implications of the identified translation equivalence in section 7.3. Also, I will investigate the units that the one-equivalent principle could not regard as translation units in 7.4. Examining such non-translation units is important for identifying key improvements on the methodology and theory used in this study; these points will be discussed in the subsequent section, 7.5.

7.1. Revisiting research questions and hypotheses

There were two main research questions (discussed in 3.1): (a) how large or small the translation units are in English-Japanese newswire texts, rendered by experienced translators, from a lexical point of view; and, (b) whether the translation units are restricted to only one size, or a selection of the possible lengths. Consulting with the relevant literature and taking into account the research conditions, hypotheses were formulated (discussed in 3.5). For the question (a), I expected the translation unit sizes

to be larger than the morpheme and smaller than the sentence: that is, single words; phrases; and clauses. For the question (b), therefore, the sizes were expected to be distributed rather than restricted to one size. Among the three expected sizes, the phrase was likely to be the most dominant translation unit size; while the single word was likely to be the least identified. Were these hypotheses supported by the case studies of the nine frequent words?

Table 7.1 presents all the identified translation units in this study; the target words are shown in bold. There were 24 translation units in total; the sizes were single words, phrases, and clauses. Amongst the three, phrasal translation units were identified the most; 19 out of 24 were at the phrasal level. The single word translation units were identified the least; only two out of 24 were at the single word level. This suggests that all the findings matched with hypotheses (a) and (b).

Table 7.1. Identified translation units and their sizes

	Single words	Phrases	Clauses
1.	*government*	*market* sources	*market* will reopen
2.	*economic*	*market* rumours	Shanghai Securities News *said*
3.		*market* economy	dealers *expected*
4.		stock *market*	
5.		gold *market*	
6.		copper *market*	
7.		*market* interest rates	
8.		Seoul stock *market*	
9.		domestic gold *market*	
10.		a *year* ago	
11.		marketing *year*	
12.		*new* company	
13.		*new* orders	
14.		*new* investments	
15.		*new* low	
16.		*new* evidence	
17.		*new* prime minister	
18.		*new* crop months	
19.		*foreign* exchange reserves	

It is worth pointing out that the value difference between the two minor sizes, i.e. single word and clause, was very small. As shown in Table 7.1, there were two single word translation units and three clausal translation units; this slight difference makes single word the least dominant size and clause the second dominant size. It is possible that this value difference disappears at some point, as more target words are investigated, eventually making the clause the least dominant size, instead of single word. However, this never happens, as long as the nine most frequent words, or subsets of these ordered by frequency, are concerned. For example, if one studies the subset of the six most frequent words (*market*, *year*, *economic*, *new*, *said*, and *told*), there would be only one single word translation unit (e.g. *economic*) and two clausal translation units (e.g. *market will reopen* and *Shanghai Securities News said*). Hence, the single word remains the least dominant size. Similarly, if one studies the set of the three most frequent words (*market*, *economic*, and *said*), there would again be only one single word translation

unit (e.g. *economic*) and two clausal ones (e.g. *market will reopen* and *Shanghai Securities News said*). The least dominant translation unit size would still be the single word. Therefore, the most minor size is always the single word translation unit within the nine frequent words, when ordered by frequency.

It was rather surprising that all the findings were supported by the hypotheses. This was because the hypotheses were based on only the four perspectives of this study: lexically defined translation units; English-Japanese comparison; newswire texts; and, experienced translators (see details in 3.5). The hypotheses did not take account of target words; they were focused on the content words which occurred most frequently in the corpus. Considering that frequent words behave differently from infrequent words (Sinclair, 1999), the finding of this study was expected to disagree with the hypotheses, to some extent. However, such disagreement was not seen; the translation unit sizes of the frequent words were identical to the translation unit sizes which I had hypothesised without taking the idiosyncrasy of frequent words into account. This indicates that the translation unit sizes do not differ between frequent and infrequent words. The idiosyncrasy of frequent words did not cause a size difference in translation unit.

This interpretation makes sense, since the idiosyncrasy of frequent words emerges only when they were examined at the single-word level. To take an example from Sinclair's study (1999), the noun *way* is a frequent word. *Way* is often used in specific contexts where no other nouns fit well, e.g. *on my way home*, *a way of –ing*, and *half way through*. Therefore, the behaviour of this noun is very different from that of other nouns. However, unlike this examination, in which *way* alone was examined in terms of its

191

usage, the examination of larger units, such as *half way through*, suggests that such units are not actually idiosyncratic. For example, the nominal phrase *half way through* appears in the contexts where other nominal phrases can occur. The sentence, *It was not until half way through* (from Sinclair (1999)), is possible to have *his last year* (*it was not until his last year* (from Google)) and *the early 20th century* (*it was not until the early 20th century* (from Google)) instead of *half way through*. This replaceability means that *half way through* is not grammatically unique; the behaviour of *half way through* as a whole is rather standard as a nominal phrase. Therefore, the difference between what the examination of *way* tells and what the examination of *half way through* tells is large and important. The first examination reveals the idiosyncrasy of frequent words when examined at the single-word level; however, frequent words are well-handled in the second methodology looking at the larger units of which they form a part. This study takes the path of the second methodology; hence, the idiosyncrasy of frequent words did not affect the findings.

7.2.   Translation unit sizes

The three sizes of translation units were not the only findings that this empirical study provided. There were other useful findings regarding translation unit size. This subsection will be dedicated to some of them. First, I will discuss about the relationship between translation unit size and parts of speech (7.2.1). Second, the phrasal translation units and their traits will be focused on (7.2.2), followed by the characteristics of the clausal translation units (7.2.3), and the single word translation units (7.2.4). These discussions will uncover the fact that there were specific types of phrases, clauses, and

single words which are translation units. Third, the identified translation unit sizes will be compared with those in the literature (7.2.5). This comparison is very helpful in recognising what new insights this empirical study has brought to the issue of translation unit size. Lastly, the shortcomings of the findings will be investigated and clarified (7.2.6).

7.2.1. Parts of speech

One useful finding of this study was that translation unit sizes vary according to the word classes of target words; noun, adjective; and, verb. As shown in Table 7.1, translation units of the nouns were identified at any of the three sizes, e.g. *government* at the single word level, *a year ago* at the phrasal level, and *market will reopen* at the clausal level. On the other hand, the translation units of the adjectives and verbs were different; they were rather selective. The former was identified only at the levels of single word and phrase, e.g. *economic*, and *new evidence*. The latter was identified only at the level of clause, e.g. *Shanghai Securities News said*. Why did this selectivity occur? Table 7.2 gives some reasons behind this. The numbers present how many items were examined at each size in the analyses of the nouns, adjectives, and verbs, respectively; e.g. the examined items regarding the target nouns are 98 in total which includes three single words, 84 phrases, and 11 clauses. The numbers in parentheses refer to how many translation units were identified at each size; e.g. 11 clauses of nouns were examined and one of them was identified as a translation unit. The grey highlight shows where the numbers of identified translation unit is one or more.

Table 7.2. Examined examples at each size

	Nouns		Adjectives		Verbs	Total
Single words	3	(1)	3	(1)	3	9
Phrases	84	(11)	91	(8)	50	225
Clauses	11	(1)	2		83 (2)	96
Total	98		96		136	330

As Table 7.2 shows, most translation units were identified where there were many examined items: e.g. phrases of adjectives (91 items were examined and 8 translation units were identified); phrases of the nouns (84 items were examined and 11 translation units were found); clauses of the verbs (83 items were examined and 2 translation units were recognised); and, clauses of nouns (11 items were examined and 1 translation unit was extracted). This suggests that the more one examines, the more translation units one can extract. The converse of this is that translation units were not identified where there were only a few examined items, which was seen in the clauses of the adjectives. There were only two examined items; hence, no translation units were identified. However, there were some phenomena which are harder to explain by this logic. One of the peculiar cases was the phrases of the verbs. Although they were examined quantitatively (50 items), all of them failed to be translation units. If one compares it with the clauses of the nouns where one translation unit was identified out of only 11 examined items, this result of the phrasal verbs was rather striking (further discussion will be made in 7.2.2 and 7.4.3).

One implication from these findings is that researchers have to have target words from different parts of speech for their studies of translation unit size. Otherwise, the study does not provide a comprehensive view on the size matter. If one's target words are only

adjectives (*economic*, *new*, and *foreign*), this study suggests that the translation unit sizes will only be at the single word and phrase levels. Similarly, if one's target words are only verbs (*said*, *told*, and *expected*), the study shows that the translation unit size will only be at the clause level. Neither of these approaches in isolation would be appropriate for understanding translation unit sizes of content words. On the other hand, if one's target words are only nouns (*market*, *year*, and *government*), the study shows that the translation unit sizes will be at the single word, phrase, and clause levels, which is exactly the same as what this study found by studying all of the word classes. This suggests, therefore, that there are two ways to investigate translation unit size of content words. One is to examine words from all the word groups; this study belongs to this group. The other is to examine nouns only. As long as nouns are examined, it is likely that a comprehensive result of translation unit size regarding content words would be achieved. In other words, nouns can act as a representative of the three word groups in the study of translation unit size.

One might wonder why it is, that nouns can be a representative of all content words in the study of translation unit size. This is because nouns are a core element of the translation units. As shown in Table 7.1, most of the identified translation units have one noun at least; there is only one translation unit which does not contain any nouns: *economic*. All of the translation units except for *economic* can be identified in the analyses of nouns. For example, the phrasal translation unit *new evidence* was found in the study of *new*; however, it is also possible to identify this if one analyses the noun *evidence*. Similarly, the clausal translation unit *dealers expected* was found in the study of *expected*; it is again possible to identify this if one examines the noun *dealers*. As

195

long as the nouns are examined, all the phrasal and clausal translation units in Table 7.1 can be identified. Therefore, nouns are a good word group to focus on in the study of translation units and their sizes.

This implication is actually useful; it allows researchers to have more target words, even though a given corpus is as small as the ARC. As discussed in 3.4.4, this study focused on only three frequent words from each word group, since the methodology required the same number of the target words from each word group. Verbs are particularly inconvenient for this method; they are very infrequent (
Table 7.3). There were only two verbs which occurred more than 2,000 times in the corpus. The third most frequent verb was, exceptionally, accepted as a target word, as it could produce a 1,000 sample set even after discarding the noise data. However, if one uses the noun-only method, one is able to have more target words. As shown in
Table 7.3, there were 17 eligible nouns for the identification of translation units. Seeing that parallel corpora tend to be smaller than the ARC, on average, at the present time, this is a promising finding for translation unit size research.

Table 7.3. Frequency list of the ARC

No	Nouns	Freq	Adjectives	Freq	Verbs	Freq
1	*market*	5,125	*economic*	3,206	*said*	38,340
2	*bank*	3,921	*new*	2,653	*told*	3,398
3	*government*	3,560	*foreign*	2,575	*expected*	1,612
4	*prices*	3,335	*central*	2,193	*rose*	1,430
5	*year*	3,269	*monetary*	2,041	*closed*	1,224
6	*traders*	3,054	*international*	1,908	*added*	1,211
7	*trade*	2,829	*higher*	1,888	*saying*	1,138
8	*tonnes*	2,812	*European*	1,678	*ended*	1,096
9	*dollar*	2,635	*financial*	1,595	*fell*	964
10	*shares*	2,527	*domestic*	1,516	*continue*	747
11	*news*	2,471				
12	*futures*	2,435				
13	*growth*	2,340				
14	*stocks*	2,294				
15	*oil*	2,172				
16	*economy*	2,077				
17	*minister*	2,006				
18	*report*	1,919				
19	*currency*	1,895				
20	*company*	1,831				

7.2.2. Phrases

Some traits of phrasal translation units were revealed in this empirical study. First, all the phrasal translation units were nominal phrases (Table 7.1). The other phrasal types such as adjectival phrases (e.g. *more foreign*), prepositional phrases (e.g. *of the year*), and verbal phrases (e.g. *have said*) were not identified as translation units. Therefore, the dominance of nominal phrases is an unmistakable characteristic of phrasal translation units in this study. Did this happen because this study examined many more nominal phrases than the other types of phrases? In fact, if one looks at Table 7.4, nominal phrases actually occurred most frequently in the sample sets; hence, they were examined the most (119 items in total). This is doubtless one reason for the large number of nominal phrasal translation units.

Table 7.4. Examined collocations (phrases)

	Nouns	Adjectives	Verbs	Total
Adjectival phrases	0	2	0	2
Nominal phrases	43 (11)	76 (8)	0	119
Prepositional phrases	41	13	0	54
Verbal phrases	0	0	50	50
Total	84	91	50	225

However, as shown in Table 7.4, prepositional phrases also occurred rather substantially in the sample sets; in fact, 54 items were examined to see whether they could be translation units or not. Although this number was not as large as the examined items of nominal phrases (i.e. 119 phrases), it was surprising that none of the 54 prepositional phrases could be regarded as a translation unit (see 7.4.1 for more discussion). Similarly, verbal phrases were also examined at a substantial level (i.e. 50 phrases); but, the

198

one-equivalent principle could not identify any of them as translation units (see 7.4.3 for more discussion). These indicate that prepositional and verbal phrases are very unlikely to be translation units, however many items of such kind are examined. The other phrasal type, adjectival phrases, was different. Adjectival phrases were very infrequent in the sample sets; there were only two examined items. Therefore, there can be two possible reasons for the zero existence of adjectival translation units. Either it was caused by the infrequency of the data examined, in which case some adjectival translation units could possibly be identified in a larger sample set, or by the nature of adjectival phrases in themselves, in which case no translation units of this phrase type could be identified even in a larger set. Unlike the prepositional and verb phrases, this study cannot tell which one of these conjectures is most likely (see 7.4.2 for relevant discussion).

The last characteristic of the phrasal translation units was their grammatical structures (Table 7.5). All the 19 phrasal translation units can be divided into: noun plus noun(s), e.g. *a-h* and *k*; adjective plus noun(s), e.g. *i*, *l-s*; and, noun plus adverb, e.g. *j*. The first two structures were dominant; both structures were seen in 9 translation units each. The last one was scarce; it was seen in only one translation unit. This result reinforces what Newmark (1988b) claims; nominal phrase with 'adjective plus noun' is the most likely translation unit in languages. As far as the nine most frequent words in the ARC are concerned, this structure is indeed very dominant. However, what Newmark (1988b) misses out is that there is another nominal phrase which can be equally dominant: noun plus noun.

Table 7.5. Phrasal translation units

	Phrases	Structures
a.	***market*** *sources*	**N** + N
b.	***market*** *rumours*	**N** + N
c.	***market*** *economy*	**N** + N
d.	*stock* ***market***	N + **N**
e.	*gold* ***market***	N + **N**
f.	*copper* ***market***	N + **N**
g.	***market*** *interest rates*	**N** + N + N
h.	*Seoul stock* ***market***	N + N + **N**
i.	*domestic gold* ***market***	Adj + N + **N**
j.	*a* ***year*** *ago*	Det + **N** + Adv
k.	*marketing* ***year***	N + **N**
l.	***new*** *company*	**Adj** + N
m.	***new*** *orders*	**Adj** + N
n.	***new*** *investments*	**Adj** + N
o.	***new*** *low*	**Adj** + N
p.	***new*** *evidence*	**Adj** + N
q.	***new*** *prime minister*	**Adj** + N + N
r.	***new*** *crop months*	**Adj** + N + N
s.	***foreign*** *exchange reserves*	**Adj** + N + N

It is worth mentioning that some of the phrasal translation units were composed by three content words, e.g. *g-i* and *q-s* in Table 7.5. In particular, two of them were noteworthy: *Seoul stock market* and *domestic gold market*. Both contained smaller translation units within them. The former had the translation unit *stock market*; the latter had the translation unit *gold market*. This might be a characteristic of three-word translation units; however, this assumption requires further investigation on the rest of the four translation units: *market interest rates*; *new prime minister*; *new crop months*; and *foreign exchange reserves*. If the components such as *interest rates*, *prime minister*, *crop months*, and *exchange reserves* are translation units, this finding will be validated. However, none of them were the target words of this study; hence, this investigation was regarded to be out of the scope of this thesis. Hopefully this will be addressed in future work.

These three characteristics of phrasal translation units suggest that nominal phrases are very promising items to be translation units; therefore, they are good to focus on if one wants to extract translation units efficiently. This actually validates some studies that examine only nominal phrases, such as Wang (2006). Her target words were frequent two-word phrases composed of an adjective and noun; 30 of them were examined and 25 of them were identified as translation units. This is a very high rate, if one compares it with this study in which only 19 phrasal translation units were identified out of the examined 225 phrases (including nominal, adjectival, prepositional, and verbal phrases). Moreover, if one's corpus is tagged and aligned like Wang's corpus (HKLDC), one can save much time by focusing on only nominal phrases. A computer programme can identify all the nominal phrases in the corpus. After this automated process, one can simply carry out the one-equivalent principle on the identified nominal phrases; then, translation units will be extracted. This extraction process is likely to yield as many translation units as the one in which all types of phrases are extracted and examined, as this study did. This is because, as discussed above, prepositional, verbal, and adjectival phrases are very unlikely to be translation units. Nominal phrases are, therefore, 'guaranteed' items to focus on if one's aim is to extract many translation units within a short period.

7.2.3. Clauses

In the same way that the phrasal translation units had a predilection for one type of phrase, this study found that the clausal translation units also had such a predilection. All of the translation units at this level were subject-verb clauses (Table 7.1). Other

clausal types were not regarded as translation units: verb-object (e.g. *supported the market*) and subject-verb-object clauses (e.g. *the source told Reuters*). This dominance of the subject-verb clausal translation units can be partially explained by reference to Table 7.6; the subject-verb clauses were the most frequent in the sample sets; there were 50 of them examined in this study. The verb-object clauses occurred in the data sets less: 32 items in total were examined. Similarly, the subject-verb-object clause was infrequent; they were only 14 which appeared in the identification process (further discussion on why there were no translation units of verb-object and subject-verb-object types will be made in 7.4.5).

Table 7.6. Examined collocations (clauses)

	Nouns	Adjectives	Verbs	Total
Subject-verb	4 (1)	0	46 (2)	50
Verb-object	1	1	30	32
Subject-verb-object	0	0	14	14
Total	5	1	90	96

It is worth mentioning that Table 7.6 disagrees with part of what Newmark (1988b) claims. Newmark (1988b) maintains that one of the most likely translation units is 'verb plus object', since this is the most common collocation in languages. However, none of the translation units in this study had verb-object structure. Also, the verb-object sequence was not the most common collocation; the subject-verb was the commonest in this study. One possible reason behind such disagreement is the methodology used. Some collocations of verb-object structure were actually ignored; hence, they did not reach to the identification process stage. Example 7.1 shows one such.

