

# A Corpus-Based Study of the Order of Adjectives in Japanese Antonym Sequences: The Role of Markedness and Frequency

Phornthip Nithithanawiwat Faculty of Humanity of Social Sciences, Rajabhat Bansomdej Chaopraya University, Bangkok, Thailand nithithanawiwat@gmail.com

Received 4 September 2019 | Revised 1 March 2023 | Accepted 8 March 2023 | Published online 17 October 2023

# Abstract

This study investigates whether the order of adjectives in antonym pairs (e.g., *ii* 'good, *warui* 'bad') in single sentences of Japanese follows principles of markedness and whether the more frequently used antonym is more likely to be used first. Based on data collected from the Tsukuba Web Corpus, the results of the study were that both markedness and frequency equally influence the order of antonym pairs. This study suggests that markedness plays a role in determining the ordering of adjectives in antonym pairs. Additionally, it was found that markedness affects the order in terms of factors of positivity and magnitude.

# Keywords

antonyms in Japanese – principle of markedness – frequently used – /i/ adjectives

# 1 Introduction

Antonymy is a term used in linguistics as part of the study of the oppositeness of meaning. An antonym pair consists of two members (Lehrer 1985). In many genres, use of antonym pairs is employed to create vivid contrasts within

NITHITHANAWIWAT

the context of single sentences. Whether it be on the stage when Hamlet says, "There is nothing either *good* or *bad* but thinking makes it so" or the slogan at your local Dairy Queen "*HOT* EATS *COOL* TREATS" – pairs of antonyms co-occurring are found throughout many languages. One area of interest concerning antonym pairs has been the study of principles which dictate the preferred sequence of antonyms in various languages, essentially which one precedes the other.

Antonymy is a lexical-semantic relation or a paradigmatic relation between word pairs. Because the semantic relationship in antonymous pairs is usually asymmetrical, antonymy can be related to the linguistic principle of *markedness*. Markedness is an analytic principle in linguistics whereby pairs of forms are contrasted in terms of a linguistic feature. Greenberg (1976) noted that a "marked word" is one which is less frequent, while an "unmarked word" is more frequent. For antonyms, members of antonymous pairs are often distinguished as being either *positive* or *negative*, with the positive member being unmarked and the negative member marked. More generally some have even argued that positive members tend to precede the negative members when they co-occur in the same sentence (Lyons 1977). But further research into the issue continues to reveal several factors affecting the sequencing of antonym pairs.

One major development has been the advent of corpus linguistics, which has allowed for nuanced investigation into the order of co-occurring antonym pairs. Some research has even claimed that frequency and markedness are marginal factors affecting the preferred ordering of antonym pairs (Jones 2002), but others have argued the opposite. For example, using the concept of markedness and corpus data, Kostić (2015) examined antonym sequences in Serbian to find that factors of word frequency and markedness do influence antonym ordering. Kostić also found that the members of antonymous pairs had an asymmetrical distribution and therefore proposed other factors governing sequencing, such as temporal and visual-spatial ordering. Later findings in Kostić (2017) showed the distributional asymmetries of antonym sequences in English may also be influenced by the principle of markedness, word frequency and the factors of temporal and visual-spatial ordering. Broadening this type of research cross-linguistically, Hsu (2015) examined antonyms co-occurring in Chinese, finding that both language-internal factors and language-external factors affect the ordering of antonymous pairs. Internal factors include syntagmatic functions and morphosyllabic structure, while language external factors include some socio-cultural values with regards to the lexical pairs.

When two words are related in some way (e.g., their meaning) there is said to be 'lexical cohesion'. Antonymy is one category of lexical cohesion. In English,

the word classes with antonyms include adjective, adverb, noun and verb but most research has focused on antonym pairs of adjectives (Cruse 1975; Gross, Fischer, and Miller 1989; Justeson and Katz 1991; Muehlieisen 1997; Willners 2001; Muehleisen and Isono 2009; Paradis, Löhndorf, Weijer and Willners 2015). While a study like Cruse (1976) suggests that the forms of adjectival antonyms pairs are intuitively clear (i.e., they are fully gradable adjectives), pioneering research like Jones (2002) worked to establish the function of antonyms in English, with studies on antonym pairs in nominal and verbal word classes.

Interestingly, the study of antonym pairs in Japanese presents some additional complexity. Although in Japanese some antonyms pairs seem to be the same part of speech, such as so-called/-i/ adjectives (e.g., *ii* 'good' and *warui* 'bad') or/na/ adjectives (e.g., *kirei* 'clean' and *kitanai* 'dirty'), some pairs of antonyms are found across different parts of speech such as in *toshitotta/wakai* 'aged/young', where *toshitotta* is a noun+verb compound and *wakai* is an adjective. While previous research on Japanese such as Muehleisen and Isono (2009) studied antonyms, they did not investigate antonym sequences nor explain why one member of a pair might precede the other when co-occurring in the same sentence.

This current study therefore aims to test whether different frequency patterns of the ordering in co-occurring Japanese antonym pairs follows a markedness principle or not. It will also look at whether more frequently used antonyms of antonym pairs occur more frequently in the initial position of antonym sequences in Japanese. The results of the study could potentially help Japanese language learners to be able to order antonym pairs when writing and speaking. Additionally, the findings of the study may also be useful for linguists interested in Japanese, as it applies various theoretical perspectives from linguistics to provide explanations of antonym sequencing phenomena in Japanese.

