The ‘Fuzzy’ Boundary Between Two Types of Japanese Adjectives

by

Ana Kanza Tariq

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Arts

Department of East Asian Studies
University of Alberta

Abstract

The two kinds of Japanese adjectives, *i*-adjectives and *na*-adjectives, along with nouns, employ different forms (*-i*, *-na*, and *no*) to modify a noun. Based on such patterns, along with other grammatical characteristics identified in constructed examples, boundaries between lexical categories have traditionally been understood to be clear-cut.

However, Uehara (1998, 2003) has found a number of lexical items which inflect both as *na*-adjectives and nouns. In fact, Uehara finds that more than 70 percent of *na*-taking lexical items exhibit noun-like behaviours. Using prior research to support his claim (Rosch1978; Lakoff 1987; Taylor 1989), he has suggested that the boundaries between lexical categories might not be as clear-cut as previously conceived.

This thesis supports Uehara's proposal by highlighting a new set of data from internet discourse which suggests that the boundary between *i*-adjectives and *na*-adjectives might also be 'fuzzy', using examples in which what are traditionally considered *i*-adjectives are inflected as *na*-adjectives. It presents the results of two studies. The first looks at the factors of word length and frequency of use of lexical items and how they affect this 'improper' *i*-adjective conjugation in internet discourse. The second is a survey conducted to see how "natural sounding" native Japanese speakers consider this usage.

Acknowledgments

I would like to thank Professor Tsuyoshi Ono for being an amazing supervisor, mentor and teacher. Without his encouragement and guidance the past two years could not have been as valuable. I am also grateful to Professors Anne Commons and Xiaoting Li, whose proposed revisions helped me better my thesis and critique my own assumptions. To Saori Daiju, I am constantly grateful for the many favours I have received from her and have never returned. I would like to thank the members of University of Alberta East Asian Linguistics Research Group (2016 – 2018) for their advice and their patience when sitting through what was basically the same presentation with different titles multiple times. Thanks also to members of the East Asian studies department, be they students, administration or faculty, whose presence has enriched the past two years. I would like to thank my family and friends for their support and offers to feed me, and particularly my Grandma for being lovely.

Contents

List of Tables

List of Figures

List of Graphs

Abbreviations and symbols

COP	copula
GEN	genitive
INS	instrumental
LOC	locative
OBJ	object
PAST	past tense
POL	polite
PRE	present
PTCL	particle
SUB	subject
TOP	topic marker
(?)	not quite grammatical

Chapter 1 Introduction

1.1. Prototypicality

The traditional assumption, in linguistics, about lexical categories is that they are independent of each other. As Baker (2003: 1) says "The division of words into distinct categories or parts of speech is one of the oldest linguistic discoveries". This division is made possible by assigning characteristics to a particular group of words. Those within that set that do not possess all of these characteristics are considered exceptions, in this traditional view. Words that are members of one category are believed to possess the characteristics of that category alone, and no others. In addition, all members of a category are equally representative of it; all members of a category are considered homogeneous. Figure 1 gives a traditional view of lexical categories, as independent units, isolated from each other.

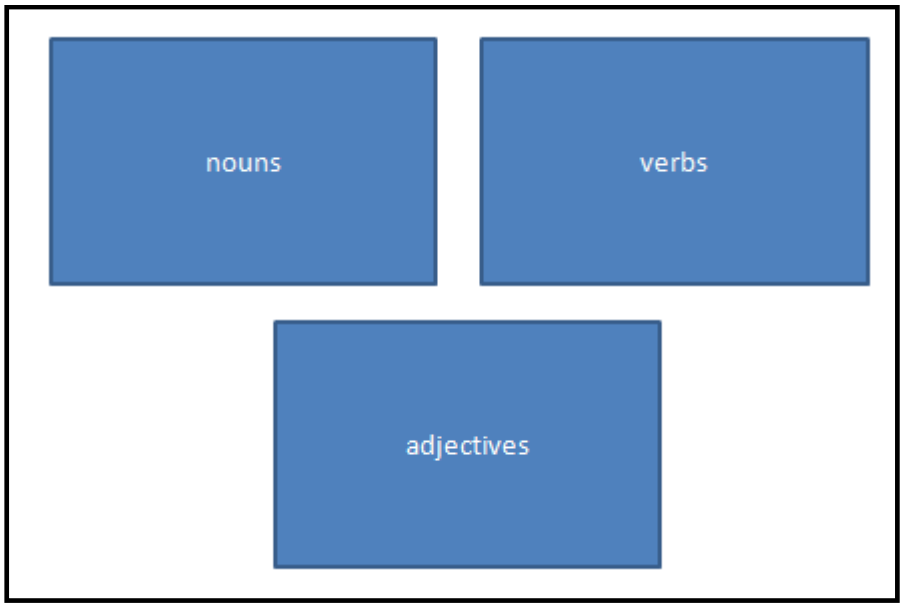


Figure 1: A traditional view of lexical categories

In contrast to this assumption of homogeneity is the concept of prototypicality

(Uehara1998, Rosch1978; Lakoff 1987; Taylor 1989). Uehara (1998) claims that within a particular lexical category, certain members are more characteristic of that category than others. Some words might possess all characteristics assigned to a particular category, others may possess only a few. Some words may possess characteristics of lexical categories they do not belong to, others might not. Instead of words existing in discrete packages under the labels of the different lexical categories, using prototypicality, we can view words as existing on a gradient, where a particular word may have characteristics of more than one category at the same time.

The concept of prototypicality is particularly meaningful in relation to language teaching. When dealing with beginners, the explanation that lexical categories are independent of each other and one rule applies to all members may be easier for students to understand. However, with advanced or intermediate level students, the concept of prototypicality would better explain why not all words display the same characteristics, instead of just writing some off as simply exceptions to a rule.

To illustrate the idea of prototypicality, in the context of the Japanese language, Uehara uses the example of nominal adjectives (often called *na* -adjectives). These are one of the two types of adjectives in Japanese, the other being *i*-adjectives. The names of these two categories reflect the form either adjective type takes in adnominal form. *i*-adjectives, when modifying nouns, do so directly, with a form that ends with *i*. For example, when modifying the noun *kuni* 'countries', the *i*-adjective *tanoshii* 'fun' forms the phrase *tanoshii kuni* 'fun countries'. *na*-adjectives, however, employ the particle *na* when modifying nouns. For example, the *na*-adjective *iroiro* 'various', when modifying the noun *kuni*, forms the phrase *iroiro na kuni* 'various countries'.

Nouns can also modify other nouns. However they require the genitive particle *no* in order to do so. For example when the noun *ajia* 'Asia' modifies the noun *kuni,* it forms the

phrase *ajia no kuni* ‘Asian countries’. Thus, *i*-adjectives, *na*-adjectives and nouns modify nouns in the following manner:

***i*-adjective**	tanoshi**i** kuni	‘fun countries’
***na*-adjective**	iroiro **na** kuni	‘various countries’
noun	ajia **no** kuni	‘Asian countries’

These lexical categories also employ different conjugational forms for their predicate use.

present tense polite	
***i*-adjectives**	tanoshii desu
***na*-adjectives**	genki desu

In present tense polite form, *i*-adjectives acting as predicates are conjugated by the addition of the present tense form of the copula *desu*. This is shown in the example above where the *i*-adjective *tanoshii* ‘fun’ when acting as the predicate of a sentence in past polite form is conjugated as *tanoshii desu* ‘(it) is fun’. *na*-adjectives, in present tense polite form, are also conjugated by the addition of the copula *desu*; however this copula directly follows the stem of the *na*-adjective. Thus, when the *na*-adjective *genki na* ‘healthy’ conjugates in present tense polite form it forms the phrase *genki desu* ‘(it) is healthy’.

These two types of adjectives are also conjugated differently in present tense polite negative form:

present tense polite negative	
***i*-adjectives**	tanoshikunai desu

***na*-adjectives** genki janai desu

i-adjectives are conjugated by substituting the final *i* for the suffix *-kunai* and then adding the present tense copula *desu*. For example, the *i*-adjective *tanoshii* takes the form *tanoshikunai desu* '(it) is not fun'. In the case of *na*-adjectives the particle *na* is replaced by the suffix *janai* and the copula *desu* is added. Thus, for the *na*-adjective *genkina* the present tense polite negative form is *genki janai desu* '(it) is not healthy'.

In past tense polite speech the two adjective types are conjugated in the following manner:

past tense polite

***i*-adjectives** tanoshi-katta desu

***na*-adjectives** genki deshita

For *i*-adjectives, the final *i* is replaced by the suffix *-katta* and the copula *desu* is added. For example, the *i*-adjective *tanoshii* when conjugated to its past tense polite form becomes *tanoshikatta desu* '(it) was fun'. However, in the case of *na*-adjectives, in past polite form, the copula *deshita* is attached to the stem of *na*-adjective. For example, the *na*-adjective *genki na* becomes *genki deshita* '(it) was healthy'.

To explain how the idea of prototypicality can help us reshape our understanding of lexical categories, Uehara (1995) calls attention to what he calls double construction cases. These are words that possess some characteristics of *na*-adjectives and some of nouns. In particular, these words, when modifying nouns, can take either the form *na* (a characteristic of *na*-adjectives) or use the genitive particle *no* (a characteristic of nouns). For example the word *iroiro* 'various', when modifying a noun (*kuni* 'countries' in the example), can take

either the particle *na* or the genitive particle *no* and lead to the same translation.

***na*-adjective**	**modified phrase**	**translation**
iroiro	iroiro **na** kuni	various countries
	iroiro **no** kuni	

1.2. 'Fuzzy' boundaries between of lexical categories

Uehara (1998) uses examples such as *iroiro* (mentioned in section 1.1) to argue against categories in language being discrete i.e., against the idea that lexical categories are independent from each other. Instead, he advocates for the boundaries between lexical categories being thought of as 'fuzzy', not clear-cut. He also uses the idea of prototypicality to deny the possibility that lexical categories are internally homogeneous i.e. he denies the idea that all members of one category will be equally demonstrative of all characteristics of said category. He quotes Teramura (1982) in saying "individual lexical items displaying ... ambiguous behaviour are located in a boundary region between categories" (Uehara1998: 104).