Example 7.1. Verb-object clauses

```
1                     He also told the committee the dollar still had …
2                Oppenheim  told a parliamentary committee examining …
3        … Madeleine Albright told a Senate committee on Thursday …
4          … economics," George told parliament's Treasury select committee.
5 … statistics institute INSEE told the National Assembly finance committee.
```

All the five lines in Example 7.1 have *told* as a main verb and *committee* as its object;

therefore, they share the same verb-object collocation: *told-committee*. However, this

collocation was neglected in the analysis of *told* due to the extraction method (see

details in 3.4.2). All the examined collocations were retrieved by WordSmith; the

outcome collocations were the adjacent sequences of two to six words. Seeing that *told*

and *committee* are not adjacent items shown in Example 7.1, WordSmith failed to detect

this verb-object collocation *told-committee* as a candidate translation unit. The

disagreement between this study and Newmark (1988b) possibly derives from this

factor.

Even if WordSmith could have recognised the collocation *told-committee*, the set in

Example 7.1 would not be a good set for seeing if *told committee* could be a translation

unit. The one-equivalent principle requires one to examine if *told committee* was always

rendered into the same Japanese translation or not. Due to the extra items before

committee, Lines 2-5 would be unsuitable for carrying out this examination; it is hard to

see how *told committee* alone was rendered into Japanese in those samples. Therefore,

Lines 2-5 should have been discarded from the examined data of this collocation of

told-committee. Since Line 1 alone would not be frequent enough to be examined, *told*

the committee would have been dismissed. This means that both WordSmith and the

requirement for the one-equivalent principle prevented this study from examining verb-object collocations as often as they appeared in the sample data.

Such treatments, however, did not apply only to verb-object collocations; rather, they applied to subject-verb collocations as well. For example, all the lines in Example 7.2 have *said* as the main verb and *source* as its subject. Due to the lack of adjacency, *source-said* in Line 5 was not detected by WordSmith, but *source-said* in Lines 1-4 were extracted. Among the four, Lines 3-4 were discarded in the manual process, because of the existence of the modifier *floor*. This left Lines 1-2 for the examination of *source-said*, which did not reach to the minimum frequency for the identification; hence, this subject-object collocation *source-said* was ignored in the analysis.

Example 7.2. Subject-verb clauses

```
1    marketing of PCs in Korea," the source said.
2     on private-sector demand," the source said.
3        crop looks good," a floor source said.
4                        A floor source said U.S. cotton was pricing …
5                        The source also said that Japan's concern about …
```

Therefore, it is fair to interpret that the finding of this study regarding verb-object sequences disagrees with Newmark (1988b). The verb-object clauses were not common in the nine frequent words; and, they could not be translation units. This incompatibility between this study and Newmark (1988b) might be caused by the language pair. Most of the examples in the study of Newmark (1988b) were in English and French. Even though he gives the impression to discussing universal translation units, it is likely that his theory on translation units mainly comes from the English-French comparison. English and French share the similarities on positions of subject, verb, and object; while,

204

English and Japanese do not. The common structure regarding subject, verb, and object in English-French texts is not necessarily identical to what is common in English-Japanese texts. The differences between this study and Newmark (1988b) are therefore inevitable to some extent.

7.2.4. Single words

This empirical study also uncovered some characteristics of single word translation units. First, only nouns and adjectives could be translation units at the single word level. Verbs were not recognised as translation units as far as the three frequent ones are concerned. This is not actually too surprising if one recalls that a translation unit is a monosemous unit (see details in 3.4.1). All the verbs examined in this study are polysemous; they have many meanings. According to a comprehensive dictionary (**Collins Cobuild Advanced Learner's English Dictionary** 2003), *say* has 21 definitions; *tell* has 15 definitions; and *expect* has 6 definitions. This polysemous nature caused 52 different ways of rendering *said* in the 1,000 sample set, 47 different translations of *told*, and 62 translations of *expected*. Therefore, it was predictable that they failed to be labelled as monosemous translation units.

One might wonder what the Collins Dictionary (2003) tells us about the two single word translation units: *government* and *economic*. Table 7.7 shows how many definitions they have; the other target words and their definitions in the dictionary are also listed for the comparison. *Economic* and *government* actually have fewer meanings than the other seven target words. Both words have only two definitions in the

205

dictionary. This means, however, that they were not monosemous in the dictionary. Then, why did the one-equivalent principle recognise them as monosemous translation units? I went back to the sample concordance lines of *economic* and *government* and carried out an examination.

Table 7.7. Numbers of definitions in Cobuild (2003)

Target words	Numbers of definitions	
economic	2	
government	2	
foreign	4	
market	6	
expected	6	(*expect*)
new	8	
year	11	
told	15	(*tell*)
said	21	(*say*)

First, Collins Dictionary (2003) defines two meanings of *economic*: (a) '**Economic** means concerned with the organization of the money, industry, and trade of a country, region, or society [emphasis in original]'; and, (b) 'If something is **economic**, it produces a profit [emphasis in original]'. The former tends to occur in nominal phrases, e.g. *The pace of economic growth* (from Collins, 2003); while, the latter does not, e.g. *The new system may be more economic* (from Collins, 2003). Most occurrences of *economic* in the sample set actually appeared in nominal phrases, which make them belong to the former type. This means that the meaning of *economic* in the set was actually only one; e*conomic* was, therefore, monosemous in the 1,000 sample examples. Second, if one looks at the 1,000 sample examples, *government* was monosemous as well. Collins Dictionary (2003) tells us that *government* refers to: (a) 'The **government** of a country is the group of people who are responsible for governing it', e.g. *fighting between government forces and left-wing rebels*; or, (b) '**Government** consists of the

activities, methods, and principles involved in governing a country or other political unit', e.g. *The first four years of government were completely disastrous*. The former occurred dominantly in the set, e.g. t*he new government*, *the next government*, *the government and political parties*, and *provided by the government*. Therefore, again, *government* was monosemous in this study.

These indicate that one can predict single word translation units by looking them up in a dictionary to some extent. For example, if this study were extended by focusing on the 17 most frequent nouns, there are seven more words which are likely to be identified as single word translation units due to their fewer definitions: *prices*; *traders*; *tonnes*; *dollar*; and, *futures*. They have two or less definitions in the dictionary (Collins, 20003); for *prices*, *traders*, and, *tones*, the present forms were consulted. Of course, monolingual dictionaries are not always useful. For example, there is the minor case in which a word is monosemous in a monolingual dictionary, but it is not from a bilingual perspective. Such cases do not allow the word to be single word translation unit. One classic example of this is the English word *bone* (Teubert, 2004b). According to Teubert (2004b), most monolingual English dictionaries tell that *bone* has one meaning; while, it has three different German translations, depending on if the *bone* is fish or non-fish related, etc. Therefore, even if a given word is monosemous in a dictionary, it is still possible that it cannot be regarded as a monosemous translation unit. However, this monosemy checking in a monolingual dictionary helps one to make a relatively promising hypothesis of single word translation units, as shown in Table 7.7. One drawback is that this method is possible only when one's target words are single words: there are no phrasal or clausal dictionaries as comprehensive as word-based ones.

7.2.5. Four variables

The finding of the translation unit sizes can be compared with the literature on the topic. The main focus here is to find out if this empirical study brings new insight to the issue of translation unit size or not. The comparison was carried out along the four variables: lexically defined translation units; English-Japanese comparison; newswire texts; and, experienced translators. First, according to Bennett (1994), Newmark (1988b), and Vinay and Darbelnet (1995), the lexical translation units tend to be the sizes of morpheme, word, phrase, clause, and sentence (Figure 7.1). This study overlapped with their claims; the identified sizes in this study were single word, phrase, and clause. One might ask why this study could not identify translation units at the morpheme and sentence levels. This was a methodological issue, as discussed in 3.4.8 for the former level and in 3.5 for the latter.

Figure 7.1. Translation unit size and definitions

Studies	Definitions	mo.	w.	ph.	cl.	sen.	par.	tex.
Bennett (1994)	Lexical							
Newmark (1988b)	Lexical							
Vinay and Darbelnet (1995)	Lexical							
Teubert (2001)	Lexical							
Toury (1986)	Lexical							

On the other hand, Teubert (2001) and Toury (1986) make claims for more restricted translation unit sizes than Bennett (1994), Newmark (1988b), Vinay and Darbelnet (1995), and the current study (Figure 7.1). Teubert (2001) identifies only single word and phrasal translation units; while Toury (1986) recognizes only phrasal and clausal translation units. Does this mean that Teubert (2001) and Toury (1986) disagree with the

208

current study which found translation units at three levels (i.e. single word, phrase, and clause)?

The findings of Teubert (2001) and Toury (1986) *are* compatible with this study. First, the difference between Teubert (2001) and this study derives from the methods used. Teubert (2001) consulted a bilingual dictionary in his identification of translation units. He examined newspaper texts to see if one could translate a word by looking up its translation in a bilingual dictionary (see details in 2.4.1). Considering that bilingual dictionaries normally have only single words and phrases as entries, this method allowed him to identify only single word and phrasal translation units. It was inevitable that his findings could not include clausal translation units. The zero existence of clausal translation units was, therefore, caused by this methodological drawback; hence, the result of Teubert (2001) is not in direct conflict with this study.

Second, the difference between this study and Toury (1986) disappears if one has a close look at his experimental data. Toury (1986) maintains that translation units are usually at the phrasal and clausal levels; however, he does not deny the possibility of single word translation units. Indeed, his data (1986) show that there were a few translation units at the single word level: the English word *astronomy* was rendered into its Hebrew correspondent *astronomya*, and the English word *studying* was rendered into the Hebrew equivalent *ve-lamadti*. Therefore, the single word translation unit was recognized as a very minor part of the lexical translation units in his study. Therefore, none of the literature of lexical translation units shown in Figure 7.1 disagrees with the findings of this empirical study.

What this study brings to the literature is empirical evidence. Bennett (1994) and Newmark (1988b) are theoretical linguists who make claims for their translation unit sizes without much linguistic evidence. This study provided corpus evidence for their claims by agreeing with their results. The other three studies used subjective criteria to identify translation units. This study identified translations in a more objective way; the findings overlapped with the studies of Toury (1986), Vinay and Darbelnet (1995), and Teubert (2001), which means that their results were validated with more evidence. Therefore, the sizes of lexical translation units are theoretically and empirically examined, and these views are in agreement. Lexical translation units are likely to be single words, phrases, and clauses.

Another set of literature can be compared with this study from the perspective of text type. This study used a corpus of newswire texts for identifying translation units. There are two empirical studies in which translation units were examined in news texts (Figure 7.2): Teubert (2001) and Alves and Gonçalves (2003). The former, as discussed above, identifies translation units at the level of the single word and phrase; the latter identifies translation units at the levels of phrase only. Both studies overlap with the findings of this study. The only difference between this study and Teubert (2001) is the point of clausal translation units; the reason for this was discussed above. The difference between this study and Alves and Gonçalves (2003) is presence of translation units at the levels of single word and clause in this study. This is also due to a methodological issue. Alves and Gonçalves (2003) examined only one sentence: *By 1998, the year of the U.S. Embassy bombings in Africa, he had acquired the lean, wolfish look of a revolutionary* (see details in 2.2.2). Seeing that the clausal and single word translation

units were minor sizes even in the present study, where 9,000 sentences were examined, it is not surprising that Alves and Gonçalves (2003) could not locate any translation units at these levels.

Figure 7.2. Translation unit size and text types

Studies	Text types	mo.	w.	ph.	cl.	se.	par.	tex.
Tancock (1958)	Technical documents	■						
Teubert (2001)	Newspaper texts	■	■					
Alves and Gonçalves (2003)	News magazines		■					
Toury (1986)	TV scripts			■	■			
Vinay and Darbelnet (1995)	Academic articles			■	■			
Tancock (1958)	Literature					■	■	
Barkhudarov (1993)	Poetry							■

What this study provides to the literature of news texts is this: translation unit sizes can be clauses, as well as single words and phrases, in news texts. Neither Teubert (2001) nor Alves and Gonçalves (2003) managed to identify this size, because of methodological drawbacks. However, this study overcame the difficulties of setting up an objective methodology and succeeded in examining clausal translation units. This finding is actually reasonable, if one thinks that news texts use the same clauses repeatedly. All the clausal translation units identified in this study are good examples of this: e.g. *market will reopen*; *Shanghai Securities News said*; and, *dealers expected*. All of them are frequent clauses in news texts; due to this repeated usages, their translations are likely to be established already in the translation community. Therefore, such frequent clauses are likely to be translation units.

Third, a comparison can be made from the perspective of translators as well. This study used translation data rendered by experienced translators. The relevant literature were:

Barbosa and Neiva (2003); Toury (1986); and, Vinay and Darbelnet (1995); all shown in Figure 7.3. All overlapped with the finding of this study in their assessments of translation unit size. The only difference is that this study identified single word translation units. However, if one looks back to the literature carefully, none of these authors actually deny the existence of single word translation units. For example, Toury (1986)'s study, as discussed above, actually identified a few single word translation units, e.g. *astronomy* and *studying*. Vinay and Darbelnet (1995) examined a paragraph of an academic article and identified 51 translation units. Although most of them were at the levels of phrase and clause, there was one single word translation unit: the French word *commençant* rendered into the English word *starting*. Similarly, Barbosa and Neiva (2003) state that '[f]ewer of the units … consisted exclusively of single words'. All the literature, therefore, agrees with this study: single word translation unit is a minor, but existent, size for experienced translators. Therefore, this study reinforced the findings of the literature: for some occasions, experienced translators do use single words as translation units.

Figure 7.3. Translation unit size and translators

Studies	Translators	mo.	w.	ph.	cl.	sen.	par.	tex.
Barbosa and Neiva (2003)	Experienced			██	██	██		
Lörscher (1991)	Experienced			██	██			
Toury (1986)	Experienced			██	██			
Vinay and Darbelnet (1995)	Experienced			██	██			

Lastly, this study extracted translation units for an English-Japanese comparison. It is now appropriate to discuss if Koller's view can be validated or not (see details in 2.4.2): 'a translation between unrelated languages will usually involve larger units than if SL and TL [=target language] are closely related' (cited and translated in Shuttleworth and

Cowie, 1997). As discussed in 2.4.2, this view accords relatively well when the language pairs are both Indo-European languages: English-German; English-French; and, English-Portuguese in Figure 7.4. However, Koller's claim is not validated when the language pairs involve one outside of the Indo-European language family. For example, Toury (1986)'s English-Hebrew study identifies phrasal and clausal translation units. Since Hebrew is not an Indo-European language, the translation units in this language pair should be larger than the ones in a closer language pair, e.g. English-German. However, the identified translation unit sizes of Toury (1986), as shown in Figure 7.4, are no larger than those which most of the studies of Indo-European languages identified.

Figure 7.4. Translation unit size and language pairs

	Language pairs	mo.	w.	ph.	cl.	sen.	par.	tex.
Goldman-Eisler (1972)	English-German				�			
Lörscher (1991)	English-German			▇	▇			
Goldman-Eisler (1972)	English-French				▇			
Teubert (2001)	French-English		▇	▇				
Alves and Gonçalves (2003)	English-Portuguese				▇			
Barbosa and Neiva (2003)	English-Portuguese				▇	▇		
Toury (1986)	English-Hebrew				▇			

The English-Japanese comparison is another language pair involving a non-Indo-European language; Japanese is not related to the Indo-European language family (Miller, 1980; Masaomi Kondo and Wakabayashi, 1998; Shibatani, 2006). This study identified translation units at single word, phrase, and clause, which are similar to most of the literature on Indo-European languages in Figure 7.4. Therefore, Koller's claim was not supported by this empirical study either. Koller's view is only validated within the Indo-European language family, but not outside of this family.

7.2.6. Shortcomings

This study identified three sizes of translation units: single word, phrase, and clause. This finding, however, does not deny the possibility that other sizes can be translation units in the ARC. This is because this study examined only the three sizes stated; while the other possible sizes, i.e. morpheme, sentence, paragraph, and text, could not be examined due to the methodology used (discussed in 3.4.8). Therefore, the result cannot tell us about the latter four sizes in a strict or direct sense.

However, the findings of this study can give some hints about the four sizes which were not identified as translation units in this study, if one compares them with the literature. According to the linguists who identified more than one translation unit size (Toury, 1986; Barkhudarov, 1993; Vinay and Darbelnet, 1995; Teubert, 2001; Barbosa and Neiva, 2003), the minor sizes are always one size smaller and/or larger than the dominant ones. For example, Barbosa and Neiva (2003) maintain that the most dominant size is the phrase, followed by clauses, then, lastly, single words. The minor sizes are immediately smaller and larger than the dominant size. Similarly, this study can be extended along the same lines; the dominant size was the phrase, whilst the minor sizes were single words and clauses, which were one size smaller and larger. No literature shows that minor sizes appear much smaller or larger than the dominant sizes.

Appling this to the discussion of the four sizes (i.e. morpheme, sentence, paragraph, and text), it is possible to make some assumptions about their possibilities to be translation units. First, the translation units at the sentence level are expected to occur rarely in the

214

ARC. This is because the dominant size in this study was phrase. Seeing that single word and clause are the sizes which are immediately smaller and larger, units at the sentence level are unlikely to be another minor size of translation unit in the ARC; this goes for even larger sizes as well, i.e. paragraph and text. Second, the translation units at the morpheme level are expected to be very rare as well. Morpheme is not at the level one size smaller than the dominant translation unit size identified in this study. Even if this study had examined some morphemes, it would be unlikely to identify some translation units at this level. Therefore, it is reasonable to conclude that single word, phrase, and clause are the main three sizes of translation units in the ARC. The other four translation unit sizes are unlikely to occur in the corpus.

7.3. Translation pairs

Along with the translation unit sizes, this study identified the translation equivalent sizes in the corpus. Table 7.8 presents all the identified translation equivalents in this study; their translation units are shown in parentheses (target words and their corresponding parts are shown in bold). What Table 7.8 shows is that the translation equivalents are not necessarily at the same size as the translation units; e.g. the phrase *a year ago* corresponded to the single word *zennen*. Some translation pairs underwent shifts in translations. Due to such shift occurrences, the frequencies of the translation equivalents at each size were different from the ones of the translation units at each size. There were 19 phrasal translation units; while, there were 15 phrasal translation equivalents. Similarly, there were only two single word translation units; while, six translation equivalents were recognized at this size. This made the single word

translation equivalents the second dominant size and the clausal ones the least dominant

size, which was a different result from that of the translation unit sizes.