#### 2 Theoretical Background

This section reviews the concept of antonymy and the concept of markedness. Moreover, it presents a short introduction to Japanese adjectives and the syntactic frameworks of co-occurring antonyms.

#### 2.1 Antonymy, Markedness, and Preferred Pairings

Antonyms concern the structure of vocabulary and many definitions have been given for 'antonym'. Palmer (1976) defined antonymy as 'oppositeness of meaning'. Lyons (1977) defined antonym as a technical term which presents two

words of opposite sense. Lehrer and Lehere (1982) observed that antonymy can have both a wide and a narrow meaning. In the wide sense, an antonym pair can consist of words with equivalent but opposite senses (e.g., male/female, *high/low*), while the narrow sense refers to gradable antonyms like *large/small*, good/bad. Antonyms can also be found at the ends of dichotomous scales like *long/short, rich/poor*; the main point is that antonymy is particularly a binary opposition (Murphy 2003). But such a view is problematical because it encourages the idea that antonyms contrast only on a single dimension, while in reality they may contrast with other words on other dimensions at the same time (see Leech 1981). Gross, Fischer and Miller (1989) distinguished conceptual antonymy into direct antonyms and indirect antonyms. Direct antonyms are lexical pairs (e.g., *heavy/light*), while indirect antonyms are not lexical pairs (e.g., *heavy/weightless*). They noted that people can recognize direct antonyms faster than indirect antonyms. While many linguists (Palmer 1976, Lyons 1977) have attempted to classify antonyms into different types - namely gradable antonyms (e.g., *beautiful/ugly*), complementary antonyms (e.g., *alive/dead*) and converse antonyms (e.g., buy/sell) - for this current study, I will use the definition of antonyms merely as pair of words which are opposite in meaning.

The concept of markedness is used in several fields of linguistics. "Markedness" was proposed and defined by Trubetzkoy and Jakobson in 1930 (see Andersen 2001). Markedness at the level of vocabulary can be used to talk about a distinction between different features. For example, in English the word *bitch* is marked for gender and applies only to female canines, while the term *dog* is not marked for gender and encapsulates canines of all genders. Andersen (2001) believes that markedness often reveals asymmetrical relationships, consisting of abstract content and complexity. Moreover, he insists that with a theory of markedness it is important to make observations which can be managed in a systematic study, ultimately believing that markedness is a principle of cognitive organization reflected in human behavior.

Givón (1970) found that adjectival pairs in English can be marked or unmarked in terms of their positivity or negativity. For example, positivity was associated with unmarked generic terms used to express things which were 'abundant,' 'extensive' or 'numerous' (e.g., *long, large, high*), whereas terms expressing things which were 'meager' and 'restricted' (e.g., *short, small, low*) are negatively marked. Lyons (1977) stated that the principle of markedness is an important concept in linguistic studies on language structure. Moreover, he noted that an asymmetry of semantic distinction can be found in antonym pairs. But also, words with positive meanings tend to appear before ones with negative meanings when antonym pairs co-occur in the same sentence.

Challenging the use of the linguistic concept of markedness, Haspelmath (2006) says markedness, as it is used in linguistics, actually has several senses.

For illustrative purposes I will present a few of these senses. *Complexity* refers to the sense of formal markedness that is overt coding, such as in English where present tense verbs are unmarked but past tense forms with an -ed suffix are marked. *Difficulty* refers to the sense of constructional iconicity, in the sense of ease in acquiring the structure. For instance, English singular-plural pairs such as *girl-girls* are iconic as unmarked words, while *sheep-sheep* is non-iconic as marked words. Abnormality refers to the sense of restricted distribution. For example, with lexical semantics, English opposition, dog is an unmarked word and is general. It can be combined with the adjectives (e.g., male dog, female dog) while the term marked for gender *bitch* is less common and cannot take on adjectives indicating gender. Finally, in a broader sense, there is a feeling that there is *multidimensional correlation* of markedness, i.e., the idea that universally, comparable linguistic structures present the same markedness values for different markedness dimensions. However, with all these different senses, Haspelmath (2006)'s point is that the concept of markedness is ambiguous and should be abandoned by linguists and that *frequency* of use is more straightforward and clearly defined. Despite Haspelmath's critique of 'markedness', the concept may perhaps still be a useful framing for targeted study of many aspects of language.

One example of using markedness is Lehrer (1985). In this study, the concept of markedness was used to classify the properties of an antonym, and ultimately there are a proposed eight criteria of markedness that can predict unmarked words of antonym pairs. The eight criteria are summarized in the criteria of markedness according to Lehrer (1985).

- a. Unmarked word is the neutralization of an opposite in questions (How *big* is your house? *Small*.)
- b. Unmarked word is an opposite in nominalizations (e.g., *\*oldness/ newness*)
- c. Only the unmarked word appears in measurement phrases (e.g., five feet *tall*)
- h. If one word of the pair consists of an affix added to the antonym, the affixed form is marked (e.g., *friendly/unfriendly*)
- i. Ratios can be used only with the unmarked word (e.g., \*twice as *young*)
- j. The unmarked word is given as a positive value and the marked word is negative (e.g., *good/bad*)
- k. The unmarked word denotes more of a quality, the marked word denotes less (e.g., *large/small*)
- l. If there are asymmetrical entailments, the unmarked word is less likely to be biased or committed, for example, A is *better* than B, A and B could be *bad*, B is *worse* than A, B must be *bad*, but A may be *bad* as well

NITHITHANAWIWAT

Lehrer's treatment of antonymy looks at all the ways in which a member of an antonym pair may be marked. Interestingly while antonymy pairs can be marked or unmarked in various ways, it is suggested that it is the antonym with the most unmarked properties which is likely to appear first when they co-occur.

