Translation Units in Japanese-English Corpora:

The Case of Frequent Nouns

Fumiko Kondo[1]

1.   Introduction

The translation unit is an elusive notion. In particular, the issue of translation unit size remains unsettled. Some theorists maintain that translation units can occur at any language unit length, such as morpheme, single word, phrase, clause, sentence, and so on (Newmark 1988b, Barkhudarov 1993). On the other hand, some theorists argue that the translation unit is restricted to one language unit length, which is the clause (Bell 1991) or the text (De Beaugrande 1978). To enable a quantitative investigation of this issue, I carried out a corpus-driven analysis on the most frequent nouns in an English-Japanese parallel corpus: *market*, *year*, and *government*. Translation unit was re-defined based on Teubert (2004) for this quantitative analysis. I looked up 1,000 examples of each noun and identified translation units with the collocational information ('a frequent co-occurrence of words' (Sinclair 1996: 80)).

2.   Rethinking the translation unit

2.1.   The translation unit and its different senses

When one reads the literature on translation units, one must be aware that there are mainly two senses in which the term 'translation unit' is used. One refers to an inseparable unit in translating; the other to a translator's focus of attention. The former is typically defined as a 'minimal stretch of language that has to be translated together, as one unit' (Newmark 1988b: 54). It is a lexical unit and the main concern is the local context. The latter is usually defined as 'segments of the source text […] to which, at a given moment, the translator's focus of attention is directed' (Alves and Gonçalves 2003: 10). It is a cognitive unit and the main concern is the wider context. Linguists often do not make themselves clear regarding which sense their translation units refer to; even worse, some mix up both senses when arguing what a translation unit is. This is not a desirable situation, especially for discussion of translation unit size. The lexical sense of the translation unit is likely to relate to smaller units, e.g. words and

[1] University of Birmingham
   e-mail: fumiko.kondo@gmail.com

phrases; while the cognitive sense of the translation unit is likely to be larger, e.g. whole texts, paragraphs, sentences, and clauses. The two views are on different levels and should be distinguished clearly.

Malmkjær (2006) agrees with this dichotomy between lexical and cognitive senses of translation units. Based on a traditional division between translation products and translation process, Malmkjær calls the lexical sense a 'product-oriented' translation unit and the cognitive one a 'process-oriented' translation unit (2006: 92). The former can be identified by observing original or translated texts (i.e. translation products); while the latter can be identified by observing the mental processes occurring in a translator's mind (i.e. the translation process). Bennett (1994) also distinguishes lexical translation units from cognitive ones. He coined the terms 'translation atom' for the lexical translation unit ('the smallest segment that must be translated as a whole' (1994: 13)) and 'translation focus' for the cognitive translation unit ('the section of text which the translator focuses on at any one time' (1994: 13)). Bennett uniquely argues that there is another sense of translation unit: 'translation macro-unit'. According to the definition ('the largest linguistic unit which the translator needs to consider (1994: 13)), this is also a cognitive unit which is larger than the translation focus.

This dichotomy between lexical and cognitive senses of translation unit is crucial to this study. This is due to corpora being able to assist in the investigation of the lexical sense of translation units only. Corpus linguistics is a lexical discipline. Corpus data provides lexical evidence of how a word was translated into another language; however, it does not provide cognitive evidence of what is happening in a translator's mind. Given that here I will be using a quantitative corpus linguistic methodology, I will limit myself to investigating only the lexical sense of translation units.

2.2. Definitions of translation units

Many scholars have defined the lexical sense of the translation unit (Toury 1986, Newmark 1988b, 1988a, Barkhudarov 1993, Bennett 1994, Vinay and Darbelnet 1995, Shuttleworth and Cowie 1997, Malmkjær 1998, Teubert 2002, 2004, Malmkjær 2006). The standard definition is Barkhudarov's: 'the *minimal* language unit in the source text that corresponds to an equivalent in the target text' (1993: 40). There are four features of the translation unit on which theorists argue differently. The first is whether a translation unit is a unit of the original or translated text. Most theorists state that a translation unit is a unit of the original text. Only Malmkjær defines it as a unit of the translated text; 'from a product-oriented point of view, the unit of translation is the target-text unit that can be mapped onto a source-text unit' (1998: 286). However, few researchers have yet to subscribe to this position. Moreover, she subsequently changed her approach and more recently defined it as a pair of original

and translated units; '[w]hen the translation product is at issue, translation units are pairs of source and target text segments' (2006: 92).