Table 7.8. Identified translation equivalents and their sizes (i)

	Single words	Phrases
1.	*zennen* (*a year* ago)	*shijoo* suji (*market* sources)
2.	*seefu* (*government*)	*shijoo* no uwasa (*market* rumours)
3.	*keezai* (*economic*)	*shijoo* keezai (*market* economy)
4.	*shin* gaisha (*new* company)	kabushiki *shijoo* (stock *market*)
5.	*shin* yasune (*new* low)	kin *shijoo* (gold *market*)
6.	*shin* shushoo (*new* prime minister)	doo *shijoo* (copper *market*)
7.		*shijoo* kinri (*market* interest rates)
8.		sooru kabushiki *shijoo* (Seoul stock *market*)
9.		kokunai kin *shijoo* (domestic gold *market*)
10.		shijoo *nendo* (marketing *year*)
11.		*shinki* juchuu (*new* orders)
12.		*shinki* tooshi (*new* investments)
13.		*aratana* shooko (*new* evidence)
14.		*shin*koku gengetsu (*new* crop months)
15.		*gai*ka junbi (*foreign* exchange reserves)

	Clauses
1.	*torihiki* wa saikai sa reru (*market* will reopen)
2.	shanhai shooken hou ni *yoru* to (Shanghai Securities News *said*)
3.	diiraa ra wa ... *yosoo shi te iru* (dealers *expected*)

One advantage of this empirical study was to be able to investigate which translation

pairs underwent shifts and which ones did not. Three shifts were focused on in this

study: unit-shifts; structure-shifts; and, class-shifts. The analyses showed that more than

half of the 24 translation pairs required one, or more than one, shift (Table 7.9, *k-y*). On

the other hand, the rest of the ten translation pairs did not undergo any shifts (Table 7.9,

a-j). The comparison between the translation pairs in the two groups revealed four

characteristics of shifts in the ARC: (a) the shifts did not occur in the N + N translation

units, e.g. *a-g*; (b) the shifts occurred in the translation units which have at least one

adjective, e.g. *l-t*; (c) the shifts always occurred in the clausal translation units, e.g. *v-y*;

216

and, (d) the shifts occurred in most of the three-word translation units, e.g. *k-o*. Considering that the shifts were seen where 'linguistic incompatibilities between SL [=source language] and TL[=target language]' lie (Shuttleworth and Cowie, 1997), these four characteristics can be taken to show which linguistic features are compatible and incompatible between English and Japanese.

Table 7.9. Translation units and their shifts occurrences

	Translation units	Unit	Structure	Class
a.	***market** sources*	not shifted	not shifted	not shifted
b.	***market** rumours*	not shifted	not shifted	not shifted
c.	***market** economy*	not shifted	not shifted	not shifted
d.	*stock **market***	not shifted	not shifted	not shifted
e.	*gold **market***	not shifted	not shifted	not shifted
f.	*copper **market***	not shifted	not shifted	not shifted
g.	*marketing **year***	not shifted	not shifted	not shifted
h.	*Seoul stock **market***	not shifted	not shifted	not shifted
i.	***government***	not shifted	not shifted	not shifted
j.	***new** evidence*	not shifted	not shifted	not shifted
k.	***market** interest rates*	not shifted	shifted	not shifted
l.	*domestic gold **market***	not shifted	shifted	not shifted
m.	***foreign** exchange reserves*	not shifted	shifted	not shifted
n.	***new** crop months*	not shifted	shifted	not shifted
o.	***new** prime minister*	shifted	shifted	not shifted
p.	***new** company*	shifted	shifted	not shifted
q.	***new** low*	shifted	shifted	not shifted
r.	***new** orders*	not shifted	shifted	not shifted
s.	***new** investments*	not shifted	shifted	not shifted
t.	***economic***	not shifted	shifted	shifted
u.	*a **year** ago*	shifted	shifted	not shifted
v.	***market** will reopen*	not shifted	shifted	-
w.	*Shanghai Securities News **said***	not shifted	shifted	-
y.	*dealers **expected***	not shifted	shifted	-

7.3.1. Shift occurrences

The first characteristic of the shift occurrences tells us that N + N phrases are compatible between English and Japanese. This means that what is expressed in this form in English was expressed in the same form in Japanese. For example, the N + N

phrasal translation unit *market sources* was rendered into the N + N phrasal translation equivalent *shijoo suji* without changing its structure (i.e. two-word phrases composed by two nouns), size (i.e. phrase), and class (i.e. nominal phrase). This was the only compatibility that this study identified between English and Japanese. The other characteristics of the shift occurrences, on the other hand, show incompatibilities between the two languages. First, the adjective usage was found to be incompatible, e.g. *l-t*. English adjectives were rarely rendered into the corresponding Japanese adjectives, as far as the most frequent adjectives were concerned. The translation unit (*i*) *new evidence* (Table 7.9) is the only example in which the English adjective *new* corresponded to the Japanese adjective *aratana*. All the other adjectives in translation units corresponded to Japanese nouns, e.g. *l*, *r-s,* or prefixes, e.g. *m-q*. The adjectives are incompatible, causing structure-shifts in most cases.

Second, the structures of clauses were found to be another incompatibility between English and Japanese. All the clausal translation units underwent structure-shifts, e.g. *v-y* (Table 7.9). The clausal translation unit (*v*) *market will reopen* shows that the future tense is not compatible from English to Japanese, which reinforces Kindaichi's claim (1988) that the Japanese future tense is not used as often as the English one. Unless it is necessary, the future tense does not appear in Japanese, which causes the change in structure. The translation unit (*w*) *Shanghai Securities News said* demonstrates that a subject in English becomes an object in Japanese. *Shanghai Securities News* had a subject status in the translation unit; while it turned to be an object in its translation equivalent *shanhai shooken hoo ni yoru to* 'based on *Shanghai Securities News*'. The last clausal translation unit (*y*) *dealers expected* demonstrates that the verb position is

incompatible. As discussed in 3.5, English verbs appear after the subject; while, Japanese verbs appear around the end of the sentence. Therefore, this incompatibility seems inevitable.

The other identified incompatibility was the structures of three-word phrases, e.g. *k-o*. Most of the three-word translation units became two-word equivalents: (*k*) *market interest rates* into *shijoo kinri*; (*m*) *foreign exchange reserves* into *gaika junbi*; (*n*) *new crop months* into *shinkoku gengetsu*; and, (*o*) *new prime minister* into *shin shusho* (Table 7.9). The Japanese equivalents were all two-word phrases composed by two nouns, i.e. N + N phrases. This was because these translation units contained words such as *interest rates*, *foreign exchange*, *new crop*, and *prime minister*, which are represented by one word each in Japanese: *kinri*, *gaika*, *shinkoku*, *shusho*, respectively. On the other hand, there were two three-word phrases which turned into three-word equivalents: (*l*) *domestic gold market* into *kokunai kin shijoo* and (*h*) *Seoul stock market* into *sooru kabushiki shijou*. The three-word structure remained through translation; however, these were relatively minor cases. Therefore, it is possible to assume that the three-word phrases are often incompatible between English and Japanese.

This discussion shows that translation pairs are useful resources for finding out such compatibility and incompatibilities between two languages. The N + N phrase belongs to the former category of compatibility; while, the adjectival usage, clausal structure, and three-word phrases belong to the latter. It is likely that translation equivalence is established when a given translation unit is N + N phrase. On the other hand, translation equivalence is not likely to be established when a given translation unit is a clause,

three-word phrase, or unit which contains one adjective or more. This finding can trigger an interesting, cross-linguistic discussion, if one compares these results with studies on the other language pairs, which is a part of my on-going interest, as mentioned in 3.2.

7.3.2. Variations of translation units

A translation unit is an inseparable unit to be rendered as a whole. In order to extract such units objectively, this study introduced the one-equivalent principle, which made the definition of translation unit narrower than the one above: a translation unit is a monosemous unit which has only one translation equivalent in the corpus (see the details in 3.4.8). This means that a translation unit has a one-to-one relationship with its translation equivalent. For example, the translation unit *market sources* corresponded to the equivalent *shijoo suji* (Figure 7.5). No other items got involved in this translation pair. However, the analyses identified some translation units that had several variations, and established many-to-one relationships with their equivalents. For example, the translation unit *market economy* had the five variations sharing the same rendering: *shijoo keezai* (Figure 7.5; '~' stands for a modifier insertion).

Figure 7.5. Translation equivalence: *market sources* **and** *market economy*
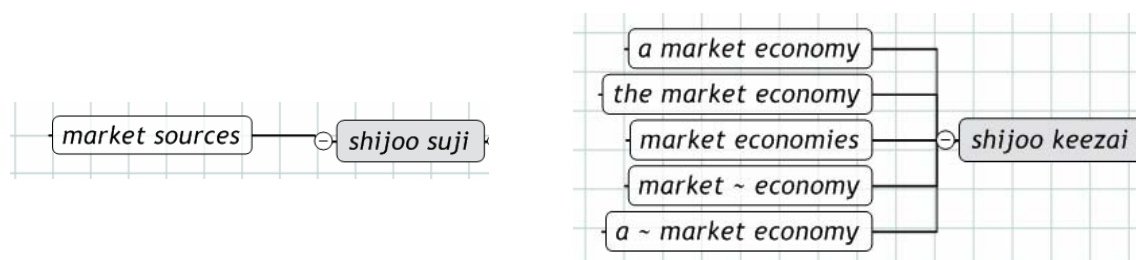
Among all the 24 translation pairs, which type of translation equivalence was seen more, one-to-one or many-to-one? Figure 7.6 is a summary of all the findings related to translation equivalence. The pairs at the left side had one-to-one translation equivalence; while, the pairs in the middle and right side had many-to-one translation equivalence. The numbers in parentheses show how many variations each translation unit has, e.g. the translation unit *new evidence* had two variations, which made it possible to establish a two-to-one relationship with its equivalent *aratana shooko*. Figure 7.6 shows that there were only six translation pairs which established a one-to-one translation equivalence. On the other hand, the rest of the 19 translation pairs exhibited a many-to-one translation equivalence. Therefore, most of the translation equivalence in the ARC was established at the many-to-one basis, as far as the frequent content words are concerned.

Figure 7.6. Translation equivalence

One-to-one   Many-to-one

market sources (1)	*economic* (3)	*government* (6)
marketing year (1)	*market rumours* (3)	*market economy* (5)
Seoul stock market (1)	*new company* (3)	*stock market* (4)
new crop months (1)	*new prime minister* (3)	*gold market* (4)
foreign exchange reserves (1)	*dealers expected* (3)	*new orders* (4)
a year ago (1)	*copper market* (2)	*Shanghai Securities*
	new investments (2)	*News said* (4)
	new low (2)	
	new evidence (2)	
	market interest rates (2)	
	domestic gold market (2)	
	market will reopen (2)	

One of the frequently occurring variations was article variation. The analyses showed that most translation units were resistant to article changes; the changes did not cause different translations. For example, *gold market*, *a gold market*, and *the gold market* were all rendered into *kin shijoo*. No matter which articles appeared before a given translation unit, its translation was not changed. Similar cases can be found in many translation units, e.g. *a-k* in Table 7.10 ('-' means that it was not examined because they occurred with no articles). This happened because definite and indefinite articles are not usually translated into Japanese (**English-Japanese Dictionary For The General Reader**, 1999; **Genius English-Japanese Dictionary**, 2001; **Lighthouse English-Japanese Dictionary**, 2007). Therefore, translation of Japanese nouns is often carried out by rendering lexical words such as *gold market*, ignoring articles *the* and *a*.

Table 7.10. Article variations

	Translation units	Article
a.	*market rumours*	resistant
b.	*market economy*	resistant
c.	*gold market*	resistant
d.	*domestic gold market*	resistant
e.	*government*	resistant
f.	*economic*	resistant
g.	*new company*	resistant
h.	*new orders*	resistant
i.	*new evidence*	resistant
j.	*new prime minister*	resistant
k.	*Shanghai Securities News said*	resistant
l.	*stock market*	(resistant)
m.	*copper market*	(resistant)
n.	*Seoul stock market*	(resistant)
o.	*market will reopen*	(resistant)
p.	*marketing year*	(resistant)
q.	*new low*	(resistant)
r.	*a year ago*	not resistant
s.	*market sources*	-
t.	*market interest rates*	-
u.	*new investments*	-
v.	*new crop months*	-
w.	*foreign exchange reserves*	-
y.	*dealers expected*	-

The translation units (*l*) *stock market*, (*m*) *copper market*, (*n*) *Seoul stock market*, (*o*) *market will reopen*, and (*p*) *marketing year* did not give concrete evidence of the same claim. They always appeared with the definite articles in the sample data, which made it impossible to examine if their article changes caused different translations or not. However, closer examination on the concordance lines presented that *the* were not rendered at all in the translations. Due to this zero translation of their articles, they were assumed to be resistant to article changes as well. The same can be said in the case of (*q*) *new low*; it occurred with the indefinite article, which was lost in translation. Hence, the translation units (*l*)-(*q*) support the same finding of the ones (*a*)-(*k*); article changes are unlikely to cause different translations. Thus, most of the translation pairs actually suggest that article changes did not matter in translation. This was one contributing

factor to the finding that a substantial amount of translation pairs were identified as having many-to-one translation equivalence.

Another frequently occurring variation was the modifier variation: a modifier appearing before a translation unit. As shown in Table 7.11, nine translation units proved that their modifier occurrence did not cause different translations. For example, *market interest rates* was always rendered into *shijoo kinri* no matter whether it occurred alone, e.g. (*c*) *market interest rates*, or with the modifier, e.g. (*c'*) *capital market interest rates*. The modifier did not matter in rendering the translation unit. On the other hand, the translation units (*h*) *market rumours* and (*i*) *copper market* demonstrated that the modifiers affected the rendering of the translation units in some cases. However, they remained as minor cases among the 24 translation units.

Table 7.11. Modifier variations (before translation units)

	Translation units	Modifier (before)
a.	*market economy*	resistant
b.	*stock market*	resistant
c.	*market interest rates*	resistant
d.	*government*	resistant
e.	*new orders*	resistant
f.	*Shanghai Securities News said*	resistant
g.	*dealers expected*	resistant
h.	*market rumours*	resistant/ not resistant
i.	*copper market*	resistant/ not resistant

This resistance of translation units was unexpected. Let's look back at how an item was identified as a translation unit in this study. Lines 1-3 (Example 7.3), for example, are some concordance lines where the item *stock market* was examined to see if it could be

a translation unit or not. *Stock market* in Lines 1-3 were all rendered into the same translation, which made the item to be labelled as a translation unit. This suggests that the translation of *stock market* should be always the same, wherever it occurs and whatever it co-occurs with. *Stock market* might appear at the end of the sentence; it might appear with different verbs from the ones in Lines 1-2 and with different prepositions from the one in Line 3. Yet, *stock market* should be rendered into the same translation. That is what a translation unit is in this study. However, I assumed that the translation of *stock market* might differ in a context like Line 4, the extra modifier *red-hot* inserted within the phrase. This assumption was grammatically motivated. *Stock market* is a phrase. When an extra item *red-hot* appears in the phrase, the phrasal structure and meaning have changed. How is it possible that *stock market* in Line 4 still maintains the same translation as the one in Lines 1-3?

Example 7.3. Different context of *stock market*

```
1                              "The stock market is up and that's  probably …
2   … percent to help support the stock market, the Shanghai Securities News …
3    … raise more money from the stock market, it said.
4      … to cool down the red-hot stock market, though analysts said it had …
```

Therefore, the result seen in Table 7.11 was rather surprising. It indicates that such grammatically closely related modifiers were unable to affect the rendering of translation units. A translation unit, therefore, is an unaffected unit. The translation equivalence between translation unit and its equivalent is robust. It is not broken by any co-occurring elements in most of the time. It is important to point out, however, that this study examined cases of one modifier insertion. Two-modifier insertions such as *the prevailing sluggish stock market* (Line 1, Example 7.4) and *the island's rallying stock*

market (Line 2) were not investigated. Whether they could break the translation equivalence or not requires further research.

Example 7.4. Two-modifier insertion of *stock market*

```
1    … because of the prevailing sluggish stock market.
2       … on Friday the island's rallying stock market was healthy and that …
```

The next most frequently occurring variations were the singular and plural variations. These were seen in six translation units, *a-f* (Table 7.12). For example, (*a*) *market rumours* and its singular variation (*a'*) *market rumour* corresponded to the same translation equivalent *shijoo no uwasa*. This could happen because Japanese nouns do not normally express the difference between one and more than one (Kindaichi, 1988; Baker, 1992; Bunt, 2003; Makino and Tsutsui, 2008); hence, *shijoo no uwasa* can mean both *market rumours* and *market rumour*. Considering this nature of nouns, it is not too surprising that singular and plural variations often shared the same translation. However, Table 7.12 shows that the four translation units were not, in fact, resistant to the singular/plural changes, e.g. *g-j*. For example, the translation unit (*g*) *market sources* and its singular variation (*g'*) *a market source* did not share the same translation. The singularity of (*g'*) *a market source* was expressed in the translation: *aru* 'a certain'. The pilot study identified that the indefinite article tended to be rendered into *aru* when a given noun denotes people such as *source*. Similarly, the translation unit (*i*) *new company* and the plural variation (*i'*) *new companies* did not share the same translation. The plurality of (*i'*) *new companies* was expressed in the translation: *kakusha* 'each company'. These indicate that Japanese nouns express the singularity and plurality in

226

some rarer cases. When this happened, however, has not been thoroughly investigated at this stage.

Table 7.12. Singular and plural variations

	Translation units	Singular/plural
a.	*market rumours*	resistant
b.	*new investments*	resistant
c.	*market economy*	resistant
d.	*stock market*	resistant
e.	*market will reopen*	resistant
f.	*government*	resistant/ not resistant
g.	*market sources*	not resistant
h.	*a year ago*	not resistant
i.	*new company*	not resistant
j.	*new low*	not resistant

The least occurring variation was the modifier variation where a modifier appears in the middle of a translation unit. This was seen only in the three translation units, e.g. *a-c* (Table 7.13). For example, *market economy* was always rendered into *shijoo keezai* in the context of (*a*) *market economy* alone, as well as in the context of the modifier variation (*a'*) *market sector economy*. The intervening modifier *sector* did not cause a different translation. The same resistance was seen in the analyses of (*b*) *new low* and (*b*) *new orders* as well. What these demonstrated was, again, the robust relationship that the translation units have to the equivalents. Even though the elements in the translation units were separated by a modifier insertion, the rendering of the translation unit remained the same. Adjacency did not matter. On the other hand, there are more examples which showed they were not resistant to this type of modifier change. For example, (*d*) *new prime minister* corresponded to *shin shushoo*; however, when the modifier *first* intervened, i.e. (*d'*) *new first prime minister*, the translation of *new* was changed. Adjacency between the elements in translation units mattered in this

227

translation pair. A similar nature was found in the translation units (*b*) *new low*, (*c*) *new orders*, (*e*) *gold market*, (*f*) *new company*, and (*g*) *dealers expected*.

Table 7.13. Modifier variations (intervening)

	Translation units	Modifier (middle)
a.	*market economy*	resistant
b.	*new low*	resistant/ not resistant
c.	*new orders*	resistant/ not resistant
d.	*new prime minister*	not resistant
e.	*gold market*	not resistant
f.	*new company*	not resistant
g.	*dealers expected*	not resistant

In summary, this investigation identified that among the 24 translation pairs, six had one-to-one translation equivalence and 19 had many-to-one translation equivalence. The latter occurred due to the nature of resistance that some translation units have. The analyses examined the four types of resistance: article change, singular/plural change, and two kinds of modifier changes. The resistance to article changes was most supported by the translation pairs; while resistance to the intervening modifier was least supported by the translation pairs.

7.4.   Non-translation units

The previous section (7.3) has demonstrated that translation equivalence is a special relationship between a translation unit and its equivalent. It is robust; most of the pairs are resistant to some changes. And it is recurrent; wherever it occurs, the translation unit is always rendered into the translation equivalent. This section will examine the items which were not regarded as translation units. In other words, the focus is the items

which could not establish such translation equivalence with any Japanese expressions. The main query will be to investigate why they failed to be translation units.