Uehara mentions Teramura's suggestion that "Japanese categories are situated along a continuum, whereby they are not strictly distinguished from each other, and individual lexical items displaying variable or ambiguous behaviour are located in a boundary region between categories" (Uehara 2003: 370). The behaviour of adjectives in Japanese is indicative of this claim.

As we saw above, Japanese has two kinds of adjectives, *i*-adjectives and *na*-adjectives. *i*-adjectives modify nouns directly, with the noun simply following the *i*-adjectives. This behaviour is similar to how verbs in Japanese modify nouns. The following is a comparison

between how the *i*-adjective *tanoshii* 'fun' and how the verb *wasurerareta* 'forgotten' modify the noun *kuni*.

***i*-adjective**	tanoshii kuni	'fun countries'
verb	wasurerareta kuni	'forgotten countries'

However, *na*-adjectives require the particle *na* and, sometimes, the genitive particle *no* when modifying a noun. This behaviour is similar to that of nouns when modifying a noun. The following is a comparison between how the *na*-adjective *iroiro* and how the noun *ajia* modify the noun *kuni*.

***na*-adjective**	iroiro **na** kuni	'various countries'
noun	ajia **no** kuni	'Asian countries'

The similarity in behaviour[1] between these categories is translated as *i*-adjectives' resembling verbs whereas *na*-adjectives resembling nouns by Backhouse (2004). According to this understanding, these lexical categories must exist in relation to each other in the manner illustrated in Figure 2.

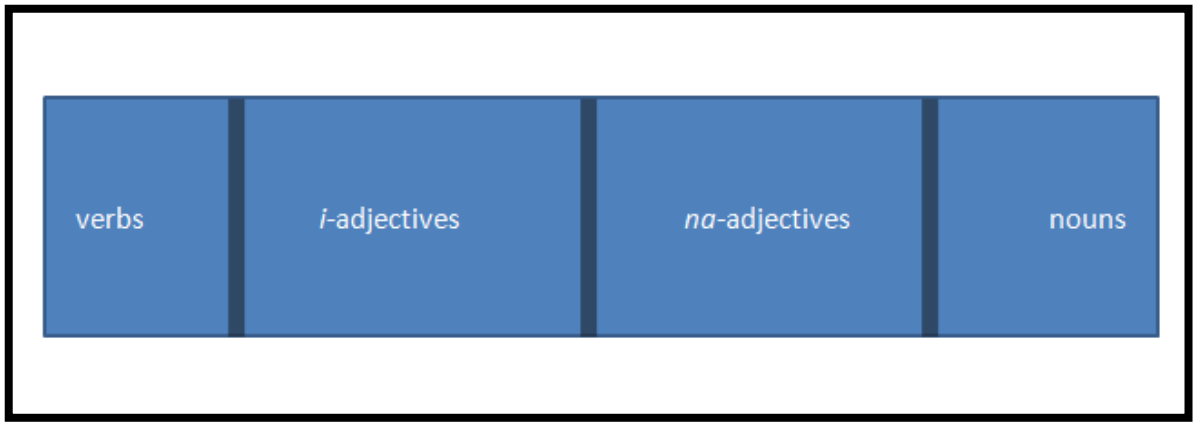


Figure 2: Categories in Japanese "are situated along a continuum" (Uehara 2003)

Figure 2 is a representation of Uehara's (2003) and Backhouse's (2004) understanding of the relationship between lexical categories. Verbs and *i*-adjectives, and nouns and *na*-adjectives display similar characteristics and so are presented bordering each other on the continuum, with 'fuzzy' boundaries between them. *i*-adjectives and *na*-adjectives show little distinction from each other except in morphology and so are presented as neighbouring categories.

The current work argues for the idea that *i*-adjectives and *na*-adjectives, like Uehara's double construction cases, also share a 'fuzzy' boundary, a region where words which possess characteristics of both *i*-and *na*-adjectives exist. Furthermore, an important purpose of this study is to ascertain whether certain properties of particular *i*-adjectives affect how often they are conjugated using the 'improper' conjugation.

1.3. Examples from internet discourse

As mentioned above, this thesis stems from the hypothesis that there is a 'fuzzy' boundary between *i*-adjectives and *na*-adjectives. Examples found in internet discourse seem to support this hypothesis. These examples were searched for using the *na*-adjective past polite conjugation[2,3], which requires the copula *deshita* to be added to the stem of the adjective when conjugating *i*-adjectives. The use of this particular conjugation was supported by a pilot study, which indicated that out of 10 other possible *na*-adjective conjugations, speakers are more likely to use the past polite conjugation when 'improperly' conjugating *i*-adjectives. This conjugation is henceforth referred to as either the *deshita* conjugation, as it requires the addition of the copula *deshita* when conjugating the *i*-adjective or the 'improper' conjugation. The term 'improper' was employed as a reminder that this usage, despite its presence in native speaker online posts, would traditionally be considered ungrammatical. The examples

gathered appeared on a variety of internet forums which differed in the degree to which the text would likely have been edited.[4]

The following example is from a forum where it is thought unlikely to have been edited. It was part of a question posted on a query website. The author, when making a medical inquiry, stated:

ketsueki	kensa	de	tetsuto	suuchi	ga	**takai**	**deshita**
blood	test	LOC	iron	numerical value	SUB	high	was

'In the blood test, the iron content was high.'

(From https://detail.chiebukuro.yahoo.co.jp/qa/question_detail/q13174355778)

Here the author, when conjugating the *i*-adjective *takai,* has used the *deshita* form as in the phrase *takai deshita* 'was high' (?). The 'proper' *i*-adjective conjugation would have been *takakatta desu* 'was high'.

improper			**proper**	
takai	deshita	instead of	takakatta	desu
high	COP:POL:PAST		high:PAST	COP:POL:PRE

In contrast with the previous example, the following example was found on a forum where it is thought more likely to have been reviewed before posting as it was found on a website of a technical nature, where the author was discussing features of different radios.

AM	kando	ga	kyoukutanni	**warui**	**deshita**
AM	sensitivity	SUB	extremely	bad	was

'The AM sensitivity was extremely bad.'

(From http://nice.kaze.com/av/st-g7-no2.html)

In this example, the author has used the phrase *warui deshita* ‘was bad’ (?) when conjugating the *i*-adjective *warui* ‘bad’ instead of the proper *i*-adjective conjugation which would have resulted in the phrase *warukatta desu* ‘was bad’.

improper			**proper**	
warui	deshita	instead of	warukatta	desu
bad	COP.POL:PAST		bad:PAST	COP:POL:PRE

The following is an example that was found on a relatively more formal forum, a newsletter for an NGO. In such a genre, the likelihood of the text being edited is higher than those in the previous two examples. When talking about a recent excursion a group had undertaken the author stated:

kakien	de	wa	tawawani	minotta	ookina	fuyugaki	o
Persimmon.Garden	LOC	TOP	abundantly	ripen	big	persimmon	OBJ

hasami	de	hitotsu zutsu	daijini	tori	**tanoshii deshita.**	
scissors	INS	one-by-one	carefully	pick	fun	was

‘Picking fully grown persimmons one at a time using scissors was fun.’

(From http://kimikagenomori.jp/)

Here the author, when conjugating the i-adjective *tanoshii* ‘fun’ has used the *deshita* conjugation forming the phrase *tanoshii deshita* ‘was fun’ (?) instead of the proper *i*-adjective conjugation *tanoshikatta desu* ‘was fun’.

improper		**proper**
tanoshii deshita fun COP.POL:PAST	instead of	tanoshikatta desu fun:PAST COP:POL:PRE

As the examples above show, this usage can be found in internet discourse even on sites where texts are likely to have been edited or checked by the author themselves or an editor. If this 'improper' use had been considered wholly 'unnatural' to Japanese grammar, it was likely to have been corrected or even removed. The fact that they remained in the text suggests that speakers did not find this usage absolutely 'improper'.

The next section explores some possible reasons why speakers might use this conjugation.

1.4. Possible reasons

Because the posts authored by native Japanese speakers and because of the large amount of such examples that were found and because they appeared on a variety of forums which differ in the degree of editing required, it becomes hard to judge all of them as simply mistakes.[5] What reasons might speakers have to use these 'improper' conjugations? Some possible reasons are listed below:

1. One obvious reason would be Uehara's proposal that the idea that as the boundaries between linguistic categories are not discrete; it is not surprising to find examples where the behaviour associated with one category influences the behaviour of some members of a nearby category.
2. The frequency of occurrence of the *deshita* conjugation might also have an impact on how acceptable the 'improper' use of it is to speakers. The *deshita* conjugation is much more frequent than the *i*-adjective past polite *-katta desu* conjugation. It is

used as a past polite conjugation both for *na*-adjectives, a much larger lexical category than *i*-adjectives due to its open class status (Uehara 1998), and for nouns, the largest of all lexical categories.

3. As mentioned in section 1.3, the present tense polite form of *i*-adjectives only requires the addition of the present tense copula *desu* to the *i*-adjective. For example the *i*-adjectives *tanoshii* 'fun' when acting as the predicate of a sentence in past polite form can be conjugated as *tanoshii desu* '(it) is fun'. It may seem natural to some speakers that switching from the present tense copula *desu* to the past tense copula *deshita* is a good enough indication of a change in tense.
4. It may arguably be easier to conjugate *i*-adjectives using the *na*-adjective conjugation. The 'proper' *i*-adjective conjugation requires an internal change, as the final *i* is replaced by the form *-katta*, followed by the addition of the copula *desu*.

adnominal form	**substituting the hiragana *i* with the form *-katta***	***i*-adjective in 'proper' past polite conjugation**
warui	warukatta	warukatta desu

However, conjugating them using the *na*-adjective conjugation simply requires the addition of the copula *deshita*.

adnominal form	***i*-adjective in 'improper' past polite conjugation**
warui	warui deshita

5. Lastly, speakers might find the form partially acceptable because some frequent *na*-adjectives have stems that end in a hiragana *i.* For example, the *na*-adjective *kirei* 'clean'.

adnominal form	**past polite form**
kirei na ‘clean’	kirei deshita ‘(it) was clean’

kirei is the 5th most frequent *na*-adjective in internet discourse (Matsushita 2011) and, thus, speakers are likely to often come across phrases like this. Because speakers may not be always aware of the lexical category of words some speakers might start conjugating *i*-adjectives in a similar manner out of habit or because they misconstrue them as *na*-adjectives with the stem final *i* at the end.[6]

1.5. Factors discussed

Some properties of particular *i*-adjectives which were considered likely to influence the usage of the ‘improper’ conjugation, were examined during a pilot project. Of these, the two that showed some correlation with the frequency of the ‘improper’ use were explored in depth in this study.