Further research in Jones (2002) looked at 53 antonym pairs in English and found that seven basic factors affect the sequencing of antonym pairs, namely: morphology, positivity, magnitude, chronology, gender, phonology and idioma*ticity*. Other minor factors having a marginal influence on antonym sequence were frequency and markedness. Regarding morphology, antonymous pairs tend to order their root word to precede their derivation (e.g., advantage precedes disadvantage). If an antonymous pair has one member with a positive meaning and the other with a negative meaning, well then, the positive meaning tends to precede the negative meaning (e.g., good precedes bad and *confirm* precedes *deny*). Similarly, the concept of size is significant in antonym sequencing in that words showing greater magnitude occurred first (e.g., *long* precedes short). Chronology refers to how antonymous pairs which have a temporal aspect are sequenced accordingly, for example, *begin* precedes *end*. With regards to gender the bias tends to be male terms precede female, for example, he before she and him before her. Another factor which apparently impacts antonym ordering is phonology with shorter words tending to occur before longer words (e.g., fact precedes fiction). Additionally, the role of idiomaticity cannot be overlooked as some antonymous pairs have particular coinages, such as war and peace, which clearly flaunt a factor like positivity. Finally, Jones (2002) also acknowledges the role of frequency and markedness but holds that they were less influential than the criteria outlined above. One example of this is the case of *rich* and *poor*, where although *poor* is the higher-frequency word, *rich* is more likely to precede it when they co-occur.

Like Jones (2002), Kostić (2015) was able to outline several factors which influenced the preferred order of antonyms pairs in written Serbian. Key for Kostić was the idea that the preference of a particular sequence in the same sentences correlates to the concept of markedness. But more specifically, she found that the factors influencing antonym ordering were *frequency*, *positivity*, *neutrality*, *temporal and visual-spatial order*, *morphology*, *quantity* and *gender*. Interestingly, in Kostić (2017) these factors were applied to a study of English antonyms and the resulting findings differed from Jones (2002). The most significant factors Kostić argued were factors of temporal and visual ordering as well as the principle of markedness and more general word frequency. One thing to keep in mind, though, is the possibility that different factors can have differing levels of significance in different languages, as with Hsu (2015)'s study of Chinese where the factor of positivity was the most outstanding factor.

### 2.2 Japanese Adjectives

Japanese grammar as it is formally taught divides words into two major groups (Kishimoto and Uehara 2016). The first group is *jiritsugo* 'independent categories' which consists of nouns, verbs, adjectives, adjectival nouns, adverb prenominal modifiers, conjunctions and interjections. The second group called *fuzokugo* 'dependent categories' consisting of auxiliaries and particles. Japanese adjectives are categorized into two subtypes. The first type is called the/-i/ adjective, canonical adjective or, simply, adjective (A), and the second type is called the/na/ adjective, nominal adjective (NA) or adjectival noun (AN) (Backhouse 1984; Nishiyama 1999; Spencer 2008). The following are examples of how Japanese adjectives can modify nouns:

(1)	/-i/ adjective	Noun
	atarashii	kuruma
	new	car
	'new car'	

(2) /na/ adjective Noun shizuka (na) machi quiet city 'quite city'

The present-tense form of the/-i/ adjective has/-i/ appear stem-finally, while the present form of the/na/ adjective adds/na/ to the final syllable of the source lexeme. The two subtypes of Japanese adjectives are different in their grammatical patterns in that/na/ adjectives share various properties with nouns, whereas the/-i/ adjective can be inflected for tense similar to a verb (see Ohkado 1991). Backhouse (1984) working on identifying the grammatical properties of adjectives in Japanese noted that/na/ adjectives are different in terms of morphology and syntax. In the case of a conjunction between the/-i/ adjective and the/na/ adjective, the first/-i/ adjective changes the ending/-i/ as *kute*, for instance, *oishikute*, *yasui* 'delicious and cheap'. In contrast, the/na/ adjective will add de in the ending word such as hansamu de yasashii 'handsome and kind'. Muehleisen and Isono (2009) expound that the/-i/ adjectival antonyms are gradable antonyms, whereas the/na/ adjectival antonyms are complementary and other types of antonyms. But comments in Backhouse (1984) also point out that the phenomenon of cross-classification between/na/ adjectives and nouns is confusing, for instance, shiawase 'happiness' is a noun and means 'happy' in the/na/ adjective form.

## 2.3 The Syntactic Frameworks of Co-Occurring Antonyms

The syntactic frameworks of co-occurring antonymous pairs in discourse are identified by many linguists. Jones (2002), using written English corpora, showed the syntactic frameworks of antonymous pairs in one sentence by dividing them into major classes: *ancillary antonymy* and *coordinated antonymy* and minor classes: *comparative antonymy, distinguished antonymy, transitional antonymy, negated antonymy, extreme antonymy* and *idiomatic antonymy.* An example of each of the syntactic frameworks of co-occurring antonymous pairs are as follows:

- a. Ancillary antonymy is structured as parallelism and ellipsis.
  - It is meeting **public** *need* not **private** *greed*.
- b. Co-ordinated antonymy appears with the conjunction as a framework: both *X* and *Y*; *X* as well as *Y*.
  - The policy is to recruit *skilled* and *unskilled* workers.
- c. Comparative antonymy appears when one antonym is measured against the other as a framework: more *X* than *Y*; *X* is more [adj.] than *Y*; *X* rather than *Y*.
  - Sometimes I feel more *masculine* than *feminine* and I don't like it.
- d. Distinguished antonymy refers to the inherent dissimilarity between a word pair as frameworks: the difference between *X* and *Y*; separating *X* and *Y*; a gap between *X* and *Y*.
  - The gap between the *rich* and the *poor* has widened.
- e. Transitional antonymy shows changing from one location or state to another or a movement as frameworks: from *X* to *Y*; turning *X* into *Y*.
  - It is easy to slip from *legal* to *illegal* trade.
- f. A negated antonym is the co-occurring of an antonym pair within a framework that negates one antonym as a device to augment the other: *X* not *Y*; *X* instead of *Y*.
  - ...to facilitate the re-establishment of *peace* not *war*.
- g. An extreme antonymy unites the outermost margins of a semantic scale as frameworks: the very *X* and the very *Y*; too *X* or too *Y*.
- h. ... except when the soil is too *wet* or too *dry*.
- i. Idiomatic antonymy is found in idioms and proverbs that would be recognized as familiar.
  - Penny *wise* and pound *foolish*.

With regards to Japanese, Muehleisen and Isono (2009) investigated antonymous adjectives in Japanese discourse. Looking at twelve pairs of 'basic' antonyms, they identified the common grammatical constructions in which they occur and their discourse functions. Moreover, they found that the discourse functions of antonym pairs identified for other languages such as those in Jones (2002) were useful in studying Japanese. For example, in (1), where *simultaneous antonymy* appears in a construction where the two adjectives are past tense inflections, they modify the same noun phrase as *ichi-nichi* 'day'. Note that both antonyms *nagai* 'long' and *mijikai* 'short' are followed by the auxiliary adjective *you-na*.

(1)	naga-katta	you-na	mijika-katta	you - na
	long - рт	seeming	short- рт	seeming
	ichi-nichi	ga	owatta	
	one- day	SB	was-finished	
	'The day tha	it had seem	ed long and sh	ort was finished.'

Or for disjunctive antonymy, see how the particle ka is used for combining with nouns or sentences to mark alternatives, as shown in (2).

(2)	tsuyo-ku	furumau	ka	yowa-ku	furumau	ka
	Strong-CT	behave	IP	weak-ст	behave	IP
	'Should I b	ehave stron	gly or	should I beh	ave weakly?'	

Moreover, studies like Martin (2004) found that there is a syntactic framework for displaying antonym pairs as follows: [*kore wa* ADJ *ga, sore wa* ADJ], like in (3).

(3) *kore wa ookii ga, sore wa chiisai* this TOP big but that TOP little 'This is big but that is little'

Previous research into the Japanese syntactic frameworks for presenting antonymous pairs is useful as it outlines targets for further investigation.

### 3 Research Methodology

This current study aimed to continue previous research on Japanese adjective pairs with acknowledgment of the aforementioned factors influencing them in Japanese. This study aimed to divide pairs of words into unmarked words and marked words by adapting concepts from Lehrer (1985), especially where the criteria were applicable to Japanese.

This study focuses on/-i/ adjectives because they are gradable and are seemingly more easily identified as being either marked or unmarked. This study analyzes only/-i/ adjectives with the selected words meeting a frequency level from 0.1% to 1.0 % (110,000–1,100,000) in a web corpus. Data for this study was taken from the Tsukuba Web Corpus<sup>1</sup> during September - November 2017. This is one of the largest corpora of the Japanese language and it consists of data from NINJAL-LWP (National Institute for Japanese Language and Linguistics-Lago Word Profiler) and BCCWJ (Balanced Corpus of Contemporary Written Japanese). The Tsukuba web corpus has stored about 110 million words from many URLS, HTML pages, magazines, legal documents, textbooks and even works of poetry. The first step was the selection of adjectives from the web which met frequency requirements. From this step, a total of 25 adjectives were selected. Then, 25 adjectives having the opposite meaning were chosen from Kitahara and Togo (1989)'s Dictionary of Antonyms and Contrasting Words, a work which contains about 13,000 words. The opposite words were put in a word list and pairs of words which had more than one antonym were counted. For example, tanoshii/ kurushii 'happy/painful' and tanoshii/tsurai 'happy/bitter' were counted as two separate pairs. From these 25 pairs 19 had the frequency requirements. Table 1 provides the 19 pairs that were searched in the database. Column A presents the words which met the frequency requirements (0.1% to 1.0%) in the database. Column B shows their antonym. Column 3 presents the number of occurrences for the words in Columns A and B. The pairs in Table 1 were then screened as to whether or not they co-occurred in Japanese sentences from Tsukuba Web Corpus. Additionally, whether they followed Muehleisen and Isono (2009)'s syntactic frameworks was also noted. The hypothesis tests for a population proportion (p-value below 0.05 is statistically significant) were employed in order to analyze the data. An additional step here was to assess if the sample from the population would display the true proportion for the whole population. The 19 antonym pairs needed to be frequently used words and they needed to be frequent enough to ensure a sampling distribution of p was reasonable. The data was analyzed based on hypothesis tests for a population proportion. This study determined that both

 $np \ge 5$  and  $nq \ge 5$ .

The sample size is calculated using the following formula:

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

<sup>1</sup> https://tsukubawebcorpus.jp/.