7.4.1. Prepositional phrases

First, this study could not identify any prepositional phrases as translation units (see 7.2.2 for relevant discussion). The analyses examined 54 prepositional phrases in total. The prepositions that the 54 phrases contained were: *in*; *of*; *for*; *from*; *with*; *by*; *on*; *over*; *about*; *as*; *into*; *out of*; *per*; *since*; *than*; and, *to* (in frequency order). Table 7.14 shows how many prepositional phrases of each target word were investigated (*government* and *economic* are not covered, because they were single word translation units and the examinations at the phrasal level were not carried out). *Year* was the target word for which the most prepositional phrases were examined: 29 phrases, such as *of this year*, *per year*, and *for the year*. *Said*, *told* and *expected* were the words for which no prepositional phrase was found.

Table 7.14. Examined prepositional phrases

Target words	Prepositional phrases
year	29
market	12
foreign	10
new	3
said	0
told	0
expected	0
Total	54

The analyses of the prepositional phrases identified three ways in which they could fail to be translation units. The first case can be seen in the examples of *from last year*. This

phrase appeared four times in the sample data of *year*. All of them are shown in

Example 7.5 (the words in parentheses are the translations of the underlined items;

words in bold correspond to the prepositions).

Example 7.5. Non-translation units: *from last year*

```
1   … about 200,000 tonnes from last year, the China …     (zennen kara)
2   … high stocks left over from last year.                (zennen kara)
3    … to be little changed from last year, the official … (zennen to)
4  … by 23.2 million tonnes from last year as a result of … (zennen o)
```

In all the concordance lines, the translations of *last year* were identical: *zennen*.

However, the translations of *from* were variable. It was rendered in three different ways:

kara 'from' in Lines 1-2; *to* 'with' in Lines 3; and, *o* as an object marker in Line 4. It

was because of this diversity of translation that the one-equivalent principle could not

identify the phrase *from last year* as a translation unit. A similar case was seen in the

other prepositional phrases as well, e.g. *to the market* (Example 7.6). Again, *the market*

was rendered into *shijoo* all the time; however, the translations of *to* were diverse: *ni* 'at'

in Lines 1-2; *e* 'to' in Line 3; and, *o* as an object marker in Line 4.

Example 7.6. Non-translation units: *to the market*

```
1      … foreigners, returned to the market today and bought …   (shijoo ni)
2 … 75,000 barrels per day on to the market for the remainder … (shijoo ni)
3    … the return of Iraqi oil to the market would not hurt price. (shijoo e)
4    … 1.60 marks was a magnet to the market," said Seth Garrett,… (shijoo o)
```

In the examples of *from last year* and *to the market*, the prepositions were translated

into Japanese on every occasion. However, there were some cases where this did not

happen, e.g. *of next year* (Example 7.7). Again, the translations of *next year* were *rainen*

in all the concordance lines. However, the translations of *of* were not the same between the five examples. In Lines 1-2, *of* corresponded to *no* 'of'. On the other hand, in Lines 3-5, *of* did not correspond to any items in the translations. *The second half of next year* in Line 3 was rendered into *rainen koohan* 'late next year'; *of* was not rendered. Similarly, *the beginning of next year* in Line 4 and *the first half of next year* in Line 5 were translated without *of*: *rainen hajime* 'early next year, *rainen kamihanki* 'first half next year', respectively. Due to these zero translations of the prepositions, the phrase *of next year* was not identified as a translation unit.

Example 7.7. Non-translation units: *of next year*

```
1        … or the first half of next year," Huang was …    (rainen no)
2       … published the first of next year, said a USDA …  (rainen no)
3        … by the second half of next year.                (rainen)
4          … by the beginning  of next year.               (rainen)
5     … during the first half  of next year and $420 in the … (rainen)
```

The third case was seen in the prepositional phrase *by foreign investors* (Example 7.8). The translations of the preposition *by* remained the same: *niyoru* 'by'. However, the translations of the nominal phrases *foreign investors* were variable; in particular, the adjective *foreign* was rendered differently. It corresponded to *kaigai* 'foreign' in Line 1, *gaikokujin* 'foreigner' in Line 2, and *gaikoku* 'foreign country' in Line 3. This set of data shows that *by foreign investors* could not be identified as a translation unit, because of the diverse translations of the non-prepositional element *foreign*.

Example 7.8. Non-translation units: *by foreign investors*

```
1    … funds in baht by foreign investors, did not … (kaigai tooshika niyoru)
2       … companies by foreign investors from 1997. (gaikokujin tooshika niyoru)
3    … was mostly led by foreign investors.          (gaikoku tooshika niyoru)
```

In summary, the translation units were not identified at the size of prepositional phrase, due to: (a) translations of a given preposition were variable; (b) some prepositions were not translated; and/or, (c) translations of non-prepositional element were variable. It was often the case that two or three of these were seen in the examples of one prepositional phrase, e.g. *for the market* (Example 7.9). The translations of *for* were diverse: *nitotte* 'for' in Line 1; *no* 'of' in Lines 2-3, and, zero translation in Line 4. Also, the nominal phrase *the market* was rendered differently as well: *shijoo* in Lines 1-2; and, *sooba* in Lines 3-4.

Example 7.9. Non-translation units: *for the market*

```
1      was the worst month  for the market but I think           (shijoo nitotte)
2   the short-term outlook  for the market.                      (shijoo no)
3       to end quickly and  for the market to soon turn higher.  (sooba no)
4        about the outlook  for the market.                      (sooba)
```

This study provided empirical evidence that prepositional phrases cannot be translation units in the ARC. In other words, the analyses showed that prepositional phrases are not large enough to be translation units. This was an unexpected finding. If one thinks of a preposition at the single word level, one can easily assume that it cannot be a translation unit. This is because prepositions at this level are usually polysemous. Monolingual dictionaries tell us that they have many definitions; for example, *for* has 34 definitions according to the Collins Dictionary (2003). Bilingual dictionaries also tell us that *for* has many definitions and translations; there are 19 definitions and more than 30 translations (the Genius English-Japanese dictionary, 2001). Hence, *for* alone is very unlikely to be a translation unit.

However, prepositions at the phrasal level, i.e. prepositional phrases, were expected to be translation units. A prepositional phrase is composed by a preposition and nominal phrase. This nominal phrase was assumed to determine what *for* meant; therefore, the whole prepositional phrase is likely to be a monosemous translation unit. However, this was not validated. This study examined the six prepositional phrases containing *for*: *for the market* (Example 7.9); *for foreign firms*; *for next year*; *for the full year*; *for the year*; and, *for this year*. All of them showed that their translations of *for* were variable; the nominal phrases such as *the market* and *foreign firms* did not narrow down what *for* means, which prevented the phrases from being translation units. This suggests that their translation units should be larger units. Phrases are not large enough to focus on for investigating translation units of prepositions.

7.4.2. Adjectival phrases

There are two other phrasal types that this study could not identify as translation units: adjectival and verbal phrases. The former did not occur much in the sample sets of the target words. As shown in Table 7.15, there were only two in the analysis of *foreign*. The other target words did not contain any adjectival phrases which appeared three or more times. The latter, on the other hand, occurred much more, i.e. 50 verbal phrases in total. Most of them were extracted in the analysis of *expected*. Only a few were found in the analysis of *said* and *told* as well.

Table 7.15. Examined adjectival and verbal phrases

Target words	Adjectival phrases	Verbal phrases
market	0	0
year	0	0
new	0	0
foreign	2	0
said	0	1
told	0	2
expected	0	47
Total	2	50

Domestic and foreign was one adjectival phrase that this study examined. This phrase could not be identified as a translation unit, because its translations were variable. Some examples are listed in Example 7.10. The most dominant translation was used in only two lines: *kokunaigai no* in Lines 1-2. The other translations occurred only once.

Example 7.10. Non-translation units: *domestic and foreign*

```
1      … from domestic and foreign rivals.              (kokunaigai no)
2 … year with domestic and foreign fund managers keen …(kokunaigai no)
3     Rising domestic and foreign palm oil demand …    (naigai no)
4 … to govern domestic and foreign brokerages trading …(kokunai oyobi gaikoku no)
5          " Domestic and foreign investors are …      (kokunai … to kaigai ..)
```

As discussed in 7.3.1, Example 7.10 is another set of evidence that English adjectives were not rendered into Japanese adjectives. None of the translations included adjectives. The most dominant translation was composed by a compound noun *kokunaigai* 'the inside and outside of country' and *no* 'of' (Lines 1-2). The translation in Line 3 has the same structure with another compound noun: *naigai* 'the inside and outside'. The translation in Line 4 was the longest one. It has two compound nouns combined with *oyobi* 'and' and ended with *no* 'of'; the nouns were *kokunai* 'country inside' and *gaikoku* 'foreign country'. Line 5 differs. The phrase *domestic and foreign* was not

234

rendered together in this example. The translator did not render the phrase *domestic and foreign* as a chunk; rather, s/he took the larger nominal phrase *domestic and foreign investors* and rendered it into *kokunai tooshi ka to kaigai tooshi ka* 'domestic investor and foreign investor'. Such a diversity of rendering seen in Lines 1-5 clearly indicates that this phrase does not establish stable translation equivalence with certain Japanese expressions. Therefore, *domestic and foreign* could not be a translation unit.

It was for a different reason that the other phrase *more foreign* failed to be a translation unit (Example 7.11): *more foreign* is not an appropriate semantic unit in the data set, which caused the failure. The word *more* did not modify *foreign*; rather, it modified the noun after *foreign*. For example, in Line 1, *more foreign* co-occurred with *investment*; what *more* means in this context is 'a greater amount of investment', not 'a greater amount of foreignness'. Therefore, this nominal phrase should be segmented into the two parts: *more* and *foreign investment*, not *more foreign* and *investment*. The latter segmentation does not capture the meaning of what the original sentence is supposed to mean. Therefore, *more foreign* is not a meaningful unit. The same can be said in Lines 2-6. None of the examples used *more foreign* as 'a greater amount of foreignness'. Therefore, this adjectival phrase *more foreign* was not a unit where meaning lies either. As long as this set of *more foreign* is examined, *more foreign* is an inappropriate semantic unit; hence, there was no way that it could be a translation unit.

Example 7.11. Non-translation units: *more foreign* (i)

```
1 … car and  auto parts markets to more foreign products, U.S. trade officials …
2    … direct financing and allow more foreign investment in the local bond …
3   … economic growth and attract more foreign investment.
4   … the move would help attract more foreign fund inflows.
5        … were aimed at drawing more foreign tourists to China this year as …
6 … and paperboard markets up to more foreign competition.
```

One might wonder whether the one-equivalent principle also demonstrated that this adjectival phrase *more foreign* was not a translation unit. Example 7.12 shows how *more foreign* was rendered; the expressions in bold correspond to *more*. In Lines 1-5, *more* was not rendered at all. *Foreign* was rendered; however, the translations were variable. Line 6 is a rare example where *more* was rendered: *yori issoo* 'more and more'. Due to the variety of translations, the one-equivalent principle could not identify this phrase *more foreign* as a translation unit either.

Example 7.12. Non-translation units: *more foreign* (ii)

```
1    … parts markets to more foreign products, U.S. trade …        (gaikoku)
2 … financing and allow more foreign investment in the …          (gaikoku)
3  … growth and attract more foreign investment.                  (kaigai kara no)
4 … would help attract more foreign fund inflows.                 (gaikoku kara no)
5    … aimed at drawing more foreign tourists to China this …     (gaikoku jin)
6      … markets up to more foreign competition.                  (yori issoo kaigai)
```

This finding does not mean that the phrase of '*more* + adjective' always fails to be a translation unit. Example 7.13 shows some evidence from the ARC that it is possible for this type of phrase to be a translation unit; the expressions in bold correspond to *more*. This is because '*more* + adjective' in Example 7.13 is an appropriate semantic unit. *More equitable* in Line 1 is used for 'a greater amount of equitableness' in the context of *more equitable distribution*. Therefore, the segmentation should be *more equitable* and *distribution*. Similarly, *more satisfactory* in Line 2 means 'a greater amount of

236

satisfaction; and, *more flexibility* in Lines 3-6 means 'a greater amount of flexibility'.

Therefore, '*more* + adjective' in Example 7.13 is an appropriately segmented unit;

hence, they are promising candidates of translation units. If one examines *more flexible*

in Lines 3-6 with the one-equivalent principle, the phrase was always rendered into the

same translation *yori jyuunan na* 'more flexible'. As long as this set is concerned,

therefore, the adjectival phrase *more flexible* is a translation unit.

Example 7.13. *More* + adjectives

```
1       … with that a more equitable distribution of income …   (yori koohee na)
2   … a resumption of more satisfactory growth.                 (yori juubun na)
3 … China was taking more flexible measures to make …          (yori juunan na)
4 … it said adding a more flexible exchange policy would …      (yori juunan na)
5    … banks to have more flexible operations is an …           (yori juunan na)
6 Germany must adopt more flexible labour laws in order …       (yori juunan na)
```

In summary, there were two adjectival phrases examined in this study. Neither of them

could be identified as translation units. One phrase *domestic and foreign* demonstrated

that the translations were variable. Other phrase *more foreign* presented that it was not a

semantic unit. Both required larger units to be monosemous translation units.

7.4.3. Verbal phrases

This study examined 51 verbal phrases. Most of them had a syntactic sequence of 'BE +

-ed' (e.g. *was expected*, *is expected to grow*). Only five phrases had the other, 'HAVE +

-ed' (e.g. *have said*, *had told*). The one-equivalent principle could not identify any of

them as translation units. The translations were too variable. For example, the verbal

phrase *are expected* occurred in 14 lines in the data of *expected* (excluding the phrases

of '*are expected to* + verb', e.g. *are expected to reach*, which will be discussed later).

Among the 14 translations, the most frequent one was the noun *mitooshi* 'prospect'

(Lines 1-2, Example 7.14); however, the occurrence was only five (36 percent). The rest

of the translations were even more infrequent, e.g. *mikoma reru* 'are expected' in Line 3

(occurring once), *yosoo sa reru* 'are anticipated' in Line 4 (occurring three times), *yosoo*

sa re te iru 'are being anticipated' in Line 5 (occurring three times), and *mi rare te iru*

'are being looked' in Line 6 (occurring once).

Example 7.14. Non-translation units: *are expected*

```
1          Record harvests  are expected in three of China's …(mitooshi)
2 … and regional employees  are expected.                    (mitooshi)
3          Mostly fair skies are expected for the coming week. (mikoma reru)
4    … normal temperatures  are expected in Brazil today, …  (yosoo sa reru)
5            Regular rains   are expected in the next few days.(yosoo sa re te iru)
6 … major policy revisions   are expected , dealers said …     (mi rare te iru)
```

It is interesting to see whether the passive expression remained in translation or not.

According to Makino and Tsutsui (1986), *rareru* and *reru* are common passive markers

(the passive markers are shown in bold in Example 7.14). The passive markers were not

seen in Lines 1-2; while, they were seen in Lines 3-6, although there was a grammatical

difference between Lines 3-4 and Lines 5-6. The pair of Line 4 and Line 5 presents this

difference clearly. *Expected* in both lines was rendered into the same main verb: *yosoo*

suru 'anticipate'. However, the difference is the endings. Line 4 has the passive marker

reru; while, the Line 5 has the passive marker *re* followed by the progressive form *te iru*.

Such a translation including both passive and progressive forms was not rare; this

translation seen in Line 5 occurred three times. Also, Line 6 is another case as well: *mi*

rare te iru 'are being looked'. Therefore, Example 7.14 shows that: (a) the translations of the verb *expect* were variable, e.g. the noun *mitooshi* 'prospect', the verbs such *mikomu* 'expect' in Line 3, *yosoo suru* 'anticipate in Lines 4-5, and *miru* 'look' in Line 6; (b) the translations of the passive expressions were variable, e.g. zero translations in Lines 1-2, the passive forms in Lines 3-4, and the passive and progressive forms in Lines 5-6.

The comparison of Example 7.14 and Example 7.15 reveals one part of the nature of Japanese verbs. The translations of *are expected* overlap with the ones of *is expected*. The same translation was shared in Lines 1-2 and 7, as well as in Lines 3 and 8, in Lines 4 and 9, and, in Lines 6 and 10. This happened because Japanese verbs do not change according to number and gender of the subject; 'English verbs change their endings according to the person doing the action ('I go', 'she goes'), but this is not the case with Japanese verbs' (Bunt, 2003). No matter if the subject is singular such as *it* in Line 9 or plural such as *temperatures* in Line 4, the present passive form of *expected* could correspond to *yosoo sa re ru* 'BE-present anticipated' (henceforth, reflecting this nature, 'BE-present' will be used in the paraphrases of Japanese expressions, instead of 'are' or 'is').

Example 7.15. Non-translation units: *is expected*

```
7          … (APEC) forum  is expected next week to …         (mitooshi)
8                    Rain  is expected in some dry regions …(mikoma reru)
9 … national economy, "it  is expected that performance of …(yosoo sa reru)
10 … governor Eddie George  is expected around 0945 GMT, …    (mi rare te iru)
```

It was surprising, though, that this translation *yosoo sa re ru* 'BE-present anticipated' occurred as the translations of *was expected* and *were expected*, as well. Lines 1 and 7 (Example 7.16; the past tense marker *ta* is marked in bold) are the linguistic evidence for it. This indicates that the four verbal phrases shared the same translation: *are expected*; *is expected*; *was expected*; and, *were expected*. Does Japanese language have no past tense of verbs? The concordance lines of *was expected* was examined. This phrase occurred in 14 lines. Some of them used the past tense: *yosoo sa re te i ta* 'BE-past anticipated' in Line 3, *yosoo shi te i ta* 'BE-past anticipating' in Line 4. However, such rendering was seen only in 5 lines; the rest of the 9 cases, *was expected* was translated into the non-past expressions, e.g. *yosoo sa re te iru* 'BE-present being anticipating' in Line 2, *mikoma reru* 'BE-present expected' in Line 4 and *iu* 'say' in Line 5. Moreover, in the concordance lines of *were expected*, all the translations were non-past expressions, e.g. e.g. *mikoma reru* 'BE-present expected' in Line 7 and *omoo* 'think' in Line 8. Therefore, the past tense is not very compatible between English and Japanese, as far as *was/were expected* is concerned.

Example 7.16. Non-translation units: *was expected* **and** *were expected*

```
1   … company's securities  was expected and the whole …        (yosoo sa reru)
2   … Monday and more rain   was expected on Tuesday.            (yosoo sa re te iru)
3              "The fall     was expected as it is still …       (yosoo sa re te i ta)
4         The operation      was expected by the market.         (yosoo shi te i ta)
5  and further improvement   was expected this week because …    (mikoma reru)
6  … an official statement   was expected later.                 (iu)
7        and energy prices   were expected, he said.             (yosoo sa reru)
8   … crop and that rains    were expected in May during sowing.(mikoma reru)
9    … aluminium premiums    were expected.                      (omoo)
```

Such incompatibility of past tense has been discussed in some studies (Kindaichi, 1988; Makino and Tsutsui, 2008). Both claim that if it is a circumstantial event, the past tense

is not often used. One of Makino and Tsutsui's examples was the sentence *there was no wind*. They maintain that the past tense in this sentence corresponds to non-past tense, if it is 'relatively unimportant circumstantial information that has no direct bearing upon the major story line'. However, with the data from the ARC, it is rather hard to evaluate if the sentence is unimportant information or not, since one cannot hold the original text to see the whole story line. Therefore, their claim could not be examined with this data set.