1.5.1. Word length

To find out the influence of word length on how often a particular *i*-adjective gets conjugated in the ‘improper’ conjugation, the number of mora in a word was used as a measure. Kubozono (1999) defines mora as a “unit of duration in Japanese” and says that they are “equivalent to a phonemic syllable”.

Because in Japanese, mora is a timing unit, the more mora a word has, the longer it is.

***i*-adjective romaji reading**[7]	**hiragana reading**[8]	**number of mora**

katai	か た い 1 2 3	3
urayamashii	う ら や ま し い 1 2 3 4 5 6	6

As *i*-adjectives with more moras might remain recognizable even when conjugated 'improperly', it was hypothesised that speakers might be more likely to conjugate longer *i*-adjectives 'improperly'.

1.5.2. Frequency of use

As is commonly known, the most frequent words in a language's lexicon are least susceptible to change. Examples of this can be found both in English and in Japanese. Usually verbs in English are conjugated in past tense by adding *ed*/*d*, for example the verb *thrive* conjugates to form *thrived* (O'Grady 2011). However, the frequent verb *go* has a past tense form *went*, which has resisted being changed to the usual paradigm. In Japanese the frequent verbs *kuru* and *suru* conjugate differently from other verbs. In the case of this study, it was thought that because frequent forms are heard more often and are a more established part of a speaker's lexicon, speakers would be less likely to accept replacements for these than for infrequent forms. Thus, it was hypothesised that speakers would find it less acceptable when very frequent *i*-adjectives use this 'improper' conjugation.

1.6. Overview of the study

Chapter 2 of this thesis discusses the internet study that was conducted to test whether the length of words and the frequency of particular *i*-adjectives in internet discourse affected whether *i*-adjectives were conjugated more often in this improper conjugation. Chapter 3

discusses a survey which comprised of asking participants to judge the degree of 'naturalness' of some constructed sentences involving 'improperly' conjugated *i*-adjectives. Chapter 4 summarises the results of chapters 2 and 3 and reassesses whether this thesis was able to support the findings of Uehara (2003).

Chapter 2 Internet Study

This chapter quantitatively examines internet discourse to see if the factors of word length and frequency affect the frequency of occurrence of *i*-adjectives being 'improperly' conjugated as *na*-adjectives. The methodology made use of posts by native Japanese speakers written on the Japanese version of the social media website Twitter (https://twitter.com/?lang=ja).

Section 2.1 discusses the selection of the forum used to collect data, the social media website Twitter and why some websites were disqualified for the purpose of this methodology. Sections 2.2 and 2.3 give details of the methodology and section 2.4 discusses the results received. Finally, Section 2.5 gives an overview of the findings of this chapter.

2.1. Selection of internet data and Twitter

The forum chosen for this study is Twitter[9], a social media website where users can write posts of up to 140 characters known as tweets.[10] Other users can then comment on those posts or 'retweet' them to share them.

As the use was first noticed in internet discourse, the most systematic approach seemed to be to use the same medium to study the phenomenon. The forum Twitter was deemed most appropriate for the methodology used. Some benefits of Twitter are listed below.

1. Like other search engines, the Twitter search engine can be used to find word strings, which may be interrupted by punctuation, spaces or paragraph changes. Unlike other search engines, as search results on Twitter load on the same page, the browser "Find" function can be efficiently used to highlight and count uninterrupted exact word strings.

2. On Twitter posts can be updated chronologically (from the most recent to those made less recently) and have a time stamp attached. This meant that if one needed to find all posts containing an exact string of keywords from a particular time period it would be relatively easy to do so.
3. Because users are required to make a profile when creating an account it is somewhat easier to find out whether or not authors of posts are native speakers of Japanese. In some cases, it may even be possible to ascertain which dialect of Japanese the author speaks.
4. Users can select a country which somewhat customizes which location their search results come from.[11] This was helpful and was used as one of the steps taken to restrict results to posts made by native speakers of Japanese, though, obviously, not all posts written within Japan are authored by natives.[12]

As indicated above, though the use of Twitter alleviated the problem of distinguishing native Japanese speakers from non-native speakers, it did not completely solve it. Some users use internet personas instead of actually identifying themselves[13] and do not declare their location, which made it difficult to identify their nativity. To overcome this problem their profiles were explored to check for any indications of them being non-native speakers. If none were found, the post authored by them was included in the study.[14]

To collect data for this study, a new Twitter account was created, using a newly created email account. No activity (such as “retweeting” or “posting” of comments or “liking”) was practiced on this account and so no inclination was shown towards any particular type of posts. Any preference towards a particular type of post would affect the search results received.

2.2. Methodology

One set of *i*-adjectives for this study was selected on the basis of the number of mora the *i*-adjective consisted of. The purpose of this set was to check whether the length of the word had any effect on the frequency of occurrence of the 'improper' usage in internet discourse. A second set was selected on the basis of frequency of use of the *i*-adjective in internet discourse and was used to check whether this factor would have an effect on the frequency of occurrence of the 'improper' usage in internet discourse.

Matsushita's *Vocabulary Database for Reading Japanese (VDRJ) Ver. 1.1* (2011)[15] was used to select *i*-adjectives for the studies conducted both in this and the next chapter and to gather information about the internet discourse frequencies of these *i*-adjectives. The version of the database mentioned contains the most common 60894 Japanese words listed according to frequency of use. The sources used for the purpose of compiling the database included printed materials[16] and the internet Q and A forum *Yahoo Chiebukuro*, otherwise known as *Yahoo! Answers* (https://chiebukuro.yahoo.co.jp/tag/tags.php?tag=Answers), a website where users post queries which are answered by other users. As data collection in the current study was limited to the website Twitter, the frequency list selected from the vocabulary database was that for which information had been gathered from the internet Q and A forum.

Searches were conducted using a Japanese writing system that uses a mix of kanji and hiragana forms of *i*-adjectives. For example, when conducting the search for the *i*-adjective *hageshii* 'intense', the form でした *deshita* (which is composed only of hiragana characters) was attached to 激しい *hageshii* (where 激 is a kanji character and all other characters are hiragana) resulting in the 'improper' phrase 激しいでした *hageshii deshita* (instead of the 'proper' phrase *hageshikatta desu*). Thus, the *i*-adjectives selected were only

those that are typically written using kanji. In addition, *i*-adjectives which were in the form of compound[17] were not selected for the current study. Such words contain a possible independent root inside them; in the case of *i*-adjectives this is often a noun. It was thought that the presence of a noun within the structure of the *i*-adjective may increase the chance of 'improperly' conjugating with *deshita,* as nouns also take the same form for past polite.[18]

2.3. Materials

2.3.1. Twitter usage

The steps taken to calculate the ratio between the 'improper' examples and the total examples (all examples containing either 'proper' or 'improper' past polite *i*-adjective conjugation) are listed below.

1. *i*-adjectives from either set (those selected to ascertain whether either word length or frequency of use in internet discourse was a factor) were conjugated using the 'improper' past polite conjugation and entered into the search bar on Twitter. For example, the 'improperly' conjugated *warui deshita*, containing the *i*-adjective *warui,* was entered into the search bar. The settings were set so the results appeared chronologically, with the most recent appearing at the top of the screen, using the "Latest" button. Though the posts shown contained both keywords *warui* and *deshita* consecutively as *warui deshita*, there were often interruptions between them in the form of spaces, paragraphs and a variety of punctuation marks. Such posts were irrelevant to the current study. Figure 3 shows some search results received when the keywords *hageshii deshita* 'was fierce' typed into the Twitter search engine as 激しいでした. Of these results, numbers 1, 2 and 4 are uninterrupted while numbers 3 and 5 are interrupted by punctuation.

Figure 3: Screenshot of Twitter results

2. Posts that did not contain any such interruptions between the consecutive keywords (e.g., *hageshii deshita* 激しいでした instead of *"hageshii" deshita*「激しい」でした) were identified (i.e., highlighted) by entering the keywords (e.g., *hageshii deshita* 激しいでした) into the Google Chrome Find function (CTRL + F)[19]. In Figure 4, result number 3,

containing the phrase *"hageshii" deshita*「激しい」でした 'was "fierce"', and result number 5, containing the phrase *hageshii. deshita* 激しい。でした are not highlighted, while lines that contained the uninterrupted phrase *hageshii deshita* 激しいでした 'was fierce' (numbers 1, 2, and 4), our target form, are highlighted.