The second point is whether a translation unit is the 'smallest' unit in the original text or not. Vinay and Darbelnet define a translation unit as 'the smallest segment of the utterance whose signs are linked in such a way that they should not be translated individually'(1995: 21). Similarly, Barkhudarov defines it as the 'minimal' (1993: 40); Newmark as the 'smallest' (1988b: 285); and Toury as the 'manageable' 1986: 83). This is actually an alternative way to distinguish the lexical sense of translation unit from the cognitive one. The former is smaller than the latter; therefore, by defining translation unit as 'smallest', linguists manage to cut off the cognitive sense of translation unit and therefore imply that a translation unit is a lexical unit. Shuttleworth and Cowie, more directly, define that a translation unit is 'the linguistic level at which ST [source text] is recodified in TL [target language]' (1997: 192). Since the term 'smallest' is not very well defined, it is preferable to define the translation unit as a 'lexical' unit.

The third point is whether the translation unit is a syntactic and semantic unit, or either of them. Barkhudarov (1993: 40) raises the issue of whether translation units are 'elements of linguistic form (structure) or content'. He, however, does not clearly state his answer and defines a translation unit as the 'language' unit, which is rather ambiguous. On the other hand, Vinay and Darbelnet clearly maintain that translation units are both syntactic and semantic units; 'lexicological units within lexical elements are grouped together to form a single element of thought' (1995: 21). Teubert also agrees with this view, based on a tenet of Saussure that 'content cannot be separated from form' (Teubert 2001: 128). The fourth point is how to call the corresponding segment of a translation unit in the translated text. The majority of theorists use the term 'equivalent': 'TL equivalent' (Newmark 1988b: 65), 'equivalent' (Barkhudarov 1993: 40), and 'translation equivalent' (Teubert 2004: 185). On the other hand, some theorists use other terms such as 'TL segment' (Toury 1986: 83), 'corresponding segment' (Toury 1986: 88), and 'TL unit' (Vinay and Darbelnet 1995: 31, Shuttleworth and Cowie 1997: 192). Taken into account that equivalence is found in translated texts, the term 'translation equivalent' is more accurate; this helps distinguish it from the other types of equivalents (Altenberg and Granger 2002: 16).

In this study, I take the view that a translation unit is (a) a unit in the original text, (b) a lexical unit, (c) a syntactic as well as semantic unit, and (d) corresponds to the translation equivalent. The definition, however, is not enough for a quantitative study since it does not give much instruction on how translation units should be identified. The two features (a) and (c) indicate the first step of identification but they do not give a clear criterion to determine whether an expression is a translation unit or not. Teubert is one theorist who indicates the criterion in his definition of translation unit; '[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' (2004: 184-5). An item can be regarded as a translation unit (i) if it has only one translation equivalent (i.e. is always translated into one translation equivalent) in the corpus, or (ii) if it has several synonymous translation equivalents in the corpus.

Teubert's definition (2004: 184-5) seems to give clear guidance on how to identify translation units. However, it has a methodological weakness; judging whether the expressions are synonyms or not is not an easy task. For example, when Wang (2006) identified translation units from a English-Chinese corpus, using Teubert's definition, her judgement of synonyms was often rather intuitive. If one looks up my target word in this study, *market*, in a English-Japanese dictionary (Genius English-Japanese Dictionary 2001), the main four Japanese translations are *ichiba* (market/bazaar), *shokuryouhinten* (grocery store), *shijou* (market/marketplace), and *souba* (market/price); translations in the parenthesis shows what they roughly mean (these are the first and second definitions from a Japanese-English dictionary (Genius Japanese-English Dictionary 2003)). The four Japanese translations are likely to appear as the translation equivalents of *market* in parallel corpora. According to the Teubert's definition (Teubert 2004: 184-5), one has to examine whether the four are synonyms or not when *market* is examined. But it is not easy. What makes the judgement tough is that all four were translated from *market*; it is highly likely that they are synonyms to some extent as they share the meaning of *market*. Then, in which cases are the equivalents regarded as non-synonymous?

