

# Annotating the Focus of Negation in Japanese Text

Suguru Matsuyoshi<sup>†</sup>, Ryo Otsuki<sup>‡</sup>, Fumiyo Fukumoto<sup>†</sup>

<sup>†</sup>Interdisciplinary Graduate School of Medicine and Engineering, University of Yamanashi,

<sup>‡</sup>Faculty of Engineering, University of Yamanashi,

4-3-11 Takeda, Kofu, Yamanashi 400-8511, Japan

{sugurum, t09kg005, fukumoto}@yamanashi.ac.jp

## Abstract

This paper proposes an annotation scheme for the focus of negation in Japanese text. Negation has its scope and the focus within the scope. The scope of negation is the part of the sentence that is negated; the focus is the part of the scope that is most prominently or explicitly negated. In natural language processing, correct interpretation of negated statements requires precise detection of the focus of negation in the statements. As a foundation for developing a negation focus detector for Japanese, we have annotated textdata of “Rakuten Travel: User review data” and the newspaper subcorpus of the “Balanced Corpus of Contemporary Written Japanese” with labels proposed in our annotation scheme. We report 1,327 negation cues and the foci in the corpora, and present classification of these foci based on syntactic types and semantic types. We also propose a system for detecting the focus of negation in Japanese using 16 heuristic rules and report the performance of the system.

**Keywords:** negation, focus of negation, annotation

## 1. Introduction

In recent years, there has been increasing attention paid to analyzing and processing complicated linguistic phenomena such as modality and negation (Morante and Sporleder, 2012; Morante and Blanco, 2012). Several text corpora annotated with them such as BioScope (Vincze et al., 2008) and FactBank (Saurí and Pustejovsky, 2009) have been developed. Negation has its scope and the focus within the scope. The scope of negation is the part of the sentence that is negated; the focus is the part of the scope that is most prominently or explicitly negated (Huddleston and Pullum, 2002). In this paper, we target the focus of negation in Japanese.

In natural language processing (NLP), correct interpretation of negated statements requires precise detection of the focus of negation in the statements. Here are two example sentences with negation cues, where a negation cue is written in a bold face, the scope is bracketed, and the focus is underlined<sup>1</sup>.

(1) [*kyou wa kuruma dewa ki mase*] **n** *deshi ta*.

(I didn't come here by car today.)

(2) *betsu ni* [*hairi taku te Hait ta*] **no de wa nai**.

(It's not because I wanted to join the club.)

As pointed out in the previous work (Huddleston and Pullum, 2002; Blanco and Moldovan, 2011a), a negated statement may have positive implicature. In Sentence (1), the statement “*kyou kuruma de ki ta* (I came here by car today)” is negated while we understand the statement “*kyou ki ta* (I came here today),” which is generated by removing the phrase “*kuruma de* (by car)” that is the focus from the original statement, is true. We can interpret from Sentence (2) that the writer in fact joined the club. If a computer detects the focus of negation in the statements precisely, this

brings progress of factuality analysis and recognizing textual entailment, whose technology is directly applicable to several NLP applications including information extraction, opinion mining and question answering.

In the field of NLP, several annotated corpora for negation scope and focus have been developed. BioScope (Vincze et al., 2008) is a pioneering corpus where negation cues and the scope are annotated. This corpus promoted development of systems for identifying negation cues and detecting the scope (Morante et al., 2008; Li et al., 2010). A subtask of \*SEM 2012 Shared task “*Resolving the Scope and Focus of Negation*” (Morante and Blanco, 2012) for detecting the scope of negation provided participants with a corpus of Conan Doyle stories where the scope of negation is annotated. Kawazoe et al. have been constructing a corpus of newspaper text in Japanese annotated with the scope of negation (Kawazoe et al., 2011).