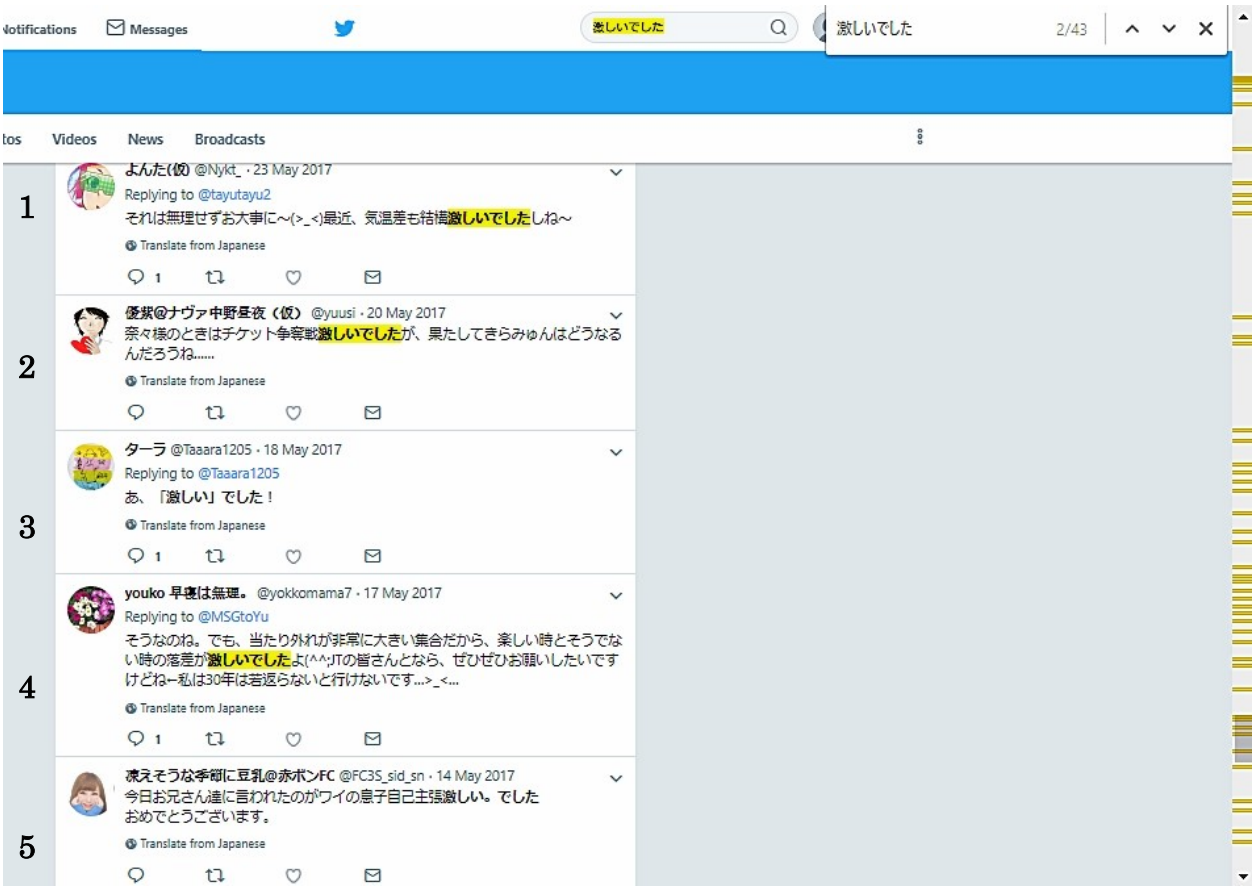


Figure 4: Screenshot of "Find" function usage on Twitter results

3. The first five most recent relevant posts written by native speakers were chosen and the time period between the most recent and fifth most recent posts was noted down.
4. *i*-adjectives were then conjugated using the 'proper' past polite conjugation and results containing the keywords were searched for. For example the *i*-adjective *hageshii* 'fierce' was conjugated to *hageshikatta desu* 'was fierce'. The settings were set to appear chronologically with the most recent post appearing first, by clicking the "Latest" button.

The Google Chrome Find function (CTRL + F) function was then used to highlight posts with no interruptions between the keywords. The 'proper' posts (posts involving *hageshikatta desu* 'was fierce') made during the period when the 5 "improper" posts (posts involving *hageshii deshita*) appeared, were counted.

5. A ratio was then calculated between the number posts containing 'improperly' conjugated past polite *i*-adjectives and the total ('improper' and 'proper') posts containing past polite *i*-adjectives.

2.3.2. Frequency of use in internet discourse

This portion of the study was conducted to assess whether the frequency of particular *i*-adjectives in internet discourse affects how frequently they are "improperly" conjugated. 3 *i*-adjectives each ranked around the 50th [20] most frequent or the 200th most frequent of all *i*-adjectives[21] in the internet portion of Matsushita (2011) were selected for the purpose of this part of the study. To eliminate the potential confounding factor of the word length of the *i*-adjective interfering with the results, all *i*-adjectives selected contained 4 mora. *i*-adjectives of this length were selected because 4 mora long *i*-adjectives were frequent enough for three adjectives of around the same frequency to be found within close proximity.

Table 1: *i*-adjectives selected to check whether frequency of occurrence in internet discourse is a factor

***i*-adjective**	**English**	**ranking among *i*-adjectives**
kanashii	sad	50
komakai	small	52
hageshii	violent, furious, tempestuous	53
tootoi	precious, valuable, priceless, noble, exalted, sacred	180

tsutanai	poor-quality, shoddy, crude	204
hakanai	fleeting, transient, short-lived, momentary, ephemeral, fickle, vain	218

Column 1 of the table gives the *i*-adjectives in romaji. Column 2 gives the meaning of the *i*-adjective in English. Column 3 gives the internet ranking among *i*-adjectives only.

2.3.3. Word length: number of mora

This portion of the study was conducted to assess the number of mora within an *i*-adjective affects how frequently they are "improperly" conjugated. 3 *i*-adjectives each of either 3 or 6 moras were selected for the purpose of this part of the study. To eliminate the potential confounding factor of frequency of use in internet discourse, *i*-adjectives were used which were relatively close to each other in the frequency rankings. The *i*-adjectives selected ranked between 123 – 177 in the ranking among *i*-adjectives. This was one of the few regions where at least 3 *i*-adjectives with 3 and 6 moras could be found.

Table 2: *i*-adjectives selected to check whether the number of mora a words is comprised of is a factor

***i*-adjective**	**English**	**number of mora**	**ranking among only *i*-adjectives**
arai	rough, rude, wild	3	123
nigai	bitter	3	132
nibui	dull, blunt	3	134
wazurawashii	troublesome, annoying, complicated	6	135
ichijirushii	striking, remarkable, considerable	6	147
magirawashii	confusing, misleading,	6	177

	equivocal, ambiguous, easily mixed up		

Column 1 of the table gives the *i*-adjectives in romaji. Column 2 gives the meanings of the *i*-adjectives in English. Column 3 gives the number of mora the *i*-adjective is composed of. Column 4 gives the internet frequency ranking among only *i*-adjectives, according to Matsushita (2011).

2.4. Results

2.4.1. Frequency of use in internet discourse

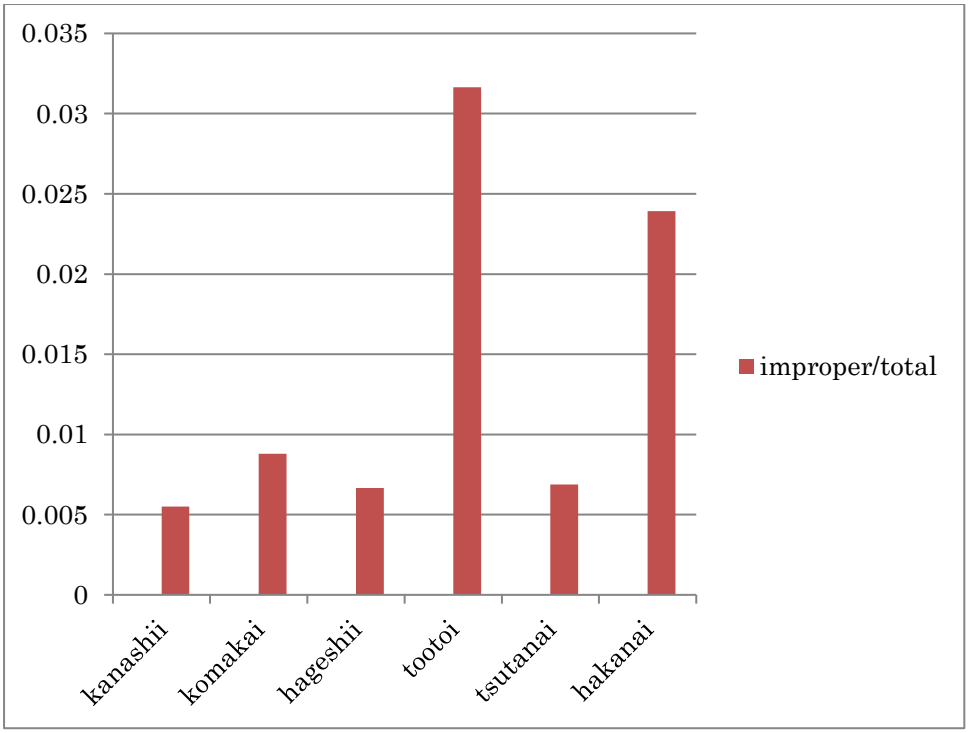
Table 3 shows results for *i*-adjectives used to ascertain whether the frequency of use in internet discourse influenced the frequency of occurrence of the 'improper' usage in internet discourse.

Table 3: Number of 'proper' and total examples for 'frequency of occurrence in internet discourse' data

***i*-adjective**	**frequency ranking (*i*-adjectives)**	**improper forms**	**proper forms**	**total (improper+proper)**	**improper/total**
kanashii	50	5	904	909	0.005501
komakai	52	5	564	569	0.008787
hageshii	53	5	746	751	0.006658
tootoi	180	5	153	158	0.031646
tsutanai	204	5	722	727	0.006878
hakanai	218	5	204	209	0.023923

The 6 *i*-adjectives used for this part of the study are listed in column 1. The internet frequency ranking of particular *i*-adjectives among *i*-adjectives only is given in column 2. Column 3 shows the most recent number of relevant 'improper' examples found. Column 4 shows the frequency of 'properly' conjugated examples posted within the same time period as when the examples in Column 3 were posted. Column 5 shows the total number of examples obtained by adding 'proper' and 'improper' examples. The ratio between the 'improper' and total examples is shown in column 6. A higher ratio meant that the likelihood of using the 'improper' form was higher.

As can be observed from the table, *i*-adjectives with a high frequency of use in internet discourse were, in general, less likely to be conjugated 'improperly' when compared to less frequent *i*-adjectives. For the more frequent *i*-adjectives (top 3 adjectives in the table), the ratios between the number of improper and total examples were 0.005, 0.008, and 0.006. In contrast, the ratios for the relatively less frequent *i*-adjectives (bottom 3 adjectives in the table), were 0.031, 0.006, and 0.023. The same information is graphically presented in Graph 1, which illustrates the correlation between the frequency of use in internet discourse and ratio between the number of improper examples over total examples.



Graph 1: Graphical representation of Table 3

In Graph 1, the x axis shows *i*-adjectives selected for the 'frequency of use' set, with the more frequent 3 on the left and the less frequent 3 on the right. The y axis shows the ratio between the amount of 'improper' and the amount of total examples with the higher bar indicating less of a difference between the number of 'improper' and number of 'proper' examples found over the same period of time.

Though the results mostly agree with the hypothesis, an exception from the expected result is the *i*-adjective *tsutanai* 'poor-quality, shoddy, crude' which is relatively infrequent (ranked 204 among *i*-adjectives) but has a ratio significantly smaller than other low frequency *i*-adjectives, *tootoi* 'precious, valuable, priceless, noble, exalted, sacred' and *hakanai* 'fleeting, transient, short-lived, momentary, ephemeral, fickle, vain'.