In order to tackle this issue of synonymy, I have attempted to narrow down what a translation unit is by adjusting Teubert's definition (2004: 184-5). In this study, a translation unit is (a) a unit in the original text, (b) a lexical unit, (c) a syntactic as well as semantic unit, (d) corresponding to the translation equivalent, and (e) a monosemous unit which has only one translation equivalent. A unit which has more than one translation equivalent was not considered as a translation unit. Based on this new definition of translation unit, I identified the translation units of the most frequent nouns (*market*, *year*, and *government*) in the ARC in order to inform an investigation of translation unit size.

2.3.   Hypothesis on translation unit size

Before presenting the investigation, it seems appropriate to discuss what has been argued about translation unit size. The literature suggests several possible translation unit sizes (Table1): phoneme, morpheme, word, phrase (including 'collocation' (1*b*), 'combination of words' (1*d*), 'units above word level' (1*d*), or 'several words' (1*e*, *f*)), clause, sentences, and text. The underlined unit size is that which the theorist believes to be the most typical.

Theorists	Size
a. Toury (1986: 83)	morpheme, word, <u>phrase</u>, <u>clause</u>
b. Newmark (1988b: 65)	morpheme, <u>word</u>, collocation, group, <u>clause</u>
c. Baker (1992)	morpheme, word, units above word level
d. Barkhudarov (1993: 41-5)	phoneme, morpheme, word, <u>combination of words</u>, sentence, and text
e. Vinay and Darbelnet (1995: 22-3)	morpheme, word, <u>several words</u>
f. Teubert (2001: 144-5)	single word, <u>several words</u>

Table 1: Selected views on translation unit size

All the theorists in Table 1 agree that (a) single words can be translation units, and (b) the commonest size is the phrase. How can this argument be applied to my target words, *market*, *year*, and *government*? First, I discarded the possibility of phonemes as translation units of *market*, *year*, and *government*. According to Barkhudarov (1993), phonemes can be translation units only when the names of people and places are translated; my sample in this study however, did not contain these types (see 3.2). Also, I discarded the possibility of morphemes as translation units of the three nouns. This is because the English and Japanese languages are unlikely to share morphological similarities. I also discarded the possibility that the single word *market* would be a translation unit. In order for *market* to be a translation unit, it would have to have only one translation equivalent in the corpus. However, a comprehensive bilingual dictionary (Genius English-Japanese Dictionary 2001) shows more than 10 translations of *market*. It is unlikely that a large sample of corpus lines of *market* has only one translation equivalent. Similarly, both *year* and *government* have more than 10 translations in the dictionary; therefore, the single words *year* and *government* cannot be translation units either.

On the other hand, phrases are likely to be translation units of *market*, *year*, and *government*. As Vinay and Darbelnet state, 'a unit of translation provides a limited context: it forms a syntactic unit where one element determines the translation of the other' (1995: 27); therefore phrases give context to render a word appropriately. In this sense, clauses and sentences also give context to determine the translations. However, I discarded the possibility of units beyond sentences as translation units. This was because my corpus, the ARC (see below), is a sentence-aligned parallel corpus. It does not allow the user to trace back to the original texts; therefore, contexts wider than the sentence level is impossible to examine. Thus, it was hypothesised that translation units of *market*, *year*, and *government* were likely to be found at the levels of phrase, clause, and sentence. To clarify my argument, these three levels were distinguished in this paper using the Hallidayan rank scale (Eggins 1994: 129-138); 'prepositional phrase, adverbial phrase, and nominal phrase' belong to phrases; 'finite, non-finite, dependent clause, subordinate clause, and relative clause' belong to clauses; and the sentence is an orthographic unit divided by periods.

3. Data and Methodology

3.1. The ARC

Corpora for identifying translation units should be chosen carefully. In particular, the directionality is an important concern. This is simply because the translation unit is a unit in the original texts. For example, if one would like to identify English translation units, one has to use parallel corpora consisting of English originals and their Japanese translations; while, if one would like to identify Japanese translation units, one has to use parallel corpora consisting Japanese originals and their English translations. My target words were English; therefore, the former type of parallel corpus was required. The Alignment of Reuters Corpora (ARC) at the National Institute of Standards and Technology (NIST) is a uni-directional parallel corpus consisting of (British) English original Reuter news texts and their Japanese translations (Ian Soboroff personal communication), and is therefore suitable for this study. The ARC is the largest English-Japanese parallel corpus for this direction available in the public domain: 1.9 million for English texts and 2.2 million morphemes for Japanese texts[2], published in 1996 and 1997[3] (Utiyama and Isahara 2003).