B

Occurrences	in the corpus
Α	В

А

		Α	В
ooi 'many'	<i>sukunai</i> 'few'	1,035,357	278,384
<i>ii</i> 'good'	<i>warui</i> 'bad'	775,868	259,495
<i>takai</i> 'high'	hikui 'short'	554,669	141,126
takai 'expensive'	<i>yasui</i> 'cheap'	554,669	103,616
<i>ookii</i> 'big'	<i>chiisai</i> 'small'	436,255	140,345
<i>tsuyoi</i> 'strong'	<i>yowai</i> 'weak'	338,403	66,963
atarashii 'new'	<i>furui</i> 'old'	256,164	68,585
<i>nagai</i> 'long'	<i>mijikai</i> 'short'	229,447	79,969
<i>muzukashii</i> 'difficult'	yasashii 'easy'	227,596	2,322
<i>tanoshii</i> 'fun'	<i>kurushii</i> 'painful'	197,716	33,697
<i>tanoshii</i> 'fun'	<i>tsurai</i> 'painful'	197,716	65,901
hayai 'fast'	<i>osoi</i> 'slow'	186,554	63,129
<i>umai</i> 'skillful'	<i>mazui</i> 'unskillful'	162,452	11,557
<i>fukai</i> 'deep'	asai 'shallow'	149,429	20,461
<i>chikai</i> 'near'	<i>tooi</i> 'far'	146,079	53,982
hiroi 'wide'	semai 'narrow'	139,746	50,711
omoshiroi	tsumaranai	130,398	10,463
'interesting'	'uninteresting'		
ureshii 'happy'	kanashii 'sad'	118,966	32,856
oishii 'delicious'	<i>mazui</i> 'unpleasant'	118,171	11,557

where

 $\hat{p} = \frac{x}{n}$  is sample proportion p = population proportion

n = population size

For example, if n = 10 (total sentences), np = (10) (0.5) = 5, nq = (10) (0.5) = 55, then the normal approximation of the sampling distribution was considered adequate.

The aim was to ensure that sample size was sufficiently large to calculate the value of the population proportion in order to test the study's hypothesis. If a pair occurred in fewer than 10 sentences, they were not used in the study. After testing, only one pair from Table 1 was eliminated (*muzukashii* 'difficult' and *yasashii* 'easy') leaving a total of 18 anonymous pairs. All 18 antonym pairs were divided into unmarked words and marked words based on Lehrer (1985)'s framework. From this each antonym pair was examined in order to find which factors affected the preferred order of antonym sequences. A benefit of using statistical analysis with existing corpus data is that it allows for quick results. Another positive side to the method is that there is a large sized digital dataset for analysis, which provides quantifiable and testable results.

## 4 Results

The concept of markedness based on Lehrer (1985) was employed to categorize words. It was found that there were 11 antonym pairs (11 out of 18 pairs, 61 percent) where unmarked words preceded marked words. For example, umai 'skillful' precedes mazui 'unskillful' at p-value below 0.05. Moreover, it was found that there were also seven antonymous pairs (7 out of 18 pairs, 39 percent) which presented a p-value above 0.05, such as *tanoshii* 'fun' preceding kurushii 'painful', furui 'old' preceding atarashii 'new'. There were four pairs (out of 7 pairs) that did not have a preferred sequence, such as *atarashii* 'new' which does not come before furui 'old' and hayai 'fast' does not osoi 'slow'. Therefore, the results of this study provided sufficient evidence to support the hypothesis that different frequency patterns of the ordered antonym pairs in antonym sequence in Japanese follow the markedness principle. In addition, this research found that the two factors affecting the sequencing of antonym pairs in Japanese were *positivity* and *magnitude*. Table 2 and Table 3 demonstrate that factors are based on the principle of markedness by presenting antonym pairs consisting of an unmarked and marked word.

#### (1) The Factor of Positivity

The semantic relation of antonym pairs in terms of 'positivity' was also investigated. Each pair was distinguished into positive and negative senses. It was found that the notion of positivity was an important factor in the management of markedness. The more positive word, which is typically unmarked, usually occurs in the first position of an antonym sequence. As shown in Table 2, there were 8 antonym pairs whose order was governed by the factor of positivity. However, 5 pairs which were found to have a p-value below 0.05 were considered statistically significant. For example, the word *umai* 'skillful' has

A unmarked	B marked	Total database sentences	Normal sequence (Raw freq)	Percentage	p-value
<i>umai</i> 'skillful'	<i>mazui</i> 'unskillful'	33	33	100	<0.001
<i>ureshii</i> 'happy'	<i>kanashii</i> 'sad'	22	22	100	<0.001
<i>ii</i> 'good'	<i>warui</i> 'bad'	334	325	97.3	<0.001
<i>tsuyoi</i> 'strong'	<i>yowai</i> 'weak'	285	272	95.4	<0.001
<i>oishii</i> 'delicious'	<i>mazui</i> 'unpleasant'	25	23	92.0	<0.001
<i>tanoshii</i> 'fun'	<i>kurushii</i> 'painful'	12	7	58.3	0.281
<i>tanoshii</i> 'fun'	<i>tsurai</i> 'painful'	11	6	54.5	0.381
<i>omoshiroi</i> interesting'	<i>tsumaranai</i> 'uninteresting'	20	10	50.0	0.500