This discrepancy in the results may have been caused by the fact Japanese has three different kinds of writing systems, of which only one was tested in the current study. Though, the words chosen for the study were thought to be used mostly with a system where kanji and

hiragana are used in conjunction[22], it is possible that in internet discourse, for the word *tsutanai* 'crude' other writing systems (such as hiragana or katakana only) were more frequently used by native Japanese speakers, which may have played a role in the ratio obtained for *tsutanai*.

We cannot claim without a doubt that frequency of use in internet discourse is a factor, but we see that the likelihood of 'improper' conjugation increases when considering *i*-adjectives that are less frequent.

2.4.2. Word length: number of mora

Table 4 shows *i*-adjectives used to ascertain whether word length, measured by the number of mora, influenced the frequency of occurrence of the 'improper' usage in internet discourse.

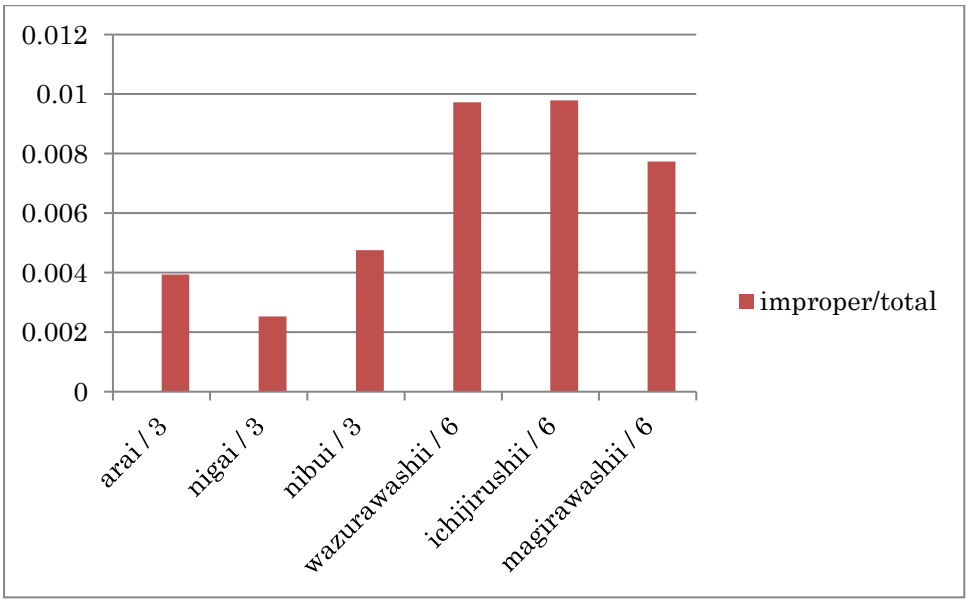
Table 4: Number of 'proper' and total examples for 'number of mora' factor

***i*-adjective**	**number of mora**	**ranking among only *i*-adjectives**	**improper examples**	**proper examples**	**total examples**	**improper/ total**
aria	3	123	5	1266	1271	0.003934
nigai	3	132	5	1972	1977	0.002529
nibui	3	134	5	1046	1051	0.004757
wazurawashii	6	135	5	509	514	0.009728
ichijirushii	6	147	5	506	511	0.009785
magirawashii	6	177	5	641	646	0.00774

The 6 *i*-adjectives used for this part of the study are listed in Column 1, on the left. Column 2 gives the number of mora the *i*-adjective is composed of. The internet frequency ranking of particular *i*-adjectives compared to other *i*-adjectives is given in the next column. Columns 4

and 5 show the amount of examples containing 'improperly' and 'properly' conjugated *i*-adjectives posted within the same time period respectively. Column 6 shows the total amount of examples ('proper' + 'improper') from within the aforementioned time period. Column 7 shows the ratio between the 'improper' and total examples. As mentioned before, higher ratios mean that the likelihood of using the improper form was higher.

Contrasting between the 2 sets of *i*-adjectives, those that are 3 and those that are 6 moras in length, there appears to be some correlation between the number of mora an *i*-adjective is composed of and the frequency of occurrence of 'improperly' conjugated *i*-adjectives. *i*-adjectives with 3 moras (*arai*, *nigai* and *nibui*) had a lower ratio between the amount of 'improper' cases and the total number of cases found (0.003934, 0.002529 and 0.004757 respectively) when compared with *i*-adjectives that had 6 moras (*wazurawashii, ichijirushii,* and *magirawashii* with ratios of 0.00972, 0.009785 and 0.00774 respectively), which indicates that longer words are more likely to be conjugated 'improperly'. The same information is presented graphically below.



Graph 2: Graphical representation of Table 4

In Graph 2, the x axis gives the *i*-adjectives selected for 'number of mora' set, with those with less mora on the left and those with more mora on the right. The y axis gives the ratio between the amount of 'improper' and the amount of total examples, with the higher bar indicating less of a difference between the number of 'improper' and number of 'proper' examples found over the same period of time.

i-adjectives composed of 6 mora had more 'improper' examples per the total number of examples, when compared with *i*-adjectives composed of 3 mora. Though, because of the limited amount of data used and produced, we cannot claim without a doubt that the number of mora an *i*-adjective is composed of is a factor, we see that the likelihood of 'improper' conjugation increases when considering *i*-adjectives with a larger number of mora.

2.5. Summary

This chapter studied two factors that may promote 'improper' conjugation of *i*-adjectives, using data from internet discourse. The purpose of this chapter was to ascertain whether the frequency of use and word length measured according to the number of mora the *i*-adjective was composed of affected the number of 'improper' examples relative to the 'proper' examples, from within a given time period. Some correlation was seen for both factors tested i.e., less frequent and long words seem to promote the use of 'improper' forms. This chapter presented a quantitative analysis of actual use by examining a relative frequency of occurrence between 'improperly' and 'properly' conjugated *i*-adjectives in internet discourse. The next chapter examines another dimension by looking at speaker judgment regarding the 'naturalness' of the 'improper' conjugation.

Chapter 3 Survey on Naturalness

The last chapter discussed a study that was conducted to see whether certain characteristics of particular *i*-adjectives affected how often they were 'improperly' conjugated in internet discourse. The factors considered were the word length, measured using the number of mora the *i*-adjective was composed of, and the frequency of use of the *i*-adjective itself, in internet discourse. The results indicated that relatively infrequent *i*-adjectives composed of a larger number of mora have a higher likelihood of being conjugated 'improperly'.

However, despite the presence of examples such as these in internet discourse, the possible correlation of the two factors discussed in the previous chapter with frequency of usage of the 'improper' conjugation and the aforementioned possible reasons speakers might have to adopt the usage (Chapter 1, Section 1.4), it is still not known whether those that read these posts judge the usage to be 'natural' to Japanese grammar; the usage could still be interpreted as a 'mistake'. Indeed, the examples found surprised most native Japanese linguists and educators who were first exposed to the data which was collected during the internet study (Chapter 2), which seemed to indicate that further support was required that such a use could be accepted by regular speakers (i.e., not engaged in Japanese language teaching), not just an error. Thus, as a method of understanding how 'natural' readers of such posts find this usage, a survey was conducted, asking Japanese speakers to judge the 'naturalness' of examples involving 'improperly' conjugated *i*-adjectives. The main purpose of the survey was to judge whether native speakers of Japanese found the use of the 'improper' conjugation with *i*-adjectives 'natural' when it was presented to them in example sentences.

3.1. Selection of *i*-adjectives

10 *i*-adjectives were selected. To eliminate bias caused by the factor of frequency of use of the *i*-adjectives in internet discourse, all *i*-adjectives selected were ranked between the most frequent and the 20th most frequent based on Matsushita (2011); they were all relatively similar in frequency. To eliminate bias due to the factor of word length, they were also all 4 moras long. The *i*-adjectives selected are given in Table 5.

Table 5: *i*-adjectives selected for survey

***i*-adjectives romaji**	**English**	**frequency ranking**
chiisai	small	3
oishii	tasty	4
tanoshii	fun	5
tawaii	cute	6
mijikai	short	7
akarui	bright	8
tsumetai	cold	9
suzushii	cool	14
yasashii	gentle	15
ureshii	happy	16

Column 1 shows the *i*-adjectives selected for the survey. Column 2 shows the meaning of the *i*-adjectives in English and Column 3 gives the ranking of the *i*-adjectives according to their frequency of use as compared to other *i*-adjectives.

3.2. Survey outline

The survey contained 15 sentences, 10 of which were relevant to the study. These sentences, containing 'improperly' conjugated *i*-adjectives, were constructed by taking example sentences from the online dictionary *Takoboto* (http://takoboto.jp/) and then replacing the

'proper' endings with an 'improper' one. The five dummy sentences were constructed for the purpose of the survey. They were included in order to eliminate response bias from the participants, as it was thought that if they became aware of the purpose of the survey it may affect the responses they gave. These dummy questions contained correct or incorrect particle usage (some being grammatical and some ungrammatical) and were irrelevant to the actual subject of the survey.

The order of the sentences was randomised, using the website "Random.org" (https://www.random.org/lists/) and, using this order, the survey was constructed on "Google Forms" (https://docs.google.com/forms). Copies of this survey were then printed out and the participants were given pens to complete this survey to make sure that the first answers could be kept track of, in case the participants changed their answers retrospectively (because the pens were un-erasable the participants had to draw a line through a previous answer if they wanted to change it). The first answers were the ones collected in case that the participants happened to have changed their answers[23].

All instructions and most questions[24] were given to participants only in Japanese. Participants were instructed to judge whether each sentence in the survey was *shizen* 'natural' sounding, *yaya shizen* 'somewhat natural' sounding or *fushizen* 'unnatural' sounding. A translation of the written instructions given to the participants is given below:[25]

> How natural sounding are the following phrases? Please click "natural sounding" for phrases that sound natural, "somewhat natural sounding" for phrases that sound somewhat natural and "unnatural sounding" for phrases that sound unnatural.

Figure 5 shows how a sentence (given in romaji rather than in the original Japanese) would have appeared on the survey.