3.2. Method

The most frequent nouns (*market*, *year*, and *government*) occurred 7,950 times, 6,402 times, 3,273 times in the ARC, respectively. Taking *market* as an example, I first discarded examples where *market* was used in proper nouns such as *Federal Open Market Committees*; as it is obvious that these are inseparable translation units, and they were not the focus of this study. If *market* appeared capitalised in the middle of a sentence, it was removed. Second, I randomly extracted 2,000 lines of *market* and their corresponding Japanese lines and identified the Japanese translations of *market* semi-automatically using ParaConc (v1.0.269) (Barlow 1995) and with manual checking afterwards. Third, I discarded duplicated examples and zero translations where *market* was not rendered into Japanese. Zero translations tell little for identifying translation units; therefore, they are not examined in this study. Finally, I selected the first 1,000 pairs of examples of *market* for analysis. I chose 1,000 pairs of samples of *year* and *government* in the same way.

I hypothesised that the single word *market* was not a translation unit. The first task was to examine if this was supported by the 1,000 samples in the ARC. The criterion was a one-to-one relationship between a translation unit and its equivalent. If *market* was always translated into one Japanese translation equivalent in the 1,000

---

[2] The size is counted by the WordSmith (version 4.0.0.365).
[3] http://trec.nist.gov/data/reuters/reuters.html

lines, 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, *market* was not considered as a translation unit, which indicated that translation units were lexical units larger than *market*.

The next task was to extract lexical units larger than *market* and to see whether they could be regarded as translation units or not. I extracted two/three/four/five/six-word collocations by the clusters function (minimum frequency was 3 within the R5-L5 window) in WordSmith (v. 4.0.0.365). There were 363 two-word, 169 three-word, 45 four-word, 7 five-word collocations, and 1 six-word collocation. Not all the collocations were examined. I first removed collocations which did not contain *market* (the 'collocate' list function in WordSmith can contain spurious hits without the search word). Second, I discarded non-syntactic collocations since translation unit was defined as a syntactic unit. The 'collocate' list consists of any adjacent collocations regardless of their grammatical relationship so that some entries are ungrammatical units. After these filters, I was left with 40 two-word, 43 three-word, 22 four-word, 4 five-word collocations, and 1 six-word collocation They were examined to see if any of them could be regarded as translation units or not based on my criteria.

There is one thing to note regarding my criteria for translation units. I made one exception to my criteria in that an item was regarded as a translation unit, even though an item had more than one translation equivalent, only if one of them was very dominant (85 percent of the examples). For example, *market sources* occurred in fourteen lines; thirteen 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. By saying creative translations, I refer to 'the creating of new words' and 'the novel collocation of existing words' (Kenny 2001: 73). The reason I set 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 was appropriate or not later in the paper.

4. Analyses

4.1. Market

In the 1,000 pairs of translations, *market* was translated into nineteen Japanese translations. Table 2 shows the top five (each translation is listed with (i) the first two definitions from the Genius Japanese-English Dictionary (2003) to provide a rough definition, (ii) parts of speech, and (iii) raw frequencies and the percentage occurrence

out of the 1,000 instances).

	Japanese translations			freq.	ratio
1	*Shijou*	(market/marketplace)	noun	789	79
2	*Souba*	(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

Table 2: *Market* and its Japanese translations

All the Japanese translations in Table 2 were nouns. The most frequent one *shijou* was very dominant, appearing in 789 lines out of the total 1,000 examples. The second one *souba* was still frequent, occurring more than 100 times in the samples. The rest of the translations appeared in less than 5 percent of the total examples. This indicates that *market* was mainly translated into either *shijou* or *souba*. Since it has more than one translation equivalent and the most dominant translation equivalent does not achieve a high coverage of 85 percent, the single word *market* was not regarded as a translation unit.

There were 110 collocations (40 two-word, 43 three-word, 22 four-word, 4 five-word collocations, and 1 six-word collocation). Based on my criteria, eleven of them could be regarded as translation units. Most of them were nominal phrases; the translation units and their corresponding translation equivalents were listed in Table 3 (the raw frequencies in parentheses and the translations of *market* are shown in bold). Word segmentation in Japanese was made by the ARC.