One might wonder if any larger verbal phrases such as *were expected to be* are rendered into Japanese without too much variation or not. The phrase *are expected to stay* gave supportive evidence for it (Example 7.17; the bold expressions correspond to *are expected*). It occurred in three lines. The most frequent translation was *suii suru mitooshi* 'prospect of staying' in Lines 1-2; and, the other translation was *tsuzuku to mi rare te iru* 'BE-present being looked to continue' in Line 3. The former achieved 67 percent of the total occurrences. Similarly, another verbal phrase *are expected to resume* showed that the most dominant translation covered 50 percent of the total occurrences: *saikai sa reru mitooshi* 'prospect of resuming' in Lines 4-5. Therefore, larger verbal units helped to have less variation of translation in some cases.

Example 7.17. Non-translation units: *be expected to* +**verb (i)**

```
1  … oil prices  are expected to stay  in a …        (suii suru mitooshi)
2  … oil prices  are expected to stay  firm this …   (suii suru mitooshi)
3   "Blue chips  are expected to stay  firm for …    (tsuzuku to mi rare te iru)
4 .. since April,  are expected to resume  in July …  (saikai sa reru mitooshi)
5 .. five months,  are expected to resume  in late …  (saikai sa reru mitooshi)
6 … PGM exports  are expected to resume  in late …   (saikai sa reru to mi rare te iru)
7 … exports now  are expected to resume   by early …(saikai sa reru to yoso sa re te iru)
```

241

However such cases were rare. Most of the '*BE expected to* + verb' had variable translations. Their most dominant translations did not usually cover more than 40 percent of the total occurrences. For example, the phrase *is expected to rise* occurred three times. All the translations were different (Example 7.18; the bold expressions correspond to *is expected*): *takamaru to yosoo sa re te iru* 'BE-present being anticipated to rise' in Line 1; *zooka suru to mikoma re* 'BE-present expected to increase' in Line 2; and, *jooshoo suru mitooshi* 'prospect of boosting' in Line 3. Larger verbal phrases did not help to reduce the variety of translations in most cases.

Example 7.18. Non-translation units: *be expected to* +verb (ii)

```
1 … for currency is expected to rise going into …  (takamaru to yosoo sa re te i ru)
2 … grain demand is expected to rise to 500 …       (zooka suru to mikoma re)
3   … percentage is expected to rise in the …       (jyooshoo suru mitooshi)
```

The translations of 'HAVE + *-ed*' were less variable than the ones of 'BE + *-ed*'. For example, the present perfect form *has told* occurred in five lines (Example 7.19). All of them were rendered into Japanese verbs with *ta*; *ta* 'shows completion, and that actions occurred in the past' (Bunt, 2003). The verbs used were *tsutae ru* 'tell' in Lines 1-2, *kata ru* 'talk' in Line 3, *tsuukoku su ru* 'report' in Line 4, and *yoosee suru* 'order' in Line 5. This shows that 'HAVE + *-ed*' and the *ta* form were compatible, although the verb *tell* failed to have a stable relationship with one Japanese verb.

Example 7.19. Non-translation units: *has told*

```
1           … William Perry  has told President Bill Clinton …(tsutae ta)
2                     China  has told visiting Singapore elder …(tsutae ta)
3 … Vice President Al Gore  has told China there would be …   (kata tta)
4 … Clinton administration  has told Canada it would take     (tsuukoku shi te ita)
5                     China  has told U.S. firm Newmont Gold …(yoosee shi ta)
```

However, the relationship between 'HAVE + *-ed*' and the *ta* form was not confirmed in the case of *have said*. The *ta* form was not seen at all in the three occurrences. The most dominant translation was the present form of the verb *iu* 'say' in Lines 1-2; the minor one was the progressive form of the verb *shime su* 'show' in Line 3.

Example 7.20. Non-translation units: *have said*

```
1       Government sources  have said if the convention was …(iu)
2       Private economists  have said the CPI, South Koreas's …(iu)
3     … Chinese economists  have said the full amount could …(shime shi te iru)
```

Therefore, the translations of verbal phrases are very complicated. First, main verbs were rendered variably. This occurred because the main verbs were still polysemous in the small context within the verbal phrases. They need to be examined in larger contexts such as clauses. The case of *dealers expected* supports this conclusion. *Expected* had 62 different translations in the sample set; however, the clause *dealers expected* had only one dominant rendering: *diiraa ra wa … yosoo shi te iru* 'dealers … BE-present being anticipating' (see details in 6.1.2). The context outside of the verbal phrase is important in making the main verb monosemous and to be rendered into one single translation (the relevant discussion will be made in 7.4.5). Second, both 'BE + *-ed*' and 'HAVE + *-ed*' were often incompatible with Japanese expressions, except the case of *has told*. It is

interesting to see if the situation would be different in larger contexts. However, the ARC was not large enough for this investigation.

7.4.4. Nominal phrases

As discussed in 7.2.2, all the identified translation units at the phrasal level were nominal phrases. However, this does not mean that any nominal phrases can be translation units in the ARC. In fact, this study examined 119 nominal phrases; and, only 19 of them qualified as translation units (Table 7.16). This subsection will demonstrate how the one-equivalent principle rejected the rest of 100 phrases as translation units.

Table 7.16. Examined nominal phrases

Target word	Nominal phrases	
new	45	(7)
foreign	31	(1)
market	33	(9)
year	10	(2)
said	0	
told	0	
expected	0	
Total	119	(19)

All the 100 phrases failed to be translation units for one single reason: their translations were too variable. *New era* in Example 7.21 is one of these examples. *Era* was rendered *jidai* all the time; however, the translations of *new* were different (shown in bold): the prefix *shin* 'new'; or, the adjectives *atarashii* 'new', *arata na* 'fresh'. The occurrences

were almost equally distributed, which did not allow this set to have one dominant translation of *new era*.

Example 7.21. Non-translation units: *new era*

```
1        … and Apple opens a  new era of cooperation between…   (shin jidai)
2 … market would usher in a  new era under China with strong …(shin jidai)
3      … over Hong Kong as a  new era for the region and said …(shin jidai)
4  Hong Kong stocks began a  new era under Chinese rule on a …(atarashii jidai)
5   … sound as it entered a  new era with the float of the …   (atarashii jidai)
6         We are entering a  new era of market principles.      (atarashii jidaii)
7    … who want to mark the  new era with a buoyant market.     (arata na jidai)
8 "This agreement begins a  new era in trade relations …        (arata na jidai)
9   … is expected to mark a  new era under China on Thursday …(arata na jidai)
```

Some translations were variable when loan words (i.e. words 'copied into one language from another language' (Trask, 1998)) were used. For example, *sentiment* in *market sentiment* was translated into the loan word *senchimento* in Lines 1-3 (Example 7.22). On the other hand, in Lines 4-6, it was rendered into the non-loan word *jiai*. Such loan word and non-loan word translations caused the variety of translations of the phrase *market sentiment*. Therefore, this phrase could not be regarded as a translation unit. Similar cases were found in many other nominal phrases, e.g. *new system* (*system* was rendered into the loan word *shisutemu*), *new positions* (*positions* into *pojishon*), *new projects* (*projects* into *purojekuto*), *new technology* (*technology* into *tekunorojii*), *new approach* (*approach* into *apuroochi*), *foreign funds* (*funds* into *fando*), and *foreign journalists* (*journalists* into *jyaanarisuto*).

245

Example 7.22. Non-translation units: *market sentiment*

```
1   … influence is on  market sentiment," Gengold        (shijoo no senchimento)
2                   "  Market sentiment did not improve … (shijoo no senchimento)
3               "But  market sentiment is still cautious …(shijoo no senchimento)
4 … gradually improve  market sentiment," said Lee …      (shijoo no jiai)
5                     Market sentiment is still good …    (shijoo no jiai)
6 … slightly dampened  market sentiment, but brokers said …(shijoo no jiai)
```

Abbreviations made some nominal phrases non-translation units as well. For example, the translation of *foreign exchange market* is normally 外国為替市場 *gaikoku kawase shijoo* (shown in Lines 1-3, Example 7.23). However, there are at least two ways to make this word short: 外為市場 *gaitame shijoo* (Lines 4-6); and, 為替市場 *kawase shijoo* (Line 7). The former took 国 and 替 out of 外国為替市場; while, the latter took 外国 out of 外国為替市場. These abbreviations occurred relatively equally in the set of f*oreign exchange market*. Therefore, the one-equivalent principle could not identify this phrase as a translation unit, since there was no dominant translation covering 85 percent of the total occurrences.

Example 7.23. Non-translation units: *foreign exchange market*

```
1    … rumours in the  foreign exchange market on Friday …(gaikoku kawase shijoo)
2     … spread in the  foreign exchange market that the … (gaikoku kawase shijoo)
3. intervening in the  foreign exchange market.          (gaikoku kawase shijoo)
4.developments in the  foreign exchange market.          (gaitame shijoo)
5 … investment to the  foreign exchange market which was …(gaitame shijoo)
6.reporters after the  foreign exchange market closed for …(gaitame shijoo)
7.intervention in the  foreign exchange market in May …  (kawase shijoo)
```

Paraphrasing was another circumstance which caused various translations of a given phrase. For example, the phrase *Britain's new Labour government* can give one good example. The commonest translation of *Britain's* was *eekoku no* 'Britain's' shown in

Line 1 (Example 7.24). However, it was also rendered into *burea* 'Blair'; the phrase *Britain's new Labour government* was paraphrased into *burea shin roodootoo seeken* 'Blair's new Labour government' (Line 2). Another example was the phrase *new team*. The typical translation of *team* was the loan word *chiimu* in Line 3. However, it was rendered into *naikaku* 'Cabinet' as well; so, paraphrasing from *new team* into 'new cabinet' occurred (Line 2). The original sentence was: *"And certainly the IMF is very happy with the agreements we've reached with the new team," he added*. This sentence was not clear whether the *team* referred to the Cabinet or not; however, it can be assumed that the translator knew that the *team* referred to the Cabinet from the previous sentences, which caused this paraphrased translation.

Example 7.24. Non-translation units: *Britain's new labour government* and *new team*

```
1          Britain's new Labour government will …(eekoku no roodootoo shin seeken)
2      by  Britain's new Labour government to … (burea shin roodootoo seeken)

3              "We have a  new team, (a new) manager in Europe.(arata na chiimu)
4 … we've reached with the  new team he added.           (shin naikaku)
```

Lastly, the diverse translations occurred when N + N phrases were rendered into N + *no* + N 'N + of + N'. For example, *foreign investors* was often rendered into *gaikoku tooshika ra* (Lines 1-3, Example 7.25). Both *foreign* and *investors* corresponded to the nouns: *gaikoku* 'foreign country'; and, *tooshika ra* 'investors'. The structure of this translation was N + N. However, some cases demonstrated that it was also translated with *no* 'of', e.g. *gaikoku no tooshika ra* 'investors of foreign country' (Lines 4-6). The structure of this translation was N + *no* + N. Both were almost equally frequent; therefore, neither of them could be a dominant translation. The one-equivalent principle rejected *foreign investors* as a translation unit.

247

Example 7.25. Non-translation units: *foreign investors*

```
1 … gold field to  foreign investors with a recent …    (gaikoku tooshika ra)
2  … Wednesday as  foreign investors snapped up …       (gaikoku tooshika ra)
3             "  Foreign investors continued selling …(gaikoku tooshika ra)
4 … attractive to  foreign investors," it said.        (gaikoku no tooshika ra)
5                  Foreign investors are to be allowed …(gaikoku no tooshika ra)
6    … shunned by  foreign investors because of the …  (gaikoku no tooshika ra)
```

Therefore, the 100 nominal phrases failed to be translation units, since their translations were variable. The main reasons for such variety were identified. These were: translating into loan words; abbreviations; N + *no* + N phrases; and paraphrasing. When they occurred, having one dominant translation which covered 85 percent or more was not likely to happen; hence, they caused many nominal phrases to be non-translation units.

7.4.5. Clauses

As discussed in 7.2.3, the three clausal types were examined in this study: subject-verb (e.g. *spokesman said*); verb-object (e.g. *supported the market*); and, subject-verb-object clauses (e.g. *the source told Reuters*). Only the subject-verb clauses could be identified as translation units. However, Table 7.17 shows that the one-equivalent principle did not identify all the examined subject-verb clauses. Fifty were investigated; three succeeded to be translation units, and the rest (forty-seven) failed.

248

Table 7.17. Examined clauses (subject-verb)

Target words	Subject-verb	
said	22	(1)
expected	19	(1)
told	5	
market	4	(1)
year	0	
new	0	
foreign	0	
Total	50	(3)

Pronouns were obstacles; in particular, third person pronouns were problematic. Example 7.26 shows one typically occurring case; the translations of the pronoun *she* is highlighted in bold. The translations of this third person pronoun were variable. *She* corresponded to the occupational positions such as *riji* 'director' in Line 1 and *iinchoo* 'chairperson' in Line 2, and the relational term such as *yuujin* 'friend' in Line 3. There was no example in which *she* was actually rendered into Japanese pronouns. This meant that the clause was unlikely to have the one dominant translation; therefore, *she said* failed to be a translation unit.

Example 7.26. Non-translation units: *she said*

1		She said she had been surprised by …	(***riji wa*** … *nobe ta*)
2		She said that release of the private …	(***iinchoo wa*** … *nobe ta*)
3		She said telephone lines to the …	(***kono yuujin ni*** *yoru to*)
4	But	she said Finland had already taken …	(*nobe ta*)

One might wonder why *she* was not translated into the corresponding Japanese pronoun. This was because '[t]here is no real third person pronoun in Japanese' (Makino and Tsutsui, 1986). There are words denoting third person, e.g. *kare* 'he', *kanojo* 'she', and *karera* 'they'; however, they joined the Japanese vocabulary fairly recently (Kindaichi, 1988). They are seen 'primarily in novels translated into Japanese and in Japanese

novels' and 'in current spoken Japanese' (Makino and Tsutsui, 1986). Unless one wants to add a certain foreignness to a text, third person pronouns are not used. Seeing that the ARC consists of newswire texts, Japanese pronouns are unlikely to be seen, as shown in Example 7.26. English pronouns tend to be paraphrased and rendered into nouns to refer to the person.

There are some cases in which pronouns were not rendered at all. *She* did not correspond to any Japanese expression in Line 4 (Example 7.26); zero translation of the pronoun occurred (only *said* was rendered into *nobe ta* 'stated'). The original clause of Line 4 was: *But she said Finland had already taken a risk when it did not join this spring*. The translation did not have a word denoting the person who said about Finland: *Shikashi, konshun ERM ni kamei shi naka tta koto de, sude ni risuku wa seotte i ru, to nobe ta* 'But, by the fact that (Finland) did not join ERM this spring, (it) has already taken a risk', (she) said. None of the items shown in parentheses appeared in the translation. This was because of the nature of subject ellipsis in Japanese; 'elements which can be understood from the context and/or from the situation can be omitted in Japanese' (Makino and Tsutsui, 1986). Therefore, it was not surprising that all the clauses including third person pronouns failed to be recognised as translation units.

Another obstacle was verbs. The translations of verbs did not stop being variable at the clausal level either (the discussion about the phrasal level was made in 7.4.3). For example, Example 7.27 shows some concordance lines of the subject-verb clause *market appeared*; the items in bold correspond to the verb *appeared*. Among the total three occurrences of this item, the translations of *appeared* were diverse. It was

250

rendered into *shiteki shi ta* 'pointed out' in Line 1, *yoo da* 'seemed' in Line 2, and the noun *moyoo* 'look' in Line 4. Therefore, the clause *market sources said* failed to be a translation unit.

Example 7.27. Non-translation units: *market appeared*

```
1. also said the  market appeared to be looking for …   (shijoo wa ... shiteki shi ta)
2          "The  market appeared not to have much …     (shijoo wa... yoo da)
3      … said the  market appeared to be biding its …     (shijoo wa ... moyoo)
```

This did not happen only to subject-verb clauses; the same was seen in the cases of verb-object clauses as well. For example, Example 7.28 is the concordance lines of *told reporters*. The translations of *told* were again diverse: *nobe ta* 'stated' Line 1; *kata tta* 'talked' in Lines 2-3; and, *shiteki shi ta* 'pointed out' in Line 4. The translations of *reporters* were less variable: *kisha dan ni* 'to the press' in Lines 1-2; and, *kisha dan nitaishi* 'towards the press' in Lines 3-4. Since the clause *told reporters* could not have one dominant translation, the one-equivalent principle could not assign it as a translation unit.

Example 7.28. Non-translation units: *told reporters*

```
1      … to Japan," Wang  told reporters in Taipei.         (nobe ta)
2 .. closed on Monday, she  told reporters.                   (kata tta)
3                Shen  told reporters that the European …   (kata tta)
4           Stiglitz  told reporters                          (shiteki shi ta)
```

The verb translations were not stable in the cases of subject-verb-object clauses either. If one looks at the concordance lines of *official told Reuters*, one can find more than 5 translations of *told* (Example 7.29). The most dominant one was *nobe ta* 'stated' in

Lines 1-2. Among the total 13 occurrences of the clause, *nobe ta* appeared only in 7 lines. The second dominant one was *kata tta* 'talked' in Lines 3-4; it occurred in 4 lines. The minors occurred once: e.g. *ni yoru to* 'according to' in Line 5 and *happyoo shi ta* 'announced' in Line 6. Such diversity ended up making the clause *official told Reuters* to be a non-translation unit.

Example 7.29. Non-translation units: *official told Reuters*

```
1 … on the new offer," the  official told Reuters.              (nobe ta)
2 … 109.00 cents a kg, the  official told Reuters.              (nobe ta)
3                     The  official told Reuters that Japan's …(kata tta)
4 … province of Batman, an  official told Reuters.              (kata tta)
5       … tax level, the  official told Reuters.              (ni yoru to)
6   … GMT on Saturday, an  official told Reuters.              (happyoo shi ta)
```

On the other hand, the translations of nouns were often less variable at the clause level. Needless to say, when a given clause had nominal phrases which had loan words, abbreviations, and paraphrasing, their translations were diverse (as discussed in 7.4.4). However, other than that, nouns in the extracted clauses corresponded to fewer items. For example, Example 7.27 shows that all the subject nouns of *market* were rendered into one translation: *shijoo wa* 'market + subject marker' in Lines 1-3. Similarly, the noun *a trader* in the clause *a trader said* did not have many different translations: *aru toreedaa wa* 'a certain trader + subject marker' in Lines 1-4; and, zero translation in Line 5 (Example 7.30). This dominant translation actually includes a loan word *toreedaa*; however, this did not cause the diversity of translation, unlike some cases seen in 7.4.4. This was because there was no example in which a non-loan word, i.e. the Japanese word for *trader*, was used.