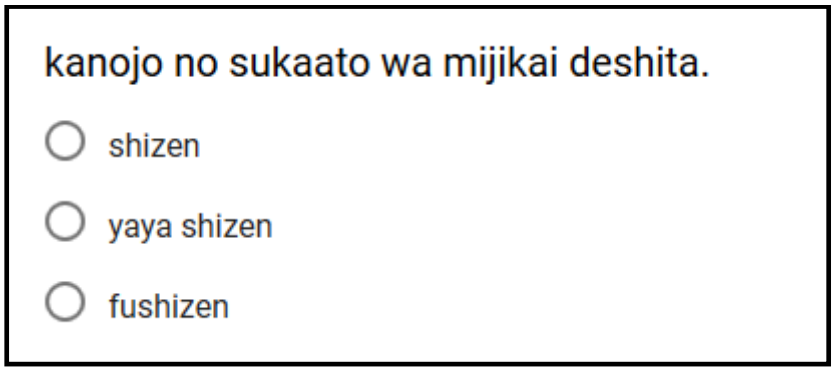


Figure 5: A sentence from the survey (in romaji)

An English translation of this sentence would be 'Her skirt was short'. The *i*-adjective *mijikai* 'short' has been conjugated in the 'improper' past polite conjugation, by adding the copula *deshita* (associated with conjugated *na*-adjectives and nouns). The participants of the survey had to judge how 'natural' the sentence sounded by picking one of the three options listed, *shizen* 'natural', *yaya shizen* 'somewhat natural' or *fushizen* 'unnatural'. The original survey questionnaire can be found in the appendix.

3.3. Participants

Participants were all students at the University of Alberta at the time of the survey. They were all native speakers of Japanese, of approximate ages between 18 and 22. There were 7 participants originating from Kansai, Chiba, Fukui, Shizuoka, Nagano and Osaka of Japan. Surveys were filled by members of the University of Alberta's Japanese Conversation Club meeting, where English and Japanese speakers gather on a weekly basis to converse with each other and improve their oral speaking skills, and were also filled by TAs of a 300 level Japanese language course.[26]

3.4. Results

Table 6 gives the response percentages for each sentence.

Table 6: Percentage of response type

sentence	English	natural	somewhat natural	unnatural	total
kanojo no sukaato wa mijikai deshita	Her skirt was short	0	57.1	42.9	100
kodomo ga kawaii deshita	The child was cute	0	57.1	42.9	100
hontoo ni oishii deshita	It was really delicious	0	57.1	42.9	100
sono tsukue wa Megu ni wa chiisai deshita	The desk was small for Meg	0	42.9	57.1	100
te ga tsumetai deshita	My hands were cold	0	42.9	57.1	100
kare no kotoba yasashii deshita	His words were kind	0	42.9	57.1	100
shiken ni gookaku shite ureshii deshita	I was pleased that I passed the exam	0	57.1	42.9	100
anata to isshoni ite tanoshii deshita	It was fun to be with you	0	57.1	42.9	100
kokage wa suzushii deshita	The shade of the trees was cool	0	42.9	57.1	100
yuube wa tsuki ga akarui deshita	The moon was bright in the evening	0	42.9	57.1	100

Column 1, on the far left, shows the sentence used in the survey (given in romaji, instead of the original Japanese). Column 2 gives the English translation of the sentence. The following 3 columns show the percentage of participants who judged the sentence to sound 'natural', 'somewhat natural' and 'unnatural' respectively. The last column gives the total percentage. Though none of these sentences containing an 'improperly' conjugated *i*-adjective were judged as being fully 'natural' by the participants, none were considered unanimously 'unnatural' either. There seemed to be some degree of 'naturalness' associated with the use as all the sentences were rated as 'somewhat natural sounding' by approximately half of the participants (3 or 4). This result contrasts with what is considered a rigid rule in traditional Japanese grammar (Kuno 1973; Martin 2004; Shibatani 1990) and indicates that there is actually flexibility and/or variation in the interpretation the behaviours of the lexical categories (a finding which overlaps with the results from Chapter 2).

Table 7 shows responses for individual participants for each of the sentences.

Table 7: Individual responses

			Participant						
#	**sentence**	**English**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	kanojo no sukaato wa mijikai deshita	Her skirt was short	U S	S N S	S N S	U S	U S	S N S	U S
2	kodomo ga kawaii deshita	The child was cute	U S	S N S	S N S	U S	S N S	S N S	U S
3	hontoo ni oishii deshita	It was really delicious	S N S	S N S	S N S	U S	U S	S N S	U S
4	sono tsukue wa Megu	The desk was small for	U	S	S	S	U	S	U

	ni wa chiisai deshita	Meg	S	NS	NS	NS	S	NS	S
5	te ga tsumetai deshita	My hands were cold	US	SNS	SNS	US	US	SNS	US
6	kare no kotoba yasashii deshita	His words were kind	SNS	SNS	SNS	US	US	SNS	US
7	shiken ni gookaku shite ureshii deshita	I was pleased that I passed the exam	US	SNS	SNS	SNS	US	SNS	US
8	anata to isshoni ite tanoshii deshita	It was fun to be with you	US	SNS	SNS	SNS	US	SNS	US
9	kokage wa suzushii deshita	The shade of the trees was cool	US	SNS	SNS	US	US	SNS	US
10	yuube wa tsuki ga akarui deshita	The moon was bright in the evening	US	SNS	SNS	US	US	SNS	US

Column 2 shows the sentence, as used in the survey (given in romaji, instead of the original Japanese). Column 3 shows the meaning of the sentence in English. The following 7 columns correspond to the responses given by the 7 participants who completed the survey. The answers have been coded, with ‘natural sounding’ coded NS, ‘somewhat natural sounding’ coded SNS and ‘unnatural sounding’ coded US.

Though none of the sentences were judged as ‘natural’, for each sentence, approximately half the participants (3 or 4 out of 7) felt that each was at least ‘somewhat natural’. Some participants were consistent with their judgments. Participants 2, 3 and 6 judged all the non-dummy sentences to be ‘somewhat natural sounding’, while Participant 7 judged all the non-dummy sentences as ‘unnatural’. However, as can be seen from the table,

some participants who regarded one particular adjective as 'somewhat natural sounding' when in the 'improper' conjugation, judged another to be 'unnatural'. Participants 1, 4 and 5 showed variability in their assessments of whether *i*-adjectives sounded natural when conjugated using the 'improper' conjugation. For example, Participant 4 reported that sentences 1, 2, 3, 9 and 10 were 'unnatural sounding' but sentences 4, 6, 7 and 8 were 'somewhat natural sounding'. This variation may indicate that there are other factors (than the number of word length and the frequency of use in internet discourse) that may affect speaker judgment when it comes to whether an 'improperly' conjugated *i*-adjective sounds "natural".[27]

The results suggest that, though there may be variation in speaker judgment, some speakers find the 'improper' usage 'natural' to Japanese grammar, to some degree. This result agrees with those received in Chapter 2, where Japanese native speakers were found using the 'improper' conjugation in internet discourse.

3.5. Summary

This survey was meant to be a more direct tool to assess whether native speakers find this usage, which would be considered ungrammatical according to traditional grammar, at all 'natural' sounding. The results indicated that many of the speakers associated a degree of 'naturalness' with the 'improper' conjugation. For all these questions participants were divided between the responses sounding 'somewhat natural' or 'unnatural', with approximately half the participants on either side. This indicates that though, in traditional grammar (Kuno 1973; Martin 2004; Shibatani 1990), rules are assumed to be fixed and inflexible; there is variability in native speaker judgment with respect to these rules. However, no concrete inferences can be made on the basis of a 7 participant survey, as the amount of data generated was too small.

In the next chapter, we go over the results of both studies (from Chapters 2 and 3), discuss how these relate to Uehara's hypothesis (2011) of the prototypicality of the lexical categories and, also, how this research can be furthered to get more concrete results.

Chapter 4 Conclusion

The traditional assumption about lexical categories is that they are independent of each other (Baker 2003). Uehara (2003) used the example of non-prototypical Japanese *na*-adjectives to counter this assumption (Uehara1998; Rosch1978; Lakoff 1987; Taylor 1989). His work discussed non-prototypical *na*-adjectives which display characteristics of the neighbouring lexical category of nouns. The purpose of this thesis was to see if whether Uehara's idea of 'fuzzy' boundaries between the lexical categories of *na*-adjectives and nouns is also supported for another boundary, that between *na*-adjectives and *i*-adjectives. Focusing on the conjugation of *i*-adjectives as *na*-adjectives in past polite form, two separate studies were conducted. The first, presented in chapter 2, was an internet study conducted on the social media website Twitter to see whether certain inherent factors meant that some *i*-adjectives were more likely to acquire characteristics of *na*-adjectives than others. In particular, it looked at whether the factors of word length and frequency in internet discourse had an impact on how likely authors of posts were to conjugate certain *i*-adjectives in this 'improper' manner. The second was a survey conducted with native Japanese speakers as participants. The purpose of the survey was to see how 'natural/unnatural' speakers find the use of the 'improper' conjugation.

4.1. Word length

Under the assumption that if the longer word that was 'improperly' conjugated it would be easier to recognise when compared with an 'improperly' conjugated shorter word, it was hypothesised that speakers would be more likely to conjugate a longer *i*-adjective using the 'improper' conjugation. A ratio between the 'improperly' conjugated examples and total past polite examples over a certain period of time was calculated. The larger the ratio was, the

more likely it was that the *i*-adjective would be conjugated improperly. We found that the longer words tended to be 'improperly' conjugated more often than shorter words. These results agreed the initial hypothesis.

4.2. Frequency

As previously stated in Chapter 1, frequently used *i*-adjectives are said to resist grammatical change more than words that are relatively less frequent (O'Grady 2011). Thus the hypothesis was that words that are relatively infrequent would be more likely to get 'improperly' conjugated. A ratio between the 'improperly' conjugated examples and total past polite examples over a certain period of time was calculated. The larger the ratio was, the more likely it was that the *i*-adjective would be conjugated 'improperly'. The results indicated that relatively infrequent *i*-adjectives tended to be 'improperly' conjugated more often than very frequent *i*-adjectives. These results agreed with the initial hypothesis.