	Translation units	Translation equivalents
a.	*market sources (14)*	***shijou** suji (13)*
b.	*market rumours (4)*	***shijou** no uwasa (4)*
c.	*market economy (3)*	***shijou** keizai (3)*
d.	*market reforms (3)*	***shijou** kaikaku (3)*
e.	*market interest rates (3)*	***shijou** kinri (3)*
f.	*The stock market (21)*	*kabushiki **shijou** (20)*
g.	*The Seoul stock market (5)*	*souru kabushiki **shijou** (5)*
h.	*The copper market (4)*	*dou **shijou** (4)*
i.	*gold market (4)*	*kin **shijou** (4)*
j.	*domestic gold market (3)*	*kokunai kin **shijou** (3)*

Table 3: Translation units of *market*

All the translation units in Table 3 were nominal phrases and their translation equivalents were also nominal phrases. There are three things to mention. First, some translation units were singular and some were plural. According to Baker (1992: 87), Japanese nouns do 'not normally indicate whether [they are] singular or plural'. Does

this mean that both *market sources* and *market source* share the same translation equivalent *shijou suji*? I looked up all the 10 translation units to see if this was the case. The 1,000 samples had two pairs: *market source/s* and *market rumour/s*. Both *market sources* and *market source* were translated into *shijou suji*; similarly, both *market rumour* and *market rumours* share the same translation equivalent, *shijou no uwasa*. This indicates that singular and plural forms did not affect translation equivalence in the case of *market*.

Second, all the translation units in Table 3 were composed by modifying and head nouns. *Market* was used as a modifying noun (Table 3*a-e*) or as a head noun (Table 3*f-j*).What happens if another modifier appears in the translation unit? Do *market economy* and *market sector economy* share the same translation equivalent as *market economy*? The 1,000 samples showed that *market economy* was translated into *shijou keizai* and *market sector economy* was also rendered into *shijou keizai*. The modifier *sector* was not translated into Japanese; it was lost in translation. Thus, this is not a good example for examining whether modifiers affect translation equivalence or not. I looked up the rest of the 9 translation units in the 1,000 samples to investigate how modifiers affect translation equivalence and found the pair of examples: *gold market* and *gold futures market*. The former had *kin shijou* as its translation equivalent; while, *gold futures market* was translated into *kin no sakimon torihiki* (future trade of gold). By having the modifier *futures*, *gold market* was not translated into *kin shijou* any longer. Also, the grammatical structure changed as well. Therefore translation equivalence is likely to be broken down by modifiers. Adjacency mattered in translation units.

Third, some translation units had the determiner *the* (Table 3*f-h*) and some did not. The reason the three translation units (*the stock market, the Seoul stock market*, and *the copper market*) had *the* was that they always occurred with *the*. On the other hand, the reason the rest of the translation units did not have *the* was that their occurrences were not restricted with *the*. For instance, both *the market economy* and *a market economy* were found in the sample and both were translated into *shijou keizai. Market rumours* appeared with *the* or without any determiners; they both were translated into *shijou no uwasa* as well (Table 4). Determiners *a* and *the* did not affect translation equivalence in the case of *market*.

Translation units	*the*	*a*	no determiner
a. *market sources (14)*			✓
b. *market rumours (4)*	✓		✓
c. *market economy (3)*	✓	✓	
d. *market reforms (3)*			✓
e. *market interest rates (3)*			✓
i. *gold market (4)*	✓	✓	
j. *domestic gold market (3)*	✓	✓	

Table 4: *Market* and determiners

There was one more translation unit of *market*: *the market will reopen (8)*. This was a clause and its translation equivalent was a clause as well: *torihiki wa saikai sa reru (7)*. It was noteworthy that *market* in the translation unit was translated into *torihiki*. This was the only one translation unit in which *market* was rendered into *torihiki*; *market* in all the other translation units was rendered into *shijou*. However, when a modifier was present, the situation changed. There were two examples of *the London market will reopen*; *market* was translated into *shijou*, not *torihiki* any longer. Adjacency matters in the case of clausal translation units of *market* as well.

4.2.   Year

The second frequent noun in the ARC was *year*. It was translated into Japanese in various ways: twenty-five translations of *year* were found. The top five are shown in Table 5.