Example 7.30. Non-translation units: *a trader said*

```
1   … Monday's USDA report," a trader said in reference to … (aru toreedaa wa)
2   … just like yesterday," a trader said.                   (aru toreedaa wa)
3      … May at the close," a trader said.                    (aru toreedaa wa)
4     … if it is available, a trader said.                    (aru toreedaa wa)
5   … everybody is bearish," a trader said.                   (-)
```

7.4.6. Summary

This study identified 299 non-translation units (i.e. 206 phrases and 93 clauses). The one-equivalent principle could not identify them as monosemous units; their translations were variable. Such diverse translations occurred due to many factors. First, when nouns were rendered into abbreviations and loan/non-loan words, their translations tended to be several. Also, when nouns were paraphrased in translation, it did not help the phrase or clause to be translation units. Second, verbs were another factor; the translations of verbs and verbal phrases were found to be diverse at the phrasal and clausal level. Main verbs in English corresponded to different main verbs in Japanese. Moreover, 'BE + *-ed*' and 'HAVE + *-ed*' were often incompatible between the two languages; 'BE + *-ed*' did not necessarily correspond to the Japanese passive forms *rareru* and *reru* and 'HAVE + *-ed*' did not always correspond to the Japanese form *ta*, which indicates completion of an action. The third person pronoun was an obstacle as well; these did not appear in Japanese news texts, which caused different rendering, either by paraphrasing the pronouns or through zero translations. The last main factor was prepositions; these were not compatible between English and Japanese. All the phrases including prepositions could not be translation units.

253

7.5. Methodology

The methodology, discussed in section 3.4, has identified 24 translation units: 10 for *market*; two for *year*; one for *government*; one for *economic*; seven for *new*; one for *foreign*; one for *said*; zero for *told*; and, one for *expected*. However, if one looks at Table 7.18, the number of extracted translation units was very low compared with the number of non-translation units. The analyses examined 330 items; only 24 of them (7 percent) were regarded as translation units. This poor extraction of translation units were caused by two characteristics of the methodology. First, this study took a 'no synonymy' position (see details in 3.4). A translation unit is a unit whose translation is always one single corresponding item in a given corpus, no matter how many times it occurred. If a given item is found to have several translations, then the item is not a translation unit. As it was already pointed out in 3.4.8, this definition is narrow (cf. Teubert, 2004b). Therefore, it was inevitable that this 'no synonymy' stance only allowed the analyses to identify a tiny fraction of the translation units, compared with those that might have been identified using the broader definition of Teubert (2004b).

Table 7.18. Numbers of translation and non-translation units

	Translation unit	Non-translation unit
Single word	2	7
Phrase	19	206
Clause	3	93
Total	24	306

Second, I set a filter of 85 percent in order to eliminate mistranslations and creative translations (discussed in 3.4.2). However, this filter was, perhaps, too strict when

infrequent collocations were examined. For example, if an item occurred six times and five of them had a Japanese translation equivalent, one might think that this translation is dominant enough. However, the coverage would be five out of six, which was only 83 percent; this item would be filtered out. Therefore, this filter setting is hard to optimise for infrequent items occurring 6 times or less. Among the examined 330 items, about half of them, i.e. 176 items (53 percent), occurred seven times or more. The other half of the collocations was infrequent ones, which were not benefiting from this filter setting. Regarding these two settings, there are some refinements to make. I will discuss the matter of synonyms in 7.5.1 and the matter of filtering in 7.5.2.

7.5.1. Synonyms

The 'no synonymy' position is radical. This needs to be softened; the methodology has to allow that a translation unit can be rendered into some synonymous items. However, the difficulty of this to practice is to establish a criterion of synonymy. It is not easy to assess if the translations are synonymous in an objective way, as discussed in 3.3.3. For example, how can one clarify that the three Japanese items denoting *new*, the prefix *shin* 'new'; and the adjectives *atarashii* 'new' and *arata na* 'fresh' (shown in bold in Example 7.31), are synonyms or not? One might say that they are synonyms if they have similar meanings. However, they were all rendered from the same word *new*; all the three items inevitably share a similar meaning. Then, any translations can be easily regarded as synonyms, which would make any examined items into translation units. A clear distinction between synonyms and non-synonyms is required.

Example 7.31. Non-translation units: *new era* **(i)**

```
1       … and Apple opens a  new era of cooperation between…    (shin jidai)
2  … market would usher in a  new era under China with strong …(shin jidai)
3     … over Hong Kong as a  new era for the region and said …(shin jidai)
4  Hong Kong stocks began a  new era under Chinese rule on a …(atarashii jidai)
5  … sound as it entered a  new era with the float of the …  (atarashii jidai)
6        We are entering a  new era of market principles.     (atarashii jidaii)
7    … who want to mark the  new era with a buoyant market.    (arata na jidai)
8  "This agreement begins a  new era in trade relations …      (arata na jidai)
9  … is expected to mark a  new era under China on Thursday …(arata na jidai)
```

The investigation of non-translation units in 7.4 suggested that there are some good candidates to be treated as synonyms, which can be easily identified: abbreviations and loan/non-loan words. As discussed in 7.4.4, they prevented some phrases and clauses from being translation units. For example, 外国為替市場 *gaikoku kawase shijoo* 'foreign exchange market' had two abbreviations: 外為市場 *gaitame shijoo*; and, 為替市場 *kawase shijoo* (Example 7.32). Seeing that they all referred to the 'foreign exchange market', they should not be treated as different translations and so make the phrase a non-translation unit. Therefore, it is appropriate that such abbreviations are regarded as synonyms. The identification of abbreviations should be straightforward; they are usually like each other. Moreover, one can simply use dictionaries for their reference.

Example 7.32. Synonymous translations: *foreign exchange market*

```
1    … rumours in the  foreign exchange market on Friday …(gaikoku kawase shijoo)
2     … spread in the  foreign exchange market that the … (gaikoku kawase shijoo)
3. intervening in the  foreign exchange market.          (gaikoku kawase shijoo)
4.developments in the  foreign exchange market.          (gaitame shijoo)
5 … investment to the  foreign exchange market which was …(gaitame shijoo)
6.reporters after the  foreign exchange market closed for …(gaitame shijoo)
7.intervention in the  foreign exchange market in May …  (kawase shijoo)
```

256

Second, loan and non-loan words can be treated as synonyms as well. This judgement is

not difficult either. All the loan words in Japanese are written in one special writing

system called *katakana*; '*[k]atakana* is used to write the numerous loanwords from

Western languages' (**Kodansha Encyclopedia of Japan**, 1983). Therefore, it is fairly

straightforward to spot loan words in Japanese texts. Considering that 'borrowed words

… make up a significant portion of Japanese vocabulary' (Makino and Tsutsui, 1995),

this treatment is likely to help more translation units to be identified. Applying this to

Example 7.33, *market sentiment* can be regarded as a translation unit. Lines 1-3 had

loan word of *sentiment*, e.g. *senchimento*; while, Lines 4-6 had non-loan word

translation of *sentiment*, e.g. *jiai*. Under this revised one-equivalent principle, they are

synonymous.

Example 7.33. Synonymous translations: *market sentiment*

```
1   … influence is on  market sentiment," Gengold        (shijoo no senchimento)
2                   "  Market sentiment did not improve … (shijoo no senchimento)
3               "But  market sentiment is still cautious …(shijoo no senchimento)
4 … gradually improve  market sentiment," said Lee …      (shijoo no jiai)
5                      Market sentiment is still good …    (shijoo no jiai)
6 … slightly dampened  market sentiment, but brokers said …(shijoo no jiai)
```

7.5.2. Filter

The filter of 85 percent was used to remove mistranslations and creative translations.

This cut-off level at 85 percent was arbitrary (as discussed in 3.4.2). I assumed that if

one examines an item with the concordance lines, most of them (i.e. 85 percent or more

of the lines) are 'standard' translations and only 15 percent or less are 'unusual'

translations. However, this assumption was not always validated, especially when

infrequent items were examined. For example, the nominal phrase *single market* was one of the infrequent items; it was seen four times in the sample set (Example 7.34). The dominant translation was *tan'itsu shijoo* 'single market' (Lines 1-3). The other one was *ooshuu shijoo* 'European market' (Line 4), which was a creative translation by paraphrasing *single* into *European*. Since this translation occurred once out of four lines, the coverage was taken as 25 percent, which was more than I assumed as the coverage of an 'unusual' translation.

Example 7.34. Case of creative translation: *single market*

```
1                "The  single market is the  basis for      (tan'itsu shijoo)
2    … policy and the  single market, all 15 member         (tan'itsu shijoo)
3  … the launch of the  single market, the EU's statistics   (tan'itsu shijoo)
4    … industry in the  single market to ensure that …       (ooshuu shijoo)
```

Although the one-equivalent principle did not identify it as a translation unit, *single market* is a unit which has to be called a translation unit. As long as the 'standard' translations (Lines 1-3) are examined, *single market* always corresponded to *tan'itsu shijoo*. Line 4 should not prevent this phrase from being assigned as a translation unit. Then, what is the refinement to make in order to eliminate Line 4? It is easy to say that the cut-off level should be lowered from 85 into 75 percent. This of course allows *single market* to be a translation unit. However, there are more cases in which 75 percent is not good enough. For example, Example 7.35 shows that the creative translation occurred one out of three (33 percent). The dominant translation was *atarashii kisee* 'new regulation'; the creative one was *arata ni doonyuu sa reru kisee* 'regulation which will be introduced newly'. In order for this phrase *new regulations* to be a translation unit, the cut-off level should be lowered again to 67 percent. However, such a cut-off level

does not have any linguistic justification behind it, which is less than ideal as a criterion for the identification of translation units.

Example 7.35. Case of creative translation: *new regulations*

```
1   … not say when the  new regulations might be completed.  (atarashii kisee)
2      China has issued  new regulations aimed at stopping …  (atarashii kisee)
3   … of accounting for  new regulations and to publish …      (arata ni doonyuu
                                                                 sa reru kisee)
```

Alternatively, there is another cut-off setting for eliminating unwelcome translations. This is based on their nature of infrequencies. 'Unusual' translations including mistranslations and creative translations do not recur; 'appropriate translations tend to be repeated while wrong translations will remain singular' (Teubert, 2004b); similarly, the creativities of one-off translations were empirically illustrated in Kenny's study (2001). Applied this to the matter of the cut-off point, the criterion of 'unusual' translations are their infrequencies. If a given translation occurs only once, it is likely to be a mistranslation or creative one. One should ignore such translations and examine if the item is a translation unit or not within the 'standard' translation examples. Under this criterion, for example, Line 4 in Example 7.34 is ignored since it was a one-off translation; the examination of Lines 1-3 yields the phrase *single market* as a translation unit. Similarly, Line 3 in Example 7.35 is ignored; the phrase *new regulations* can be identified a translation unit as well. This criterion is linguistically justified, which is more ideal than the one discussed above.

I returned to the 299 non-translation units (i.e. 206 phrases and 93 clauses) and carried out this alternative one-equivalent principle using the new synonymy and filter settings.

The result is shown in Table 7.19. A substantial amount of items were newly identified as translation units, i.e. 82 items (shown in parentheses in Table 7.19). A main benefit of this alternative method is the case of *told*. The one-equivalent principle, previously, failed to identify any translation units of *told*; however, six translation units were recognised by the alternative method: *has told*; *told a meeting*; *told an audience*; *told him*; *told the magazine*; and, *she told reporters*. The one-off filter helped them to be assigned as translation units. Moreover, it is noteworthy that, as shown in Table 7.20, this alternative method could identify 13 prepositional phrases (e.g. *for next year*, *on year*, *in new orders*, *of foreign exchange*, and *with foreign companies*) and 7 verbal phrases (e.g. *have said*, *are expected to stay*, *were expected*, and *had been expected*). One might wonder if the most dominant translation unit size is still phrase with the newly identified translation units. However, the findings remained the same; translation unit sizes are predominantly at the phrase level. The least dominant size is still that of the single word.

Table 7.19. Results of the alternative principle and target words

Target words	Translation units	
market	10	(+20)
year	2	(+9)
new	7	(+15)
foreign	1	(+16)
said	1	(+5)
told	0	(+6)
expected	1	(+11)
Total	22	(+82)

Table 7.20. Results of the alternative principle and types

Types	Translation units	
Nominal phrases	19	(+46)
Clauses	3	(+16)
Adverbial phrases	0	(+0)
Prepositional phrases	0	(+13)
Verbal phrases	0	(+7)
Total	22	(+82)

This investigation, therefore, gives empirical evidence that the two settings are worth changing in the one-equivalent principle. The objectiveness of criteria for identification of translation units still remains. Moreover, considering that the filter of 85 percent did not have any linguistic justification, this alternative method is even more appropriate in this sense; the one-off treatment was validated in some linguists' studies (e.g. Kenny, 2001; Teubert, 2004b). There are some advantages and disadvantages of this alternative method; however, this investigation will be continued in future studies.

7.5.3. Corpus size

In 3.2.2, I pointed out that one limitation of this study was the size of the ARC: 1.8 million words for English texts and 2.5 million morphemes for Japanese texts. One benchmark regarding sizes of corpora for translation research was Danielsson's research (2001) where only 7 translation pairs were identified in the 1-million-word parallel corpus (0.5 million words per language); she maintains that much larger corpora are preferable for identifying translation units. The ARC was only about 4 times larger than her parallel corpus, if the English texts are compared. Was this size large enough? If one simply compares the numbers of identified translation units, this study could identify

much more than Danielsson did (2001); this study extracted 24 of them, and more than 82 translation units with the alternative method. Therefore, the 4-times-larger corpus helps the extraction of more translation units.

One might point out that this comparison is not appropriate in a strict sense, as the identification methods of the two studies differed (discussed in 3.3.4). This is true; however, both studies aimed to extract as many as possible translation units using a computational approach. And, it is worth mentioning that the result of this study agrees with Danielsson's claim (2001); the larger corpus yields many translation units, even though this study examined far fewer items, i.e. 330 items (9 single words, 225 phrases, and 96 clauses), than Danielsson's research (2001), i.e. 12,099 items. The ratios of translation units out of the examined items show that the larger parallel corpus is a more efficient platform for identification.

More importantly, whether the ARC was large enough for studies identifying translation units or not depends on the applications. One possible application can be translation memory (i.e. 'a database … of a pair of source and target text segments' (Reinke, 2006)). Translation units are the items whose translation is only one, e.g. *market economy* and *shijou keezai*. If one encounters the translation unit *market economy* in the future, their Japanese translation should be *shijou keezai*, nothing else, at least in the ARC, or most likely in the other news texts as well. Such translation pairs are useful for translators; they are good entries for translation memory. However, for this application, the size of the ARC was not large enough. Many more translation pairs would be required in the database to make it useful in reality.

The other possible application is cross-linguistic projects on translation units. This is my on-going interest (as mentioned in 3.2); how does translation unit size in one language pair differ from the one in another language pair? As discussed in 2.4.2, Koller's cross-linguistic claim on translation unit size was only partially validated; 'a translation between unrelated languages will usually involve larger units than if SL and TL [=target language] are closely related' (cited and translated in Shuttleworth and Cowie, 1997). As far as the studies of the Indo-European language pairs, this view was found to be tenable. However, further investigation is required for the language pairs outside the Indo-European language family. My systematic review found only one study of non Indo-European language comparison: English-Hebrew (Toury, 1986). His findings disagree with Koller's cross-linguistic claim, as well as this study does. I would like to identify translation units in the other non-Indo-European language pairs such as English-Chinese and English-Korean and examine how their translation unit sizes are different from ones within the Indo-European language pairs.

For the purpose of such a cross-linguistics project, the size of the ARC would be large enough. First, this project would require more studies of non-Indo-European language pairs. In order to make the studies comparable, all the subsequent studies should use the same method. Four-million-word parallel corpora should not be hard to find. It might be categorised as a larger parallel corpus at the present time, however, the size of corpora have been increased rapidly due to the technological development. Therefore, parallel corpora of a similar size are likely to be in the public domain, or if not, they will likely arrive soon. Second, the size of the ARC could yield a substantial amount of translation units as well. The original method identified relatively fewer units i.e. 24 items;

however, as demonstrated in 7.5.2, the alternative method could identify a good amount of translation units, i.e. 82 items more, at least. With this quantity, it is possible to compare the most and least dominant sizes of translation units and draw out the significant differences and similarities between the language pairs.

7.6. Summary

This chapter has demonstrated that the study of translation units can reveal many aspects of translation between two languages. Some are worth summarising here. Regarding the nature of translation units in the English-Japanese comparison, this study found that:

(a)   Nominal phrases, subject-verb clauses, and single words are the only grammatical forms to be translation units;
(b)   translation unit sizes vary along with the word classes of the target words; and,
(c)   nouns are a core element of translation units.

The translation units identified by the stricter version of the one-equivalent principle were only three types (in 7.2). The occurrence of each type was found to be associated with whether a target word was a noun, adjective, or verb (in 7.2.1). Verbs were particularly unique in the sense that their translation units were always subject-verb clauses. This means that verbs tend to be monosemous at the clausal level in the English-Japanese comparison. Another finding was that all the identified translation units (except *economic*) had one noun at least (see 7.2.1). Nouns, therefore, play important roles in translation units. In other words, the study of translation units might

be the study of nouns and their monosemy (i.e. the study searches for the level at which a given noun can be monosemous).

These identified characteristics of translation units lead us to some useful interpretations for developing the methodology of translation unit identification:

(d)   Nouns are a better choice than adjectives or verbs to represent content words;
(e)   translation unit sizes do not differ between frequent and infrequent words; and,
(f)   the refinements of the method achieved a higher extraction rate.

Since the study of translation units is about nouns and their translations, it is suggested that linguists can focus on nominal target words if their goal is to identify translation unit size (in 7.2.1). This study guaranteed that the outcome (i.e. the most and least dominant sizes) would remain the same, and, that the efficiency (i.e. the extraction rate of translation units) would increase. Another noteworthy finding regarding the methodology was that frequent words are good target words for the study of translation unit size. It is true that frequent words usually behave uniquely; their contextual behaviours are idiosyncratic (Sinclair, 1999). However, seeing that such idiosyncrasy of frequent words is observable only when linguists examine them at the single word level, as long as linguists look at them at larger levels as well, the problem of frequent words does not disturb the result (i.e. translation unit sizes). Therefore, although this study focuses only on frequent words, the outcome of translation unit sizes is likely to be similar to those for infrequent words.

The refinements of synonyms and unusual translations are other interesting findings regarding the methodology (in 7.5.2). The pilot study proved, with linguistic justification, that these refinements can increase the extraction rate of translation units. These successful refinements were observed in the examination of non-translation units. This demonstrates that the analysis of non-translation units can be as useful and informative as the analysis of translation units for advancing the field.

Moreover, it is important to emphasize that what the refined pilot study suggested was not only a higher extraction rate of translation units. It also indicated the nature of English-Japanese translation in the ARC. They are:

(g)   Frequent occurrences of one-off translations; and,
(h)   frequent occurrences of translations with abbreviations, loan words, and non loan words.