4.3. Survey

The survey included sentences containing the 'improperly' conjugated *i*-adjectives. The survey was a method of checking whether speakers found the use 'natural' at all. As participants did not respond with a unanimous 'unnatural' when given a sentence with an 'improperly' (i.e., ungrammatically) conjugated *i*-adjective, this may be an indication that the 'rule' that conjugates *i*-adjective conjugation isn't wholly subscribed to by native speakers and, thus, isn't as well-defined in speaker's minds as previously thought. This lack of clear cut definitions with regard to grammatical rules in speakers' minds is comparable to Uehara's findings on the 'fuzziness' of the boundary between lexical categories, suggesting that, like lexical categories, people's intuition is not discrete.

The idea that speakers' judgments about grammatical rules, and, thus, grammatical rules themselves, are dynamic and not well defined within the minds of native speakers is in accordance with the findings of a study by Vance (1991), where he asked native Japanese participants to conjugate some made-up verbs as they saw appropriate. Vance found that participants had difficulty conjugating the forms properly using the rules which they had been assumed to have, and, thus, suggested a possibility that speakers might not have grammatical rules; they just have already-memorised individual forms. The results of the current study suggest that newly encountered 'improper' forms are sometimes judged as 'natural' by individual speakers, again questioning the rigid characterisation of grammar traditionally used to describe language.

4.4. Other interpretations and possible improvements

Some other possible interpretations of the result and possible improvements to this project are discussed in this section.

4.4.1. Closed versus open word classes[28]

An interpretation of why speakers might treat certain *i*-adjectives as they would *na*-adjectives when conjugating them might be the closed status of *i*-adjectives as a lexical category. As Uehara observes, "the *na*-adjective category is an open class, unlike the (*i*-) Adjective category" (2003: 378). *i*-adjectives "show strong phonological restrictions", while *na*-adjectives are not bound by similar "phonological restrictions", and, thus, adjectives borrowed from foreign languages are mostly added to the *na*-adjective category (Backhouse, 2004: 66). If speakers encounter an adjective they have not heard of previously, they may choose to conjugate it as a *na*-adjective simply as a default option, because it is the *na*-adjective category where most new adjectives go. This may also explain why our results

show relatively less frequent *i*-adjectives getting conjugated 'improperly' more, because these are *i*-adjectives that speakers may not have encountered as often. This interpretation, however, fails to explain why even in the case of the most frequent *i*-adjectives such as *takai* 'high' and *tanoshii* 'fun' (ranked 4th and 23rd among *i*-adjectives in internet discourse), utterances with 'improper' conjugation are sometimes found.

4.4.2. Phonological/morphological patterning

Whether the phonological/morphological patterning of the word influence speaker judgment about how 'natural' an *i*-adjective sounds when conjugated 'improperly' is itself also an interesting issue.[29] Backhouse (2004) summarises the 'composition of adjective types' in the following manner: *i*-adjectives are "semantically central, non-derived native members…generally gradable and occur in all major semantic types" (ibid.: 70) while *na*-adjectives are "gradable and occur in most semantic types, but…include many lexically complex members" (ibid.: 70). A notable distinction Backhouse mentions is one of "phonological restrictions". *na*-adjectives have fewer phonological restrictions than *i*-adjectives, which is an important distinction as it is one of the reasons foreign words enter the Japanese lexicon as *na*-adjectives instead of *i*-adjectives. This study was limited by the researcher's limited proficiency of the Japanese language and knowledge of the etymology of Japanese words. However, with a larger corpus of words, it would be interesting to see whether any conclusive evidence of *i*-adjectives with phonological characteristics more similar to those of *na*-adjectives could be found to sound more 'natural sounding' when conjugated using the 'improper' conjugation.

4.4.3. 'Improper' conjugation in audio data

Having looked at the use through the written medium of social media posts, it would be interesting to study the usage through the use of audio data. Some social media posts appear to have been written in a manner similar to spoken discourse, thus, in some ways a change in the grammars of internet discourse may indicate a change in the manner in which people speak. With particular regard to the Japanese language, evidence of this is the frequent dropping of particles[30] and the use of sentence ending particles such as *yo* and *ne*[31] in Tweets on Twitter and statuses on Facebook.

4.5. Limitations

This section addresses the possible weaknesses of this study.

4.5.1. Larger corpus with finer instrument

Both the studies discussed in this paper generated a very limited amount of data, which made it difficult to discuss variation found in the in results in detail. For example, some participants of the survey (Chapter 3) preferred the use of the 'improper' conjugation with certain *i*-adjectives over others, which may be an indication that there are other factors at play. This interpretation of the results is strengthened by the fact that the two factors believed to affect the 'improper' usage were kept constant during the survey. It was thought possible that a pattern in the results may indicate the presence of some unnoticed factor. However, possibly because of the small number of sentences in the survey and the small pool of participants, no such pattern was noticed.

Thus, though the effect of these other factors (including the ones mentioned in sections 4.4.1 and 4.4.2) has not been demonstrated in this thesis, this flaw could be remedied with a larger corpus, a finer instrument and a larger pool of participants. Factors, such as those

mentioned in sections 4.4.1 and 4.4.2, cannot be dismissed as potential reasons for this use because not all speakers might adopt the usage for the same reasons and multiple factors might play a role.

4.6. Non-prototypical members of lexical categories

The current study stemmed from Uehara's finding (2003) that *na*-adjectives are a good example of a lexical category where some members display prototypical behaviour. Other members of the category can be said to display non-prototypical behaviour. This study aimed to show that certain *i*-adjectives also display non-prototypicality, using the example in internet discourse where *i*-adjectives are conjugated, in past polite form, as if they were *na*-adjectives. Such *i*-adjectives can be said to exist in the 'fuzzy' region between the two lexical categories of *i*- and *na*-adjectives. 0The study also showed that the gradience of this non-prototypical behaviour can be explained by certain factors of particular *i*-adjectives (such as the word length and the frequency of use) and that longer and relatively less frequent *i*-adjectives are more susceptible to the usage. The results of the study indicated that the two factors may affect the use of the 'improper' conjugation more for certain *i*-adjectives than for others. A short survey was also conducted that supported the finding that speakers found the use 'somewhat natural' to varying degrees for different *i*-adjectives. Thus, though in traditional grammar the boundaries between lexical categories may be clear cut and the members of said categories may be homogenous in behaviour (Baker 2003), this does not seem to be the case either in the context of internet discourse or for the judgment of individual speakers.

[1] *i*-adjectives "occur as the head of intransitive predicates, directly as modifier of nouns in NPs, and as copula complement …of verbs such as *naru*" (Backhouse 2004: 50) while *na*-adjectives "occur as the complement of the copula *da* ; as copula complement with the marker *ni* of *naru* etc.; and as modifier of nouns in NPs followed by the adnominal marker *na* …, *no* …and either *na* or *no*" (Backhouse 2004:50).

[2] The use of the *deshita* conjugation to exemplify the phenomenon was initially arbitrary. However, sufficient reason has since been found to continue using this particular conjugation. A pilot project showed that out of ten *na*-adjective forms (from within the adjective types plain style non past, polite style negative, polite style past, plain style negative past, polite style negative past and the *-te* form conjunctive) the most likely to be used 'improperly' with *i*-adjectives is the *deshita* conjugation. Furthermore, in her book *Japanese: A Linguistic Introduction*, Hasegawa (2014) mentions that, due to how relatively recent the past polite *i*-adjective conjugation is, speakers might still feel uncomfortable when using it. This may mean the past polite conjugation is most likely to have variants.

[3] This conjugation in past polite form is shared between *na*-adjectives and nouns.

[4] "Edited texts", for the purpose of this thesis, are from forums where the texts containing the 'improper' usage might have been edited by someone other than the author or reviewed by the authors themselves before they were posted. These are understood to be texts where the chances of the author simply having used this particular conjugation as a mistake are low. The assumption here is that, on such a forum, the author and/or the editor find the use 'proper' or 'natural to Japanese grammar' to some degree. In contrast, an unedited text would be one where the author probably did not review the comment or did not need to have it edited by a second person. Thus, the possibility of the author having simply made a mistake when

writing the post is relatively higher. The estimated level of editing is a matter of gradience and is subjective.

[5] Even if they are genuine mistakes, it is interesting that they occur in this particular form, which is a borrowing from a nearby category. This again supports Uehara's (2003) claim that lexical categories exist on a continuum.

[6] Another reason might be the instability of the 'proper' (*-katta desu*) conjugation, which was standardized during the Meiji period (1868 – 1912) (Hasegawa 2014). Because the conjugation is relatively recent, "speakers still feel uncomfortable" when using it (Hasegawa 2014).

[7] Romaji is Japanese written using English script.

[8] In Japanese, there are 3 kinds of writing systems; the same word can be written using 3 different kinds of characters and still said to be written in Japanese. Using the example of the *i*-adjective *arai* 'rough, rude, wild', we can introduce these writing systems as katakana アライ, hiragana あらい and kanji 荒い.

[9] Before the current methodology was selected, pilot studies were conducted using Google and Yahoo search engines. A few of the issues encountered while collecting data using these forums are listed below:

1. Double quotes can be used in either search engine to extract 'exact strings', results where all keywords appear in the same order as they do within the double quotes. However, neither search engine took account of punctuation, spaces or paragraph changes between the words within exact strings. This meant that it was problematic to search for phrases where authors to conjugated *i*-adjectives

'improperly'. Interrupted phrases, which were irrelevant to the study, crowded the search results. An example where the *i*-adjective and the copula are uninterrupted is shown below.

高校野球最高！今日の試合激しいでした!

kookoo yakyuu saikoo! kyoo no shiai hageshii deshita!
highschool baseball best! today GEN match tempestuous.COP.PAST
'High school baseball (is the) best! Today's match was tempestuous!'

In the example shown, the speaker has used the uninterrupted phrase (a phrase where the two keywords are not interrupted by any punctuation, spaces or paragraph changes) 激しいでした *hageshii deshita* 'was tempestuous'. The sentence used is relevant to the study; the sentence can be understood well, if the reader interprets the past polite copula *deshita* to be conjugating the *i*-adjective *hageshii* as an *i*-adjective. However when using Yahoo or Google search engines, punctuation or spaces often interrupted the text. An example of a interrupted, and thus irrelevant, example is given below.

インテンソわかりました！イタリア語で「激しい」でした。

intense wakarimashita! itariago de [hageshii] deshita.
Intenso understand.PAST Italian.language REL [intense] COP.PAST
'(I) understood what *intenso* means! In Italian (it means) [intense].'