	Japanese translations			freq.	ratio
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

Table 5: *Year* and its Japanese translations

Table 5 shows that there were two distinctive features of how *year* was translated into Japanese. First, all the Japanese translations in Table 5 were nouns, and there is no very dominant Japanese translation of *year*. The most frequent translation *nen* only appeared in 302 lines out of the total 1,000 examples; if one compares, the most frequent translation of *market* appeared in 789 out of 1,000 lines (Table 2), this is far less dominant. Second, the second translation *kotoshi* denotes 'this year'. There is a single word denoting 'this year' in Japanese, while, there is no single word denoting 'this year' in English. Thus, when *year* was translated into Japanese, the translator needs to know the context, whether *year* appeared in *this year* or just *year*. This clearly indicates that *year* is not a translation unit as it cannot be translated as a unit; it needs context to be translated appropriately. This is not restricted only to *kotoshi*. The rest of the translations (*zennnen*, *rainen*, and *sakunen*) also belong to this case. *Zennen* denotes 'the previous year'; *rainen* denotes 'next year'; and *sakunen* denotes 'last year'. These Japanese words are single words; while, there are no single words denoting 'the previous year', 'next year', and 'last year' in English. Thus, when *year* was translated into Japanese, translators have to be aware of this lexical difference and be aware of what comes before *year* to produce appropriate translations. The single

word *year* was clearly not a translation unit since (a) it has more than one translation and therefore, (b) it needs more lexical elements when it is rendered.

There were 108 collocations (19 two-word, 28 three-word, 26 four-word, 17 five-word, and 17 six-word collocations). Based on my criteria, only three of them were regarded as translation units of *year*.

Translation units	Translation equivalents
a. *a year ago (42)*	***zennen*** *(36)*
b. *the marketing year (8)*	***nendo*** *(8)*
c. *the rest of this year (3)*	***kotoshi*** *(3)*

Table 6: Translation units of *year*

All the translation units in Table 6 are nominal phrases. Their corresponding translation equivalents, however, are not nominal phrases; they are single nouns. The linguistic ranks were shifted in translation. Another attribute found in the translation units concerned singular and plural variations. In the case of *market*, the singular and plural forms did not affect translation equivalence; *market sources* and *market source* were translated into the same translation equivalent. Would it be the same result with the translation units of *year* as well? I looked up all three translation units in the 1,000 samples and found one pair: *a year ago* and *years ago*. The former was rendered into *zennen*; while, the latter was rendered into *nen mae* (year ago). They did not share the same translation equivalent, which is a different result from the case of *market*. The other two attributes (adjacency and *a/the* variations) could not be investigated: there were not enough examples of these cases in the 1,000 concordance samples.

There was one more translation unit of *year*: *compared with the same week a year ago (3)*. It was a clause and the translation equivalent was a nominal phrase with a particle, *zennen douki hi de (3)* (in comparison with the same week a year ago). The linguistic rank was shifted in translation here as well. The translation unit had a variation: *compared with the same period a year ago (1)*. It had *period* instead of *week*; however, they shared the same translation equivalent. Again, the rank was shifted in translation.

4.3. *Government*

The third frequent noun in the ARC was *government*. It had fourteen Japanese translations and all of them were nouns. The top five are shown in Table 7.

	Japanese translations			freq.	ratio
1	*seifu*	(government/administration)	noun	849	85
2	*seiken*	(power/regime)	noun	82	8
3	*toukyoku*	(authority)	noun	23	2
4	*seichou*	(government office[4])	noun	19	2
5	*naikaku*	(cabinet/ministry)	noun	8	1
			total	981	98

Table 7: Government and its Japanese translations

The most frequent one was very dominant, appearing in 849 lines out of the 1,000 examples of *government*. The second translation equivalent *seiken* only occurred 82 times. The frequency gap between *seifu* and *seiken* was quite large. The rest of the translations took up only 2 or less percent of the total examples. Based on my criteria, *government* was regarded as a translation unit since it had one dominant translation which took up 85 percent of the total examples. Unlike *market* and *year*, a single word *government* was regarded as a translation unit.

The two attributes of translation units were also examined. First, singular and plural variation did not affect translation equivalence. There was one example of *governments* in the 1,000 samples and it was translated into *seifu*. *Government* and *governments* share the same translation equivalent. Second, *a/the* forms did not affect translation equivalence either. *Government* occurred both with *the* and *a*; *a government* was translated into *seifu* in forty-six lines out of forty-eight (96 percent) and *the government* was rendered into *seifu* in 360 lines out of 385 (94 percent).