It was interesting to see that nearly three times more translation units were identified as soon as the refinements were applied (in 7.5.2). This means that one-off translations, (non) loan words, and abbreviations occurred often in the English-Japanese translation. Seeing that the ARC is a news corpus, the high occurrence of abbreviations in translated text is easily predictable. News articles about stock market are likely to be read by people who know about the field; it is understandable that abbreviations are used often. The result that loan and non-loan words were used frequently was not surprising. For example, *sentiment* in *market sentiment* corresponds to the loan word *senchimento* in some translations and to the non-loan word *jiai* in other translations. Seeing that news articles have to be timely, if a new loan word is spread between readers, it starts to be

used in the text. It is possible that one can spot both loan and non-loan translations in news texts, even though a given corpus is a collection of articles published for a short period like the ARC (i.e. one year).

However, more than half of the additional translation units were identified by the one-off filter, which was unexpected. This suggests that 'unusual' translations often happen. They included creative translations and drastic paraphrasing; mistranslations were hardly seen. One might intuitively think that news is not as creative as fiction; however, if one focuses on translations, creativity occurs in news translation as well. After all, what translation has to fulfil is to 'fill cultural gaps' (Utiyama and Isahara, 2003) and convey the original meaning, no matter which genre a text belongs to.

Lastly, but not the least, this empirical study has successfully observed many facets of the relationship between the two languages. The main findings were:

(i) Translation equivalence is robust due to some resistance to variation in translation units;
(j) shifts occur in most translation pairs; and,
(k) the identification of established differences between English and Japanese.

The observed translation equivalence found in this study was robust (in 7.3.2). In the total 24 translation pairs, a few had one-to-one translation equivalence. Their relationships were vulnerable in a sense that they would be broken by small changes such as modifier insertion. On the other hand, most translation pairs had many-to-one translation equivalence. Their relationships were rather unaffected by variation. The

267

examination of such unaffected equivalence revealed the differences between English and Japanese, e.g. article usages. Also, many translation pairs underwent one, or more than one, shift (in 7.3.1). The examination of such shift occurrences demonstrated the compatibilities and incompatibilities between English and Japanese, e.g. adjective usage. This shows that translation equivalence is a useful resource for identifying cross-linguistic relationships. The fact that the examinations identified established differences (e.g. article and adjective usages) increases our trust in the reliability of this resource.

8.   Conclusion and future work

This thesis has demonstrated how corpus study can be of assistance regarding two central issues of translation units: their identification and their size. In this chapter, I focus on two areas: I summarise the main achievements of this study (in 8.1); and, I make suggestions for relevant future research which may be of value (in 8.2).

8.1.   Achievements of the current study

8.1.1.   The variables associated with size

One achievement of this study is the identification of four variables associated with translation unit size: the theoretical viewpoint from which the 'translation units' are being considered; the language pair under investigation; the type of text; and, who is doing the translating. It is fair to say that these variables are not what I discovered in this thesis. Different linguists have written about different variables before (see details in 2.4); all the four variables were there in the literature already, although claims about their significance were usually made with little linguistic evidence. What deserves to be called an achievement of this study is, therefore, not the discovery of the variables, but the systematic examination of the variables and associated findings. The relevant literature (see in 1.1 for the selection of literature) was scrutinised for finding the key factors associated with translation unit size, and the four factors were identified. All the case studies were sorted by each factor and then investigated thoroughly to see if claims about the effects of the variables could be validated through comparison with the other

269

linguists' case studies. This comprehensive review concluded that translation units had an average tendency to respond to changes in the variables with changes in size.

It should be highlighted that this empirical study has discovered a fifth variable: the target word. As discussed in 7.2.1, translation unit sizes vary according to the word class of the target word: noun, adjective; and, verb. Translation units of nouns were identified at any of the three sizes (single word, phrase, and clause). On the other hand, the translation units of the adjectives and verbs were selective. Translation units of the former were only identified at the levels of single word and phrase, while translation units of the latter were identified only at the level of the clause. The investigation and confirmation of the importance of the five variables is very useful; it allows the construction of a clear framework, within which linguists can systematically investigate translation unit size. Moreover, the (dis)agreements amongst linguists are no longer simply based on claims and hunches, but now have linguistic justification behind them: size is strongly influenced by a combination of the five variables in a given study.

8.1.2.   The one-equivalent principle

The establishment of the one-equivalent principle is another achievement of the current study. It allows researchers: (a) to identify translation units objectively; and, (b) to identify translation units at all levels. Such a methodology was not located during the systematic review of translation unit size (see 3.3). First, regarding point (a) above, objectivity is an irreplaceable benefit: it makes the judgement of translation units unbiased. The extraction can be carried out consistently throughout considerable

amounts of data. Moreover, the objectivity makes translation units reproducible: if somebody else will do the same research with the same data, the result (i.e. the identified translation units) would be same (see 3.4.7). Because of this reproducibility, the one-equivalent principle is very useful for contrastive studies. An analysis of translation unit size in a fiction corpus (with all the other variables remaining the same as in this thesis) would reveal clear differences and similarities of translation units between news and fiction by comparing the results of the fiction corpus and the ARC. The fact that different linguists using the methodology should not create differences in the results makes it possible for one to carry out such an effective comparison, and makes the resulting conclusions much more open to the scrutiny of peers.

Second, regarding point (b) above, the one-equivalent principle allows researchers to examine items of any size, i.e. from morpheme to text. This aspect of the one-equivalent principle derives from the monosemous principle (Teubert, 2004b), which also has this characteristic. Such a method is useful for two types of studies: one in which linguists are interested in a particular size of translation unit; and, another in which linguists are interested in any possible size of translation units (i.e. they do not know which sizes the outcomes will be). The study of Wang (2006) belongs to the former type. Wang used the monosemous principle to investigate nominal phrases only; translation units were successfully identified (see details in 2.3.2 and 3.3.3). This thesis belongs to the latter type of study; I did not have any prior idea of what sizes the translation units would be identified at. Seeing that most methods constrain the identifiable sizes of translation units (see 3.3), this aspect of the one-equivalent principle is very valuable. In addition, two important factors of the principle, i.e. the objectivity and the absence of any

constraint on identifiable size, remain even after two refinements (synonymy and filter settings) were applied.

One might point out that this study could not identify translation units at the levels of morpheme, sentence, paragraph, and text (as discussed in 7.2.6); therefore, the identifiable sizes were constrained by the one-equivalent principle. However, this is not the case. It was not the one-equivalent principle that prevented this study from examining units at the levels of the sentence or larger; rather, it was the ARC (see details in 3.2.2 and 3.5). The discussion of pros and cons (section 3.2) suggested, however, that the ARC was still the best parallel corpus for this study. Similarly, this study could not examine translation units at the morpheme level either; however, this was again, not due to the one-equivalent principle, rather it was due to the settings of target words. The target words were single words in this study. I did not set up any morphemes as examined items. This was because English and Japanese do not share morphological similarities in general (in 3.4.8). Therefore, I discarded the possibility of translation units at this level; Barkhudarov (1993) also gave a linguistic justification for this discarding (see 3.4.8).

8.1.3. Translation units and their equivalents

The third achievement of this study is the translation pairs (i.e. the translation units and their equivalents) identified by the one-equivalent principle. What is useful about them is that the translation pairs are ideal for identifying similarities and differences across languages. This interpretation is motivated by Tognini-Bonelli (2002), who 'addresses

the issue of comparing words and expressions across languages'. The main messages from her study are that one should be careful for choosing what to compare, and that one cannot simply pick up two words, *X* in language A and *Y* in language B, for comparison, just because they look alike. How does one know that they are comparable? *X* might be an incomplete semantic unit; a single word is not necessarily a unit of meaning (Teubert, 2004a). If *X* does not have any concrete meaning (or, *X* has a very ambiguous meaning), it will not be possible to compare *X* with *Y*. The result of such a comparison would be that *X* and *Y* look alike, but that they are not semantically the closest match; which is often rather obvious before analysis anyway.

For example, the English word *person* and the Norwegian word *person* were examined in such a way, and it was identified that their behaviours were quite different (Johansson, 2007). Such a result, however, is easily predictable; linguists would not usually assume that two words behave exactly the same across languages, just because they look similar. It seems pointless to carry out such an analysis if one's goal is simply to identify the closest matches across languages. One picks up two bilingual items which may or may not be the closest match, compares them, and recognises that they are not the closest pair. In which case, one should keep carrying on this process repeatedly until s/he gets a pair which has no differences. Clearly, this is not efficient. Similarly, Wierzbicka (1999) compared two words, *sadness* from English and *pečal* from Russian; this comparison showed that their meanings were different. However, this finding is not surprising either; one would not expect randomly chosen words to have identical meanings across languages, even though Wierzbicka 'thinks' that the two items both roughly denote the

emotion of sadness. The problem of these studies is the biased choice of target words; therefore, the items were not comparable in the first place.

On the other hand, translation pairs are ideal to compare for contrastive analysis (the term 'contrastive analysis' is used here in a broad sense, referring to the 'study of two languages in contrast' (Baker, 1998)). Translation pairs are guaranteed to be comparable. First, a translation unit and its equivalent is the closest semantic match between the two languages (especially in the corpus). The pairs are not randomly matched by subjective judgement; rather, they are identified by an unbiased principle with concrete linguistic evidence. Second, translation units and equivalents are monosemous units; therefore, they have concrete meanings. If one compares such two bilingual expressions, then, the findings are interesting. For example, *market sources* and *shijoo suji* is a pair identified in this study. Wherever *market sources* occurred, it was always rendered into *shijoo suji*. They are, semantically, the closest match due to this linguistic evidence. The comparison of the items showed that: (a) their grammatical structures are the same, i.e. N + N; and, (b) their number sensitivity is same, i.e. both *market sources* and *shijoo suji* are plural forms (*a market source* corresponded into a different Japanese expression, *aru shijoo suji*). Therefore, the two items are identical not only in meaning, but also in grammatical features. The similarities between the two languages are identified in this comparison: N + N and number sensitivity in nouns.

Another example from this study can be the pair *new investments* and *shinki tooshi*. The one-equivalent principle identified them as identical semantic units. The comparison of the items showed that: (a) their grammatical structures are actually different, i.e. ADJ +

274

N for *new investments* and N + N for *shinki tooshi*; and, (b) the number sensitivities are also different, i.e. *new investments* is number sensitive and *shinki tooshi* is not (it corresponds to both *new investment* and *new investments*). Therefore, even though the two items are semantically the closest match, their grammatical features are different. The same meaning is expressed differently in English and Japanese. Adjectives in English do not correspond to adjectives in Japanese. Numbers on nouns do not correspond between the languages. Such cross-linguistic differences are identified effectively in this 'real' comparison.

Tognini-Bonelli (2002) actually establishes an identification method of such 'real' comparable items (she calls them 'functionally complete units of meaning'; the relevant discussion can be seen in 3.3.4). However, this method involves purely manual labour, including many subjective criteria, which are inevitable due to her thorough semantic examination. This is not ideal for contrastive linguists, who might think of identifying 'real' comparable items with this method; for them, having items to compare is just the starting point of their studies. If it takes much energy and time to only get to the start line, it is not going to be productive for them to take this route. Moreover, the subjectivity is not welcome. If the items are chosen by a biased criterion, then, this choice is as bad as the method they might usually use. Therefore, the method that Tognini-Bonelli (2002) suggests is not suitable for contrastive analysts. Instead, the one-equivalent principle overcomes such problems and provides 'appropriate' comparable items for contrastive studies more efficiently.

It is worth mentioning that translation pairs for such contrastive usages were actually validated in this study. Lots of differences between English and Japanese were identified when the translation pairs were compared: article usage, number sensitivity, future tense, subject status, and pronouns. The fact that this study could recognise these established cross-linguistic differences indicates that the translation pair is a guaranteed platform to yield good contrastive outcomes. Seeing that the differences were identified only from 24 translation units, this method can be seen to be very efficient. To take this line of argument further, studies of translation pairs can be useful when one is interested in contrastive analysis between a language pair which has not been studied before, such as English-Ainu (Ainu is a language of northern Japan, having 'no known relationship with other languages and cannot be assigned to a family' (Malmkjær, 2004)). One can carry out an extraction of translation pairs in a corpus; by examining the pairs, it is possible to quickly observe a good amount of differences and similarities across languages. Translation pairs are therefore a reliable and powerful resource for contrastive analysis.

8.2. Future work

In order to make such translation pairs more usable for contrastive analyses, there is a line of investigation worth embarking on. The current methodology takes 1,000 lines of target words and identifies their translation pairs from these. This might be too much preliminary work for contrastive analysts to carry out, since finding comparable items is not their goal, but the starting point of their research, as discussed above. Therefore, it is important to make this labour as minimal as possible. As discussed in 3.4.3, the pilot

studies showed that 100 and 500 sample lines yielded only a few translation units (which led to this study set being increased to 1,000 sample lines instead). However, these pilot studies were carried out without the refinements of synonyms and the filter for 'unusual' translations. Seeing that these refinements identified nearly three times more translation units, it would now be appropriate to carry out another set of pilot studies on fewer sample lines, with these new criteria, and see if the outcome would be satisfactory or not. If 100 sample lines are enough to extract a satisfactory amount of translation units, then, this is more promising for contrastive linguists to use for their research.

There are also future researches that one can carry out regarding English-Japanese translation units. The extensive investigations mentioned in the thesis were:

(a)   The extraction of translation units of the 17 noun target words;
(b)   the investigation of Koller's view on translation unit sizes;
(c)   the testing of the scarcity of sentence translation units;
(d)   the examination of prepositional phrases at the clausal level; and,
(e)   the examination of verbal phrases at the sentence level.

First, this study found that nouns can be a representative for all of the content words in the study of translation unit size (discussed in 7.2.1). One possible research, therefore, is the extraction of translation units with 17 target nouns. The focuses would be: how many more translation units will be actually identified; and, whether the identified translation units can find more similarities and differences between English and Japanese or not. Second, this study did not support Koller's cross-linguistic view on translation unit size (in 7.2.5): 'a translation between unrelated languages will usually

involve larger units than if SL and TL [=target language] are closely related' (cited and translated in Shuttleworth and Cowie, 1997). By identifying translation units in other non-Indo-European language pairs, further examination of this claim is possible.

Third, it was assumed from the findings of this study that the translation units at the sentence level would rarely occur in the ARC, even though the methodology would have allowed me to examine this level (in 7.2.6). However, this is an assumption from the literature; it would be ideal if one can validate it with linguistic evidence. For this purpose, another English-Japanese corpus is required, as the ARC does not contain appropriate data for consulting on this issue (see details in 3.4.3). Until the emergence of a such corpus, this research has to be pending.

Fourth, this study showed that the one-equivalent principle could not identify any prepositional phrases as translation units (in 7.4.1 and 7.4.5). The new settings made it possible to identify 14 translation units out of 54 examined items; this means that most prepositional phrases still could not be translation units at the phrasal level either. Further investigation would therefore include: are most prepositions likely to be translation units at the clause level (i.e. are they monosemous at the clausal level)? Similarly, as discussed in 7.4.3 and 7.4.5, this study has seen that verbs are identified as translation units at the clausal level (with the one-equivalent principle) and the phrasal level (with the refinements). However, there are still some verbs which could not be regarded as monosemous units at these levels. The question which needs to be clarified, thus, is: can they be monosemous at the sentence level? These investigations will give further insights to translation units in the English-Japanese comparison.

In addition to these works, there is another potentially illuminating project worth considering: the nestedness of translation units. There are a few linguists who discuss this nature of translation units, although they use different expressions. For example, Bennett (1994) maintains that 'each UT [= translation unit] is part of a larger unit'; that is, translation units can exist at different levels simultaneously. In a different manner, Newmark (1988b) also posits that translation units are nested: 'The sentence is the ''natural'' unit of translation…Within the sentence, there are five possible sub-units of translations [emphasis in original]'. Both linguists suggest that different sizes of translation units co-exist simultaneously in text. This thesis aimed to identify 'minimal' translation units. Identifying larger translation units was out of this thesis's focus. However, there were two positive signs of nestedness recognised (7.2.2). Both *domestic*

gold market and *gold market* were identified as translation units; the latter is a part of the former. The same can be said in *stock market* and *Seoul stock market*. The possible further investigation involves: (a) extracting larger units of *government* and *economic* (which were not examined in this study, since they were regarded as single word translation units), and, (b) seeing if any of their larger units such as *coalition government* and *economic reforms* can be identified as translation units or not. This should provide an insight into the frequency of occurrence of nestedness, which would furnish empirical evidence for the claims of Bennett (1994) and Newmark (1988b).

In conclusion, this study has demonstrated that translation unit is not elusive with the one-equivalent principle. If we suppose that a student asks, 'how can we find such translation units, then?', or 'what do these units look like?'; the lecturer can now say that 'an unbiased, reproducible identification method is available' and that, 'there are five variables associated with translation unit size'. The lesson sounds even more interesting if the lecturer mentions about the possibility of translation pairs being ideal comparable items for contrastive analysis. This thesis has successfully presented one possible way to 'describe, categorise, and formalise' (cf. Lefevere, 1993) what a translation unit is with the help of a parallel corpus.

APPENDICES

Appendix 1: Romanisation rules quoted *verbatim* from Iwasaki (2002)[*]

1. The system of Romanization used in this book [Iwasaki, 2002] is a modified Hepburn system with five vowels, two semi-vowels, and 20 consonants including those represented by two letter combination, *ts*, *ch*, and *sh*.

Vowels:	*a*	*i*	*u*	*e*	*o*				
Semi-vowels:		*y*	*w*						
Consonants (voiceless):	*p*	*t*	*k*	*f*	*s*	*sh*	*h*	*ts*	*ch*
Consonants (voiced):	*b*	*d*	*g*		*z*	*j*			
Consonants (sonorant):	*m*	*n*	*r*						

2. Palatalised consonants (before vowels *a*, *u*, *o*) are represented by an added y, as in *kya*, *kyu*, and *kyo* (except for *ja*, *ju*, *jo*).

3. A consonant (*p*, *k*, *ch*, *sh*, *b*, *g*, *j*, *n*, *m* and *r*) may be followed by any of the five vowels. Others are followed by a restricted number of vowels as shown below.

t	is followed by	*a*			*e*	*o*
ts				*u*		
s		*a*		*u*	*e*	*o*
h		*a*	*i*		*e*	*o*
f				*u*		
d		*a*			*e*	*o*
z		*a*	*i*	*u*	*e*	*o*

4. The 100 basic moras are arranged below as found in the traditional table of *kana* symbols.

    Vowels:
    a   *i*   *u*   *e*   *o*

---

[*] From Iwasaki, S. (2002) *Japanese*. With kind permission by John Benjamins Publishing Company, Amsterdam/Philadelphia. www.benjamins.com

Consonant + Vowels:

ka	ki	ku	ke	ko	ga	gi	gu	ge	go
sa	shi	su	se	so	za	ji	zu	ze	zo
ta	chi	tsu	te	to	da	-	-	de	do
na	ni	nu	ne	no					
ha	hi	fu	he	ho	ba	bi	bu	be	bo
					pa	pi	pu	pe	po
ma	mi	mu	me	mo					
ya	-	yu	-	yo					
ra	ri	ru	re	ro					
wa									

Palatalized Consonant + Vowels:

kya	kyu	kyo	gya	gyu	gyo
sha	shu	sho	ja	ju	jo
cha	chu	cho			
nya	nyu	nyo			
hya	hyu	hyo	bya	byu	byo
mya	myu	myo	pya	pyu	pyo
rya	ryu	ryo			

5. Moraic nasal is represented as *n* though their actual pronunciation varies depending on the following consonant […]. When the moraic nasal is followed by a vowel, an apostrophe is added to separate them (e.g. *shin'ai* 'dear') to distinguish a sequence of a nasal consonant /n/ followed by a vowel (e.g. *shinai* 'do not do').