 TABLE 2
 The factor of positivity

a more positive meaning than the negative meaning of *mazui* 'unskillful' in 100% (33 out of 33 sentences, 100 percent) of cases. This means that in 33 sentences both of *umai* 'skillfil' and *mazui* 'unskillful' are in the same sentence but *mazui* 'unskillful' never precedes *umai* 'skillful', such as *tsuyoi* 'strong' preceding *yowaii* 'weak' 95.4% (272 out of 285 sentences, 95.4 percent), while *oishii* 'delicious' preceded *mazui* 'unpleasant' 92.0% (23 out of 25 sentences, 92.0 percent). Three antonym pairs having a p-value above 0.05 were not considered statistically significant, namely *tanoshii* 'fun' preceding *kurushii* 'painful' 58.3% (7 out of 12 sentences, 58.3 percent), *tanoshii* 'fun' preceding *'tsurai* 'painful' in 54.5% (6 out of 11 sentences, 54.5 percent) and *omoshiroi* 'interesting' preceding *tsumaranai* 'uninteresting' 50.0% (10 out of 20 sentences, 50.0 percent). In contrast consider a pair like *ii* 'good' and *warui* 'bad', where *ii* 'good' precedes *warui* 'bad' 97.3% (325 out of 334 sentences, 97.3 percent) such as in example (4).

(4)	ii-warui	wa	betsu	desu
	Good-bad	ТОР	different	be - pres
	'Good-bad a	re diffe	rent'	

## (2) The Factor of Magnitude

The notion of magnitude includes size, measurement, weight, quantity and bulk. The ordering of most antonym pairs was found to be influenced by magnitude as much as the factor of positivity. For example, cases where *ookii* 'big' preceding *chiisai* 'small' such as in (5) were found in this order in 89.7% (166

A unmarked	B marked	Total database sentences	Normal sequence (Raw freq.)	Percentage	p-value
ooi 'many'	<i>sukunai</i> 'few'	2,687	2,676	99.5	<0.001
<i>ookii</i> 'big'	<i>chiisai</i> 'small'	185	166	89.7	<0.001
<i>takai</i> 'high'	<i>hikui</i> 'short'	299	247	82.6	<0.001
<i>takai</i> 'expensive'	<i>yasui</i> 'cheap'	218	171	78.4	<0.001
<i>fukai</i> 'deep'	<i>asai</i> 'shallow'	31	22	70.9	<0.01
<i>nagai</i> 'long'	<i>mijikai</i> 'short'	104	70	67.3	<0.001
<i>furui</i> 'old'	<i>atarashii</i> 'new'	47	27	57.4	0.153
<i>hiroi</i> 'wide'	<i>semai</i> 'narrow'	38	21	55.2	0.258
<i>tooi</i> 'far'	<i>chikai</i> 'near'	54	29	53.7	0.293
<i>hayai</i> 'fast'	<i>osoi</i> 'slow'	975	206	21.1	1.000

TABLE 3. The factor of magnitude

14

out of 185 sentences, 89.7 percent) of cases. As shown in Table 3, several pairs were found to occur in relation to this factor.

(5) *nayami ni ookii, chiisai wa arimasen* Suffering COP big little TOP have-NEG 'There is nothing big little in suffering'

Other examples include: *ooi* 'many' preceding *sukunai* 'few' in 99.5% (0.001) of the cases of it co-occurring (*ooi* 'many' precedes *sukunai* 'few' in 2,676 sentences in the same sentence, while *sukunai* 'few' precedes *ooi* 'many' in 11 sentences), *takai* 'high' preceding *hikui* 'short' in 82.6% (p-value <0.05) of cases, and *nagai* 'long' preceding *mijikai* 'short' in 67.3% (p-value <0.05) of cases. Finally, *fukai* 'deep' was found to precede *asai* 'shallow' in 70.9% (p-value <0.05) of the cases of co-occurrence. There were four antonym pairs which had a p-value above 0.05, which were not considered statistically significant, namely *furui* 'old' and *atarashii* 'new'; *hiroi* 'wide' and *seimai* 'narrow'; tooi 'far' and *chikai* 'near'; and *hayai* 'fast' and *osoi* 'slow'.

Another important finding was that more frequent words preceded the less frequent ones when they appeared as an antonym pair in a single sentence, as shown in Table 4.

A	В	Word frequency	A precedes B	Frequency	p-value
		A: B	In one sentence	1	I
ooi 'many'	sukunai 'few'	1,035,357:278,384	2676:11	×	<0.001
<i>ii</i> 'good'	<i>warui</i> 'bad'	775,868:259,495	325:9	×	<0.001
<i>takai</i> 'high'	<i>hikui</i> 'short'	554,669:141,126	247:52	×	<0.001
<i>takai</i> 'expensive'	<i>yasui</i> 'cheap'	554,669:103,616	171:47	×	<0.001
<i>ookii</i> 'big'	<i>chiisai</i> 'small'	436,255:140,345	166:19	×	<0.001
<i>tsuyoi</i> 'strong'	<i>yowai</i> 'weak'	338,403:66,963	272:13	×	<0.001

TABLE 4         Frequency effects	on antonym pair sequencing
-----------------------------------	----------------------------