4.4. Implications

My hypotheses regarding translation units of frequent nouns should be revisited. I expected that single words would not be translation units but units larger than that (phrases, clauses, and sentences) would be. For *market* and *year*, these hypotheses were supported to some extent by the analyses. The single words *market* and *year* could not be regarded as translation units but two/three/four/five/eight-word collocations were. Most of them were phrases; however, there were two clausal translation units and one of them was a sentence as well. As for *government*, however, the hypotheses were not supported by the analyses. The single word *government* was regarded as a translation unit.

Four attributes of translation units were revealed: rank shifting, singular/plural variations, modifier variations, and *the*/*a* variations. First, all the translation units of *market* and *government* did not shift their ranks through translation. *Market* had ten nominal phrasal translation units and one clausal

---

[4] Jim Breen's WWWJDIC Japanese Dictionary (http://www.csse.monash.edu.au/~jwb/wwwjdic.html). Translation of *seicho* was not listed in the Genius Japanese-English Dictionary (2003).

translation unit. The phrases were translated into Japanese phrases; the clause was translated into a Japanese clause. Similarly, the single word *government* was translated into a single Japanese word. No rank shifting occurred in the case of *market* and *government*. However, this was not the case for *year*. All the nominal phrasal translation units of *year* were translated into single words in Japanese; the clausal translation unit of *year* was translated into a nominal phrase in Japanese. Rank was shifted in the case of *year*.

Second, singular and plural forms in translation units did not affect the translation equivalence in the case of *market* and *government*. *Market sources* and *market source* shared the same translation equivalent. However, this was not the case for *year*. *A year ago* and *years ago* were translated into Japanese differently. Considering what Baker says (Japanese nouns do 'not normally indicate whether [they are] singular or plural' (1992: 87)), *year* is an exceptional noun in this aspect. Third, adjacency mattered in translation units. By having modifiers in translation units, translation equivalence was likely to be broken. Fourth, *a/the* forms did not affect translation equivalence. No matter whether *a* or *the* occurred before translation units, they were translated into the same translation equivalents.

5. Discussion

I have identified eleven translation units for *market*, four translation units for *year*, and one translation unit for *government*. What does this mean? If one encounters one of the translation units of *market*, one knows how it should be rendered into Japanese. If one encounters *market economy*, then it has to be translated into *shijou keizai*. One-input-and-one-output-system is simple and convenient to handle for translators. However, how often do translators encounter one of those eleven translation units of *market*? Do they cover all the 1,000 examples of *market*? I examined how many examples of the 1,000 had one of the translation units. There were only eighty-three lines out of 1,000 (8.3 percent). The eleven translation units only cover 8.3 percent of the 1,000 examples of *market*. In other words, identification based on collocation failed to extract translation units from 917 lines out of 1,000. As for *year*, I identified four translation units; however, they cover only fifteen lines of the total 1,000 samples (1.5 percent). In other words, identification based on collocation failed to extract translation units from 985 lines out of 1,000. On the other hand, *government* was quite different. The single word *government* was regarded as a translation unit and its translation equivalent was *seifu*. This covered 849 lines out of 1,000 examples (85 percent).

There were four contributing reasons for why I only found translation units for such a tiny portion of examples of *market* and *year*. First, I examined grammatical collocations occurring three times or more: 110 collocations of *market* (covering 754 lines in the 1,000 samples of *market*) and 108 collocations of *year* (covering 953 lines

in the 1,000 samples of *year*). At this point, I discarded 246 lines of *market* and 47 lines of *year* out of 1,000. Second, I set a filter of 85 percent as a criterion for determining translation units (discussed in 3.2). However, this filter was obviously not useful for less frequent collocations. If an item occurred six times and five of them had a Japanese translation equivalent, it would only be 83 percent; 85 percent is not achievable. In order to make 85 percent achievable, an item had to occur at least seven times in the 1,000 samples. Among 110 examined collocations of *market*, only forty-four of them occurred seven times or more; among 108 examined collocations of *year*, only forty-five of them occurred seven times or more. Most of the collocations were less frequent ones.