6. Geminate consonants are represented by a sequence of two identical consonants (e.g. *kitte* 'stamp', *happi* 'a happi coat') except when the germination involves *ch*, which is preceded by *t* (e.g. *matchi* 'match').

7. Dipthongs and long vowels are represented by a sequence of two vowels (e.g. *aa*, *ai*, *au*, *ae*, *ao*; *ia*, *ii*, *iu*, *ie*, *io* etc.). However, the sequences, *ei* and *ou*, are spelt as *ee* (e.g. *eega* 'movie') and *oo* (e.g. *oosama* 'king'), respectively, to reflect the actual pronunciation […].

8. In addition to the above basic syllables, innovative combinations may be used to represent recent loan words and onomatopoetic words.

fa as in	*famirii* 'family'	*fi* as in	*firumu* 'film'
fe	feminisuto 'feminist'	*fo*	*fooku* 'folk/fork'
she	*sherri* 'sherry'	*che*	*chekku* 'check'
ti	*tii* 'tea'	*di*	*diizeru* 'diesel'
je	*jetto* 'jet'		

REFERENCES

Altenberg, B. and Granger, S. (2002) "Recent trends in cross-linguistic lexical studies". In Altenberg, B.G., Sylviane (Ed.) **Lexis in Contrast**. Amsterdam, John Benjamins 3-50.

Alves, F. and Gonçalves, J.L.V.R. (2003) "A Relevance Theory approach to inferential processes in translation". **Triangulating Translation: Perspectives in process oriented research**. Amsterdam, John Benjamins 3-24.

Baker, M. (1992) **In Other Words**. London: Routledge.

Baker, M. (ed.) (1998) **Routledge Encyclopedia of Translation Studies**, London: Routledge.

Bakker, M., Koster, C. and Leuven-Zwart, K.M.v. (1998) "Shifts of translation". **Routledge Encyclopaedia of Translation Studies**. London, Routledge 226-231.

Barbosa, H.G. and Neiva, A.M.S. (2003) "Using think-aloud protocols to investigate the translation process of foreign language learner and experienced translators". In Alves, F. (Ed.) **Triangulating Translation. Perspectives in process oriented research**. Amsterdam, John Benjamins 137-155.

Barkhudarov, L. (1993) "The problem of the unit of translation". In Zlateva, P. (Ed.) **Translation as Social Action: Russian and Bulgarian Perspectives**. London, Routledge.

Barlow, M. (1995) **A guide to ParaConc** [online]. http://www.ruf.rice.edu/~barlow/pc.html [Accessed 11 April 2004]

Barlow, M. (2000) "Parallel texts in language teaching". In Botley, S.P.; McEnery, A.M. & Wilson, A. (Eds.) **Multilingual Corpora in Teaching and Research**. Amsterdam, Rodopi 106-115.

Basic Japanese-English dictionary. (1986) Oxford: Oxford University Press.

Bassnett-McGuire, S. (1980) **Translation Studies**. London: Methuen.

Batchelder, E.O. and Ohta, A. (2000) "Rules vs. Rote in Japanese verb inflection". In Melby, A.K. & Lommel, A.R. (Eds.) **LACUS 26, Fullerton, CA: LACUS.**, 39-50.

Bell, R.T. (1991) **Translation and Translating: Theory and Practice**. London: Longman.

Bennett, P. (1994) The Translation Unit in Human and Machine. **Babel**, 40: (1):

    12-20.

Bernardini, S. (2007) Collocations in translated language: Combining parallel, comparable and reference corpora. **Proceeding of CORPUS LINGUISTICS 2007**.

Brown, K., Anderson, A.H., Bauer, L., et al. (eds.) (2006) **Encyclopedia of Language and Linguistics (2nd Edition)**, London: Elsevier.

Bunt, J. (2003) **Oxford Japanese Grammar & Verbs**. Oxford: Oxford University Press.

Catford, J.C. (1965) **A linguistic theory of translation: an essay in applied linguistics**. London: Oxford University Press.

Chang, B., Danielsson, P. and Teubert, W. (2006) "Chinese-English translation database: extracting units of translation from parallel texts". In Barnbrook, G.; Danielsson, P. & Mahlberg, M. (Eds.) **Meaningful texts: the extraction of semantic information from monolingual and multilingual corpora**. London, Continuum 131-142.

Cipris, Z. and Hamano, S. (2002) **Making sense of Japanese Grammar**. Honolulu: University of Hawaii Press.

Collins Cobuild Advanced Learner's English Dictionary (2003) 2nd.Glasgow: HarperCollins Publishers.

Coulmas, F. (1996) **The Blackwell Encyclopaedia of Writing Systems**. Oxford: Blackwell.

Danielsson, P. (2001) **The Automatic Identification of Meaningful Units in Language**. PhD,

Danielsson, P. (2003) Automatic extraction of meaningful units from corpora. **International Journal of Corpus Linguistics**, 8: (1): 109-127.

Durrell, M. (2006) "Germanic Languages". In Brown, K.; Anderson, A.H.; Bauer, L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier.

English-Japanese Dictionary For The General Reader. (1999) 2nd edition.Tokyo: Kenkyusha.

Genius English-Japanese Dictionary. (2001) 3rd.Tokyo: Taishukan.

Goldman-Eisler, F. (1972) Segmentation of input in simultaneous translation. **Journal of Psycholinguistic Research**, 1: 127-140.

Grootaers, W.A. (1983) "Japanese Language". In Itasaka, G. (Ed.) **Kodansha Encyclopedia of Japan**. Tokyo, Kodansha 23.

Hackett, J. (2006) "Semitic Languages". In Brown, K.; Anderson, A.H.; Bauer,

L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 229-235.

Halliday, M.A.K. and Hasan, R. (1976) **Cohesion in English**. New York: Longman.

Hartmann, R.R.K. and James, G. (1998) **Dictionary of Lexicography**. London: Routledge.

Hatim, B. and Munday, J. (eds.) (2004) **Translation: An advanced resource book**, London: Routledge.

Hunston, S. (2002) **Corpus in Applied Linguistics**. Cambridge: CUP.

Isham, W. and Lane, H. (1993) Simultaneous interpretation and the recall of source-language sentences. **LANGUAGE AND COGNITIVE PROCESSES**, 8: (3): 241-264.

Itasaka, G. (ed.) (1983) **Kodansha Encyclopedia of Japan**, Tokyo: Kodansha.

Iwasaki, S. (2002) **Japanese**. Amsterdam: John Benjamins.

Japanese For Busy People, (1994) London: Kodansha International.

Johansson, S. (2007) **Seeing through Multilingual Corpora: On the use of corpora in contrastive studies**. Amsterdam: John Benjamins.

Jorden, E.H. (2000) "Acquired Culture in the Japanese Language Classroom". In LAMBERT, R.D. & SHOHAMY, E. (Eds.) **Language Policy and Pedagogy**. London, John Benjamins 207-220.

Kaiser, S. (2003) "Language and script in Japan and other East Asian countries: Between insularity and technology". In Maurais, J. & Morris, M.A. (Eds.) **Languages in a Globalising World**. Cambridge, Cambridge University Press.

Kamiya, T. (2001) **The Handbook of Japanese Verbs**. London: Kodansha International.

Kenny, D. (2001) **Lexis and Creativity in Translation**. Manchester: St. Jerome.

Kenny, D. (forthcoming) "Translation Units and Corpora". In Kruger, A. & Wallmach, K. (Eds.) **Corpus-based Translation Studies**. Manchester, St. Jerome.

Kindaichi, H. (1988) **Nihongo (Japanese language) vol. 2**. 2nd.Tokyo: Iwanami.

Kodansha Encyclopedia of Japan. (1983) Tokyo: Kodansha.

Kondo, F. (2000) An approach of the Japanese thematic structure. **Proceedings of Japan Association of Systemic Functional Linguistics (JASLF)**.

Kondo, F. (2006) Translatability of words denoting emotions: the case of unhappiness in a Japanese-English Parallel Corpus. **Proceedings of**

LR4Trans-III Workshop 2006 (LREC 2006).

Kondo, F. (2007) Book review of Stig Johansson (2007) 'Seeing through Multilingual Corpora: On the use of corpora in contrastive studies. **International Journal of Corpus Linguistics**.

Kondo, M. and Wakabayashi, J. (1998) "Japanese Tradition". <u>In</u> Baker, M. (Ed.) **Routledge Encyclopaedia of Translation Studies**. London, Riytkedge 485-493.

Lefevere, A. (1993) "Introduction of L Barkhudrav's "The Problem of the unit of translation"". <u>In</u> Zlateva, P. (Ed.) **Translation as Social Action: Russian and Bulgarian Perspectives**. London, Routledge 39-40.

Lighthouse English-Japanese Dictionary. (2007) 5th edition.Tokyo: Kenkyusha.

Livbjerg, I. and Mees, I.M. (2003) "Patterns of dictionary use in non-domain-specific translation". <u>In</u> Alves, F. (Ed.) **Triangulating Translation: Perspectives in process oriented research**. Amsterdam, John Benjamins 123-136.

Lörscher, W. (1991) **Translation performance, translation process, and translation strategies: psycholinguistic investigation**. Tübingen: Gunter Narr Verlag.

Lörscher, W. (1993) "Translation process analysis". <u>In</u> Gambier, Y. & Tommola, J. (Eds.) **Translation and Knowledge**. Turku, University of Turku 195-212.

Lörscher, W. (1996) A Psycholinguistic Analysis of Translation Processes. **Meta**, XLI: (1): 26-32.

Luo, X. (1999) Linguistic Contributions to the Development of Translation Studies in China. **Meta**, 44: (1): 101-109.

Makino, S. and Tsutsui, M. (1986) **A Dictionary of Basic Japanese Grammar**. Tokyo: The Japan Times.

Makino, S. and Tsutsui, M. (1995) **A Dictionary of Intermediate Japanese Grammar**. Tokyo: The Japan Times.

Makino, S. and Tsutsui, M. (2008) **A Dictionary of Advanced Japanese Grammar**. Tokyo: The Japan Times.

Malmkjær, K. (1998) "Unit of Translation". <u>In</u> Baker, M. (Ed.) **Routledge Encyclopaedia of Translation Studies**. London, Routledge 286-288.

Malmkjær, K. (ed.) (2004) **Linguistics Encyclopedia 2nd Ed**, London: Routledge.

Malmkjær, K. (2006) "Translation Units". <u>In</u> Brown, K.; Anderson, A.H.; Bauer, L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 92.

Matsubara, S., Iwashima, K., Kawaguchi, N., et al. (2000) Simultaneous Japanese-English Interpretation Based on Early Prediction of English Verb. **The Forth Symposium on Natural Language Processing**.

Matsumoto, Y., Kitauchi, A., Yamashita, T., et al. (2002) **Morphological Analysis System ChaSen version 2.2.9 Manual** [online]. http://chasen.aist-nara.ac.jp/stable/doc/chasen-2.2.9.pdf Nara Institute of Science and Technology [Accessed 28 January 2009]

Miller, R.A. (1980) **Origins of the Japanese Language**. London: University of Washington Press.

Munday, J. (2001) **Introducing Translation Studies: Theories and Applications**. London: Routledge.

Newmark, P. (1988a) **Approaches to translation**. Hemel Hempstead: Prentice Hall,.

Newmark, P. (1988b) **A Textbook of Translation**. London: Prentice-Hall.

Nida, E.A. and Taber, C.R. (1969) **The theory and practice of translation**. Leiden: E. J. Brill.

Ohara, K.H., Fujii, S., Saito, H., et al. (2003) The Japanese Framenet Project: A Preliminary Report. **Pacific Association for Computational Linguistics**.

Olohan, M. (2004) **Introducing corpora in translation studies**. London: Routledge.

Oxford English Dictionary (2000) [online]. http://dictionary.oed.com [Accessed 24 August 2005]

Quint, J. (2003) "Automatic Japanese transliteration with a formalism for presyntactic analysis". **2003 International Conference on Natural Language Processing and Knowledge Engineering, Proceedings**. 512-518.

Rayson, P. (2003) **Matrix: A statistical method and software tool for linguistic analysis through corpus comparison**. Ph.D.,

Rayson, P. (2007) **Wmatrix: a web-based corpus processing environment**. Computing Department, Lancaster University: http://www.comp.lancs.ac.uk/ucrel/wmatrix/.

Reinke, U. (2006) "Translation Memories". <u>In</u> Brown, K.; Anderson, A.H.; Bauer,

L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 61-65.

Ross, M. (2006) "Language Families and Linguistic Diversity". In Brown, K.; Anderson, A.H.; Bauer, L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 499-507.

Sager, J.C. (1994) **Language engineering and translation : consequences of automation**. Amsterdam: John Benjamins.

Scott, M. (1996) **Wordsmith Tools**. Oxford: OUP.

Shibatani, M. (2006) "Japanese". In Brown, K.; Anderson, A.H.; Bauer, L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 102-105.

Shuttleworth, M. and Cowie, M. (1997) **Dictionary of Translation Studies**. Manchester: St Jerome.

Sinclair, J.M. (1996) The Search for Units of Meaning. **Textus**, IX: 75-106.

Sinclair, J.M. (1999) "A Way with Common Words". In Hasselgård, H. & Oksefjell, S. (Eds.) **Out of Corpora: Studies in Honour of Stig Johansson**. Amsterdam, Rodopi 157-179.

Sinclair, J.M. (2004a) **How to Use Corpora in Language Teaching**. Amsterdam: John Benjamins.

Sinclair, J.M. (2004b) **Trust the Text**. London: Routledge.

Stubbs, M. (1996) **Words and Phrases: Corpus studies of lexical semantics**. Oxford: Blackwell.

Taira, M. (1998) "The japanese particle ne and its literary and linguistic implications: some translation problems". In Bush, P. & Malmkjær, K. (Eds.) **Rimbaud's Rainbow: Literary translation in higher education**. Amsterdam, John Benjamins.

Takahashi, T., Kaneko, S., Kaneda, A., et al. (eds.) (2005) **Nihongo no bunpoo [Japanese Grammar]**, Tokyo: Hitsuji Syobo.

Tancock, L.W. (1958) "Some problems of Style in Translation from French". In Booth, A.D. (Ed.) **Aspects of Translation**. London, Secker and Warburg 29-51.

Terakura, H. (1983) Noun modification and the use of *to yuu*. **The Journal of the Association of Teachers of Japanese**, 18: (1): 23-55.

Teubert, W. (1996) Comparable or Parallel Corpora? **International Journal of Lexicography** 9: (3): 238-276.

Teubert, W. (2001) Corpus Linguistics and Lexicography. **International Journal of Corpus Linguistics**, 6: (Special Issue): 125-153.

Teubert, W. (2002) "The role of parallel corpora in translation and multilingual lexicography". In Altenberg, B. & Granger, S. (Eds.) **Lexis in Contrast**. Amsterdam, John Benjamins 189-214.

Teubert, W. (2004a) "Language and Corpus Linguistics". In Halliday, M.A.K.; Teubert, W.; Yallop, C. & Čermáková, A. (Eds.) London, Continuum 73-112.

Teubert, W. (2004b) "Units of Meaning, Parallel Corpora, and their Implications for Language Teaching". In Connor, U. & Upton, T.A. (Eds.) **Applied Corpus Linguistics: A Multidimensional Perspective** Amsterdam, Rodopi 172-189.

Teubert, W. (2005) My version of corpus linguistics. **International Journal of Corpus Linguistics**, 10: (Number 1): 1-13.

Teubert, W. and Čermáková (2004) "Directions in corpus linguistics". In Halliday, M.A.K.; Teubert, W.; Yallop, C. & Čermáková, A. (Eds.) **Language and Corpus Linguistics**. London, Continuum 113-166.

Tognini-Bonelli, E. (2001) **Corpus Linguistics at Work**. Amsterdam: John Benjamins.

Tognini-Bonelli, E. (2002) "Functionally complete units of meaning across English and Italian: Towards a corpus-driven approach". In Altenberg, B. & Granger, S. (Eds.) **Lexis in Contrast: Corpus-Based Approaches** Amsterdam, John Benjamins 73-96.

Totten III, G.O. (1983) "Romanization of Japanese". In Itasaka, G. (Ed.) **Kodansha Encyclopedia of Japan**. Tokyo, Kodansha 335-336.

Toury, G. (1986) "Monitoring Discourse Transfer: A Test-Case for a Developmental Model of Translation". In House, J. & Blum-Kulka, S. (Eds.) **Interlingual and intercultural communication : discourse and cognition in translation and second language acquisition studies**. Tübingen Gunter Narr 79-94.

Toury, G. (1995) **Descriptive Translation Studies and Beyond**. Amsterdam: John Benjamins.

Tranter, N. (2008) Nonconventional script choice in Japan. **International Journal of Sociology of Language**, 192: 133-151.

Trask, R.L. (1993) **A Dictionary of Grammatical Terms in Linguistics**. London: Routledge.

Trask, R.L. (1998) **Key Concepts in Language and Linguistics**. London: Routledge.

Trask, R.L. (2000) **Introducting Linguistics**. London: Icon Books.

Tsujimura, N. (1996) **An Introduction to Japanese Linguistics**. Oxford: Blackwell.

Uchida, Y. and Nakayama, M. (1993) Japanese verbal noun constructions. **Linguistics**, 31: 623-666.

Umesao, T., Kindaichi, H., Sakakura, A., et al. (1989) **Nihongo Daijiten (Encyclopaedia of Japanese language)**. Tokyo: Kodansha.

Unger, J.M. (2006) "Japan: Writing System". In Brown, K.; Anderson, A.H.; Bauer, L.; Berns, M.; Hirst, G. & Miller, J. (Eds.) **Encyclopaedia of Language and Linguistics (2nd Edition)**. London, Elsevier 95-102.

Utiyama, M. and Isahara, H. (2003) Reliable Measures for Aligning Japanese-English News Articles and Sentences. **ACL-2003**, 72-79.

Vinay, J.P. and Darbelnet, J. (1995) **Comparative stylistics of French and English: a methodology for translation**. Amsterdam: John Benjamins.

Wang, W. (2006) **A Corpus-Driven Study on Translation Units in an English-Chinese Parallel Corpus**. MPhil,

Wierzbicka, A. (1999) **Emotions Across Languages and Cultures: Diversity and Universals**. Paris: Cambridge University Press.

Zhu, C. (1999) Ut Once More: The Sentence as the Key Functional Unit of Translation. **Meta**, 44: (3): 429-447.