In the example shown, the *i*-adjective *hageshii* is enclosed in Japanese style quotation marks (「激しい」). The two keywords (shown underlined in the example) are interrupted. The *i*-adjective *hageshii* 'intense' is quoted as:

(It) was ‘intense’ in Italian.

In other words, *hageshii* ‘intense’, in Example 2.2, is not inflected in the past tense but is still part of the search results because the search engine does not recognize the punctuation as an interruption. The search, thus, required a method using which the researcher could screen against phrases interrupted by punctuation, paragraph changes or spaces between the two keywords used in the search.

2. The search engines gave preference to the most popular pages containing the keywords mentioned in the search. This meant that there was no way of keeping the searches reliable for future studies or the results of the study easy to verify as the popularity of websites did changed even as the next search page was loaded. Furthermore, because the current study investigates language change, it was important that search results loaded in a chronological order.
3. The aforementioned search engines required the user to switch to a different page after a set number of results had been loaded onto the current page. The maximum amount that could be loaded on one page was a 100. Each time a new page was loaded it was possible that order in which the search results had been presented on the previous page had changed. The reliability of results depends on other researchers being able to get the same results if the experiment is reattempted. The inability of the aforementioned search engines to replicate previous search results made the study and the methodology unreliable.
4. The frequent appearance of language learning websites or dictionary websites in the search results was also problematic.

Figure 6: Search results received via Google Japan.

Figure 6 shows the search results for the keywords “悲しいでした” *kanashii deshita* ‘was sad’, received via Google Japan. The first five results include some from the websites (oshiete.goo.ne.jp), (hinative.com) and (soudan1.biglobe.ne.jp) all of which are question and answer forums, where authors have posted questions regarding whether this use is grammatically correct or not. These websites did not usually contain posts where speakers had conjugated *i*-adjectives as *na*-adjectives, but questions or explanations about how the conjugations should be used. For this reason they had to be discounted. Because websites like these were the most popular search results and the search engines presented the most popular websites first, comments that would actually be relevant to the study were hard to find.

5. Not all websites that appeared in the search results required authors of posts to have profiles, which meant it was sometimes hard to distinguish native Japanese speakers from proficient non-native Japanese speakers.

[10] This amount was doubled for many languages in 2017 but not for Japanese, Chinese and Korean (Newton 2017).

[11] According to the Twitter website the country selected "helps us to customize your Twitter experience, and may affect the content we are able to display".

[12] Another advantage of Twitter is that the search engine can load more posts by simply scrolling down the page (instead of switching from one page to the other). This allows a large amount of data to be loaded on the same page without the page getting refreshed and the loss of earlier data. More data being loaded onto the same page without the entire page getting refreshed meant that the browser's Find function (Ctrl +F) could be used on a large amount of data to count the number of 'exact strings' present. This is particularly helpful when counting up the amount of 'proper' examples that have occurred over a period of time as these sometimes stretched to the thousands in the case of some more frequent *i*-adjectives. In addition, because the Find function accounted for punctuation, spaces and paragraph changes between keywords, the exact strings were 'uninterrupted'.

[13] Figure 7 below shows the profile pictures and the user names of 4 Twitter users. It is unlikely that the usernames given are the actual names of the user.

Figure 7: Profile pictures and names of select users on Twitter

[14] During past presentations, the fact that Twitter posts are limited to 140 characters was brought up as a factor that may influence certain authors to switching from the 'proper' to the 'improper' conjugation, which is one character shorter. The example of the *i*-adjective *warui* 'bad' is given below.

	悪かったです	warukatta desu	'was bad (proper)'
mora	1 2 3 4 5 6		

	悪いでした	warui deshita	'was bad (improper)'
mora	1 2 3 4 5		

The *i*-adjective *warui* 'bad' when 'properly' conjugated forms the phrase *warukatta desu.* *warukatta desu* 'was bad' which is 6 characters long whereas when 'improperly' conjugated it forms the phrase *warui deshita* 'was bad' which is 5 characters long. I'd like to thank those who brought this point to my attention. However, as it was only a difference of one character and the posts themselves were often much shorter the maximum length allowed ("The average length of a tweet in Japanese is 15 characters, and only 0.4 percent of tweets hit the 140-character limit…" (Newton, 2017)), this was not considered a significant enough factor which would lead to the use of the conjugation. Though Twitter was used as the medium to study the 'improper' conjugation, it is not the only website where this usage occurs. As

discussed in chapter 1 (section 1.3), because posts containing the 'improper' conjugation have also been found on a variety of different websites, the character restriction on Twitter was determined not a driving factor in this usage.

[15] Though the Matsushita's database "Vocabulary Database for Reading Japanese (VDRJ) Ver.1.1" proved very helpful during the course of this project, we have to concede that because it was published in 2011, this ranking may have changed in the intervening time period and might not reflect the current situation

[16] They were from various fields in the humanities including literature, history, philosophy and social sciences such as law, economics. The database also used books from science and medicine.

[18] An example of a compound *i*-adjective is ***kokoro****motonai* 'uneasy', where the noun *kokoro* 'heart' is present within the *i*-adjective.

[19] I would like to distinguish between the Google search engine and Google chrome "Find" function. This function is available on most browsers but Chrome was used for this study.

[20] More frequent *i*-adjectives than this could not be used as, even for relatively short periods of time, the number of 'proper' examples available for these adjectives was so large the browser would crash while loading them.

[21] It was difficult to decide whether to use frequency among the *i*-adjectives only or to use the frequency compared to all words in the database. The frequency among *i*-adjective adjectives was chosen for the ease of the methodology. However, this issue may require further investigation.

[22] This was verified by Professor Ono Tsuyoshi who is a native Japanese speaker and an eminent linguist.

[23] This circumstance arose only once among all 7 surveys.

[24] The first two questions of the survey were asked in English. They were as follows:

1. Are you a native or a non-native speaker of Japanese?
2. Where in Japan are you from?

[25] The instructions given can be found on the copy of the survey, given in the appendix.

[26] 4 surveys were also completed by Teacher's Assistants from 100 level Japanese courses but, as those TAs had recently witnessed students learning the difference between *na*-adjective and *i*-adjective conjugations, data from those TAs was not included in the study.

[27] However, it may just be a case of an individual's preference, because as previously mentioned, participants also showed variability when assessing the dummy questions, numbered from 1 to 5. A definite inference cannot be made on the basis of this 7 response survey.

[28] This factor was brought to my attention by Miho Fujiwara.

[29] This factor was brought to my attention by Professors Hiromi Aoki and Xiaotin Li.

[30] The fact that noun phrases are often unmarked by particles is a distinguishing feature of spoken Japanese, when compared with written Japanese (Lee 2002).

[31] Sentence final particles are a common feature of Japanese dialogues (Katagiri 2007). In writing, they are used "to create a more casual and conversational tone" (Kakegawa, 2009: 310).

References

Backhouse, A. E. (2004). Inflected and uninflected adjectives. *Adjective Classes: A Cross-linguistic Typology*, 1, 50-73.

Baker, M. C. (2003). *Lexical Categories: Verbs, Nouns and Adjectives* (Vol. 102). Cambridge University Press.

Hasegawa, Y. (2014). *Japanese: A linguistic Introduction*. Cambridge University Press.

Kakegawa, T. (2009). Development of the use of Japanese sentence final particles through email correspondence. *Pragmatic Competence*, 301-334.

Kuno, S. (1973). *The Structure of the Japanese Language* (Vol. 3). Cambridge, MA: MIT Press.

Lee, D. Y. (2002). The function of the zero particle with special reference to spoken Japanese. *Journal of Pragmatics*, 34(6), 645-682.

Martin, S. E. (2004). A Reference Grammar of Japanese. University of Hawaii Press.

Matsushita, T. (2011, December 16). Vocabulary Database for Reading Japanese (VDRJ) Ver.1.1 (for Research) Top 60894 Words. Retrieved April 05, 2018, from http://www17408ui.sakura.ne.jp/tatsum/english/databaseE.html

Newton, Casey (2017). Twitter Just Doubled the Character Limit for Tweets to 280. *The Verge*. www.theverge.com/2017/9/26/16363912/twitter-character-limit-increase-280-test.

O'Grady, W., Archibald, J., & Katamba, F. (2011). *Contemporary Linguistics: An Introduction.* 2nd, [U.K.] ed. Harlow, UK: Pearson/Longman.

Shibatani, M. (1990). *The Languages of Japan*. Cambridge University Press.

Uehara, S. (1998). *Syntactic Categories in Japanese: A Cognitive and Typological Introduction*. Kurosio Publishers.

Uehara, S. (2003). A diachronic perspective on prototypicality: The case of nominal adjectives

in Japanese. *Cognitive Approaches to Lexical Semantics*, 363-391.

Vance, T. J. (1991). A new experimental study of Japanese verb morphology. *Journal of Japanese Linguistics*, 13(1), 145-166.

Appendix

The following is an exact copy of the survey that was given to the participants (Chapter 3).

Acceptability Test

Are you a native or a non-native speaker of Japanese?
Mark only one oval.

- Native
- Non-native

Where in Japan are you from?
以下の表現がどのくらい自然に聞こえるかお答えください。自然に聞こえる場合は「自然」をクリックしてください。やや自然に聞こえる場合は「やや自然」をクリックしてください。不自然に聞こえる場合は「不自然」をクリックしてください。

本が作る。
Mark only one oval.

- 自然
- やや自然
- 不自然

彼女のスカートは短いでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

猫をいる。
Mark only one oval.

- 自然
- やや自然
- 不自然

テレビで見ました。
Mark only one oval.

- 自然
- やや自然
- 不自然

子供が可愛いでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

本当に美味しいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

車にドライブする。
Mark only one oval.

- 自然
- やや自然
- 不自然

その机はメグには小さいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

チョコレートに食べます。
Mark only one oval.

- 自然
- やや自然
- 不自然

手が冷たいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

彼の言葉優しいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

試験に合格して嬉しいでした。
Mark only one oval.

- 自然
- やや自然

- 不自然

あなたと一緒にいて楽しいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

木陰は涼しいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

夕べは月が明るいでした。
Mark only one oval.

- 自然
- やや自然
- 不自然

Thank you!
Please use the space provided below to voice any concerns about the quality of the survey.