Third, I examined only lexical collocations. However, it can be argued that translation units are not necessarily lexical. Sinclair proposes four types of co-occurrence relations: collocation, colligation ('the co-occurrence of grammatical choices' (Sinclair 1996: 85)), semantic preference ('the co-occurrence of words with semantic choices' (Sinclair 2004b: 174)), and semantic prosody (an attitudinal or pragmatic meaning (Sinclair 2004a: 292)). Further investigation is required to discover whether, (i) colligations of *market* such as '*market for* N', (ii) semantic preferences of *market* such as '*market* associated with stock exchange', (iii) semantic prosodies of *market* such as '*market* in negative sense', can also be translation units or not. Finally, this study was based on 1,000 pairs of English and Japanese lines. In order to examine several-word collocations, 1,000 examples are not enough; however, it was a manageably large set if the manual translation matching of 1,000 pairs is considered.

6. References

Altenberg, B. and S. Granger (2002). 'Recent Trends in Cross-Linguistic Lexical Studies', in Altenberg, B. G., Sylviane (ed.) Lexis in Contrast, pp. 3–50. Amsterdam: John Benjamins.

Alves, F. and J. L. V. R. Gonçalves (2003). 'A Relevance Theory Approach to Inferential Processes in Translation', Triangulating Translation: Perspectives in Process Oriented Research, pp. 3–24. Amsterdam: John Benjamins.

Baker, M. (1992). *In Other Words*. London: Routledge.

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. Available on-line from http://www.ruf.rice.edu/~barlow/pc.html, (accessed: 11 April, 2004)

Bell, R. T. (1991). *Translation and Translating: Theory and Practice*. London: Longman.

Bennett, P. (1994). 'The Translation Unit in Human and Machine'. *Babel* 40, 12–20.

De Beaugrande, R. (1978). *Factors in a Theory of Poetic Translating*. Assen Van Gorcum.

Eggins, S. (1994). *An Introduction to Systemic Functional Linguistics*. London: Pinter.

Genius English-Japanese Dictionary (2001). 3rd edition. Tokyo: Taishukan.

Genius Japanese-English Dictionary (2003). 2nd edition. Tokyo: Taishukan.

Kenny, D. (2001). *Lexis and Creativity in Translation*. Manchester: St. Jerome.

Malmkjær, K. (1998). 'Unit of Translation', in Baker, M. (ed.) Routledge Encyclopedia of Translation Studies, pp. 286–88. London: Routledge.

Malmkjær, K. (2006). 'Translation Units', in Brown, K., A. H. Anderson, L. Bauer, M. Berns, G. Hirst and J. Miller (eds.) Encyclopaedia of Language and Linguistics (2nd Edition), pp. 92. London: Elsevier.

Newmark, P. (1988a). *Approaches to Translation*. Hemel Hempstead: Prentice Hall,.

Newmark, P. (1988b). *A Textbook of Translation*. London: Prentice-Hall.

Shuttleworth, M. and M. Cowie (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. (2004a). 'New Evidence, New Priorities, New Attitudes', in Sinclair, J. M. (ed.) How to Use Corpora in Language Teaching, pp. 271–300. Amsterdam: John Benjamins.

Sinclair, J. M. (2004b). *Trust the Text*. London: Routledge.

Teubert, W. (2001). 'Corpus Linguistics and Lexicography'. *International Journal of Corpus Linguistics* 6, 125–53.

Teubert, W. (2002). 'The Role of Parallel Corpora in Translation and Multilingual Lexicography', in Altenberg, B. and S. Granger (eds.) Lexis in Contrast, pp. 189–214. Amsterdam: John Benjamins.

Teubert, W. (2004). 'Units of Meaning, Parallel Corpora, and Their Implications for Language Teaching', in Connor, U. and T. A. Upton (eds.) Applied Corpus Linguistics: A Multidimensional Perspective pp. 172–89. Amsterdam: Rodopi.

Toury, G. (1986). 'Monitoring Discourse Transfer: A Test-Case for a Developmental Model of Translation', in House, J. and S. Blum-Kulka (eds.) Interlingual and Intercultural Communication : Discourse and Cognition in Translation and Second Language Acquisition Studies, pp. 79–94. Tübingen Gunter Narr.

Utiyama, M. and H. Isahara (2003). 'Reliable Measures for Aligning Japanese-English News Articles and Sentences'. *ACL-2003*, 72–79.

Vinay, J. P. and J. Darbelnet (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. *School of Humanties, the University of Birmingham,* MPhil thesis.