A	В	Word frequency A: B	A precedes B In one sentence	Frequency	p-value
<i>atarashii</i> 'new'	<i>furui</i> 'old'	256,164:68,585	20:27	*	0.846
<i>nagai</i> 'long'	<i>mijikai</i> 'short'	229,447:79,969	70:34	×	<0.001
<i>tanoshii</i> 'fun'	<i>kurushii</i> 'painful'	197,716:33,697	7:5	×	0.281
<i>tanoshii</i> 'fun'	<i>tsurai</i> 'painful'	197,716:65,901	6:5	×	0.381
<i>hayai</i> 'fast'	<i>osoi</i> 'slow'	186,554:63,129	206:769	*	1.000
<i>umai</i> 'skillful'	<i>mazui</i> 'unskillful'	162,452:11,557	33:0	×	<0.001
<i>fukai</i> 'deep'	<i>asai</i> 'shallow'	149,429:20,461	22:9	×	<0.01
<i>chikai</i> 'near'	<i>tooi</i> 'far'	146,079:53,982	25:29	*	0.706
<i>hiroi</i> 'wide'	<i>semai</i> 'narrow'	139,746:50,711	21:17	×	0.258
omoshiroi 'interesting'	<i>tsumaranai</i> 'uninteresting'	130,398:10,463	10:10	×	0.500
<i>ureshii</i> 'happy'	<i>kanashii</i> 'sad'	118,966: 32,856	22:0	×	<0.001
<i>oishii</i> 'delicious'	<i>mazui</i> 'unpleasant'	118,171: 11,557	23:2	×	<0.001

 TABLE 4.
 Frequency effects on antonym pair sequencing (cont.)

The results of the statistical test showed only 11 pairs that had a p-value less than 0.05. There were only three more frequently used words, namely *atarashii* 'new', *hayai* 'fast' and *chikai* 'near' which did not precede their antonyms. For a case like *atarashii* 'new' which is more frequent than *furui* 'old,' – *furui* 'old' was

the first in 27 sentences out of 47 sentences. Similarly, for a pair like *hayai* 'fast' and *osoi* 'slow' the more frequent word was only first in 206 sentences out of 769 sentences. Lastly, the pair of *chikai* 'near' and *tooi* 'far' only had *chikai* – the more frequent word – appear first in 25 sentences out of 54 sentences.

## 5 Discussion

The results showed that unmarked antonyms precede marked antonym pairs 61% of the time (11 out of the 18 pairs). This result offers support for the conclusions of both Lyon (1977) and Lehrer (1985). The results also show the role that markedness can play in framing linguistic research. There were however some adjectival antonym pairs which did not fit Lehrer's conception of markedness regarding antonyms. For example, while it might be expected that hayai 'fast' precedes osoi 'slow' most of the time, this study found that osoi 'slow' is mentioned more often than hayai 'fast'. Positivity and magnitude, factors which were relevant in other studies (Jones 2002; Kostić 2015; Hsu 2015), were found to be relevant for this study as well. However, the results differ some from Jones (2002) and add support for Kostić (2015). According to Jones (2002), markedness and frequency are a marginal criterion in preferred ordering, whereas Kostić argued that frequency is the most important factor in her study. In this study it was apparent that more frequently used words of antonym pairs occurred first within antonym pair sequences in Japanese. When considering syntactic frameworks, it is interesting to note that osoi 'slow' precedes haya 'fast' more often in the pattern of X kare Y kare, like in (6).

(6) oso kare haya kare minna shinimasu.
 slow or fast or everybody die -PRES
 'Slow or fast, everybody will die.'

Another common syntactic framework is X ku Y'ADJ iku + -i/ADJ' like in (7)

(7) tanoshiku kurushii taiken deshita
 fun painful one's experience be-PAST
 'It is fun and painful one's own experience'

Some sentences do not use a coordinator, namely syntactic frameworks like xy 'AdjAdj', *X*, *Y* 'Adj, Adj' line in examples (8) and (9). These syntactic frameworks seem similar to those outlined in Muehleisen and Isono (2009).

NITHITHANAWIWAT

- (8) hayaiosoi wa nai fast slow P-TOP have-NEG 'There is nothing fast slow'
- (9) *nayami ni ookii, chiisai wa arimasen.* sorrow P-DAT big little P-TOP have-NEG 'There is not *big, small* in sorrow'

From this study both markedness and frequency influence the sequencing of antonym pairs when they co-occur in the same sentence. Since unmarked words were more frequent than the marked words and the unmarked words are in the first position of antonym sequences in Japanese, this study's findings mostly support Greenberg (1976)'s proposal that frequency can distinguish marked and unmarked words when they co-occur in the same sentence. These results seem to show that, despite criticism of "markedness" in Haspelmath (2006), use of the concept here allowed for a productive look at sequences of marked and unmarked antonym pairs in Japanese. Moreover, this study provides some evidence that factors which influence Japanese's ordering of antonym pairs may be different from other languages. For example, for English, Jones (2002) found that *old* and *new* were in that order 72% of the time. But this study only found that *furui* 'old' preceded *atarashii* 'new' 57% of the time. Jones (2002) framed the ordering in terms of chronology, explaining that *old* precedes *new* because it is thought that things which are *old* come before things which are new, but Kostić (2017) showed that concepts like old and new are more likely to be governed by temporal or visual-spatial ordering. Further research should be directed to other adjective types in Japanese, as well as to other languages in order to gain insight into cross-linguistic similarities or differences regarding the sequencing of antonym pairs. Additionally, because this study was based only on written data further work with spoken corpus data may highlight interesting similarities or differences.

## 6 Conclusion

This corpus-based study used data from the Tsukuba Web Corpus to investigate factors affecting antonym sequencing when antonym pairs co-occur in one sentence. Both markedness and frequency were found to equally influence antonym sequences which co-occur in single sentences. This study concluded that the factors which govern the ordering of 11 antonym pairs in one sentence were positivity and magnitude. These factors were determined to be

18

based on markedness and frequency. The study reveals that frequently used and unmarked words are more likely to appear first when co-occurring within antonym sequences in Japanese. The results also point to the possibility that Japanese sequencing of antonym pairs was only marginally different from what is found in Western languages studied in this manner. The study found both the markedness and frequency are equivalently important for sequencing antonyms pairs in Japanese, but it focused only on/-i/ adjectives.

#### Acknowledgements

I would like to express my sincere gratitude to Dr. Isara Choosri, Assoc. Prof. Dr. Sophana Srichampa and Asst. Prof. Dr. Pattama Patpong, for their valuable suggestions and useful feedback on this research.

#### References

- Andersen, Henning. 2001. "Markedness and Theory of Linguistic Change". *Linguistics Theory* 219: 21–57.
- Backhouse, A.E. 1984. "Have All the Adjectives Gone?" *Lingua* 62, issue 3 (March): 169–186. https://doi.org/10.1016/0024-3841(84)90074-3.
- Cruse, D.A. 1976. "Three Classes of Antonyms in English". *Lingua* 38, issues 3–4: 281–292. https://doi.org/10.1016/0024-3841(76)90015-2.
- Givón, Talamy. 1970. "Notes on the Semantic Structure of English Adjectives". *Language* 46, no. 4 (December): 816–837.
- Greenberg, Joseph H. 1976. Language Universals. The Hague: Mouton.
- Gross, Derek, Fischer, Ute, and Miller, A. George. 1989. "The Organization of Adjectival Meanings". *Journal of memory and language* 28, issue 1 (February): 92–106, https://doi.org/10.1016/0749-596X(89)90030-2.
- Haspelmath, Martin. 2006. "Against markedness". *Journal of Linguistics* 42, no. 1 (March): 25–70.
- Hsu, Chan-Chia. 2015. "A Syntagmatic Analysis of Antonym Co-Occurrences in Chinese: Contrastive Constructions and Co-occurrence sequences". *Corpora* 10, no.1: 47–82, https://doi:10.3366/cor.2015.0066.
- Jones, Steven. 2002. Antonymy: A Corpus-Based Perspective. New York: Routledge.
- Jones, Steven, and Murphy, M. Lynne. 2005. "Using Corpora to Investigate Antonym Acquisition". *International Journal of Corpus linguistics* 10, issue 3 (September): 401–422, https://doi:10.1075/ijcl.10.3.06jon.
- Justeson, John S. and Katz, Slava M. 1991. "Co-occurrences of Antonymous Adjectives and Their Contexts". *Computational Linguistics* 17: 1–19.

- Kishimoto, Hideki, and Uehara, Satoshi. 2016. "Lexical Categories". In *Handbook of Japanese Lexicon and Word Formation*, edited by Kageyama. Taro and Kishimoto Hideki, 51–91. Berlin: De Gruyter Mouton.
- Kitahara, Yasuo and Tōgō, Yashio. 1989. *Hantaigo taishougojiten ('Dictionary of Antonyms and Contrasting Words')*. Tokyo: Tokyodo Shuppan.
- Kostić, Nataša. 2015. "Antonym Sequence in Written Discourse: A Corpus-Based Study". *Language Sciences* 47:18–31, https://doi:10.1016/j.langsci.2014.07.013.
- Kostić, Nataša. 2017. "The Distributional Asymmetries of English Antonyms in Language Use". *Brno Studies in English* 43, no.1 (January): 5–32, https://doi:10.5817 /BSE2017-1-1.
- Leech, Geoffrey. 1981. Semantics. New York: Penguin Books.
- Lehrer, Adrienne. 1985. "Markedness and Antonymy". *Journal of Linguistics* 21, no.2 (September): 397–429.
- Lehrer, Adrienne and Lehrer Keith. 1982. "Antonymy". *Linguistics and Philosophy* 5: 483–501.
- Lyons, John. 1977. *Linguistic Semantics: An Introduction*. Cambridge: Cambridge University Press.
- Martin, Samuel E. 2004. *A Reference Grammar of Japanese*. Honolulu, HI: University of Hawaii Press.
- Muehleisen, Victoria Lynn. 1997. *"Antonymy and Semantic Range in English"*. PhD diss., Northwestern University.
- Muehleisen, Victoria, and Maho Isono. 2009. "Antonymous Adjectives in Japanese Discourse". Journal of Pragmatics 41, issue 11 (November): 2185–2203, https://doi .org/10.1016/j.pragma.2008.09.037.
- Murphy, M. Lynne. 2003. *Semantic Relations and the Lexicon*. Cambridge: Cambridge University Press.
- Nishiyama, Kunio. 1999. The *Morphosyntax* and Morphonology of Japanese Predicates. PhD diss., Cornell University.
- Ohkado, Masayuki. 1991. "On the Status of Adjectival Nouns in Japanese". Lingua 83, issue 1 (January): 67–82, https://doi.org/10.1016/0024-3841(91)90052-7.
- Palmer, F.R.1976. Semantics a New Outline. Cambridge: Cambridge University Press.
- Paradis, Carita, Simone Löhndorf, Joost van de Weijer, and Caroline Willners. 2015. "Semantic Profiles of Antonymic Adjectives in Discourse". Linguistics 53, no.1: 153–191.
- Spencer, Andrew. 2008. "Negation in Japanese: A Case of Morphosyntactic Mismatch". Lingua 118, issue 7 (July): 997–1017, https://doi.org/10.1016/j.lingua.2007.04.004.

Willners, Caroline. 2001. Antonyms in Context. Lund: Lund University.