Research on the focus of negation in NLP is much less than research on the scope of negation. Blanco and Moldovan have constructed a corpus annotated with the focus of negation (Blanco and Moldovan, 2011a). They use PropBank (Babko-Malaya, 2005) as textdata and propose a procedure of annotating the text with the focus of negation on top of predicate-argument structure in PropBank. They also propose methods of detecting the focus of negation by using machine learning and heuristic rules (Blanco and Moldovan, 2011a; Blanco and Moldovan, 2011b). Another subtask of the above \*SEM 2012 Shared task for detecting the focus of negation was done by using the annotated corpus (Morante and Blanco, 2012). In this task, Rosenberg and Bergler proposed a method of detecting the focus of negation by using four heuristics (Rosenberg and Bergler, 2012). Matsuyoshi et al. deal with the focus of negation as an element for extended modality annotation in Japanese (Matsuyoshi et al., 2010). However, they annotated a very small number of instances because negation focus is not the main element for extended modality.

This paper proposes an annotation scheme for the focus of

<sup>1</sup>The translation of the Japanese sentences into English is in parentheses.

negation in Japanese text. As a foundation for developing a negation focus detector, we have annotated textdata of “Rakuten Travel: User review data” and the newspaper sub-corpus of the “Balanced Corpus of Contemporary Written Japanese” with labels proposed in our annotation scheme.

## 2. Proposed Scheme

### 2.1. Definition of Negation Cues in Japanese

The Japanese language has several types of negation cues. We classify these into the following three types roughly and six types in detail:

#### Function word

auxiliary verb “*nai* (not)” and “*zu* (not)”; suffix “*-nai* (not)”; prefix “*hi* (un-), “*fu* (un-), “*mu* (no), “*mi* (un-), “*han* (anti-), and “*i* (another)”

#### Compound auxiliary verb representing negation

“*no de wa nai* (it is not . . . that)”, “*wake de wa nai* (it is not . . . that)” “*wake ni wa ika nai* (cannot)” and so on.

#### Content word

adjective “*nai* (does not exist)” and noun “*nashi* (no existence)”

Development of compound auxiliary verbs for representing negation is one of the significant features that the Japanese language has.

### 2.2. Pseudo-Negation Phrases

There are idiomatic phrases that literally include negation cues but do not negate propositions. We call these phrases pseudo-negation phrases. In annotating text in Japanese, we pay attention to them. We classify pseudo-negation phrases into the following two types:

**Idiom** “*mono tari nai* (insufficient)”, “*shikata ga nai* (there is no way)”, “*omo-wa zu* (in spite of myself)” and so on.

**Compound functional expressions that do not negate propositions** “*nakere ba nara nai* (have to)”, “*nai to ike nai* (have to)”, “*ka mo shire mase n* (may)”, “*ni mo kakawara zu* (although)”, “*dake de naku* (not only)” and so on.

We have collected these phrases from Japanese dictionaries (Matsumura and Editors, 1998; Nishio et al., 2000) and books in linguistics (Morita and Matsuki, 1989; Jamasi, 1998).

### 2.3. A Guideline for Annotation

We created a guideline for judging which segment is the focus of a target negation cue as follows.

1. Look around the broad context of the sentence including a target negation cue.
2. Generate a statement from the original sentence by removing the target negation cue and an argument segment. Judge the segment as the focus of the negation if you can infer that the generated statement is true.

3. Try procedure 2. for each segment in the sentence.

4. If you can not find the segment that is the focus, no segment in the sentence is prominently negated. We call this situation the case that “the focus of negation is the whole scope of the negation” as a matter of practical convenience. In the case, the sentence does not carry positive meaning.

### 2.4. Tags for a Negation Cue and its Focus

We annotate a negation cue with the following three tags:

#### Surface forms of morphemes

##### Morpheme IDs

**Part of speech** auxiliary verb, suffix, prefix, negative compound auxiliary verb, adjective and noun.

We annotate the segment that is the focus of negation with the following seven tags:

#### Surface forms

##### Morpheme IDs

**Syntactic type** a classification of negation focus based on constituent morphemes and predicate-argument structure.

- “*ga* (subject)”-case, “*wo* (direct object)”-case, “*ni* (indirect object)”-case, “*de* (locative)”-case and so on.
- Adverbial phrase, Adnominal “*no* (of)” phrase, Adnominal verb phrase.
- “*te* (because)”-clause, “*to* (when)”-clause and so on.

This type can be identified by a parser and a predicate-argument structure analyzer. Such information may be effective in detecting the focus of negation.

**Semantic type** a classification of negation focus based on semantic interpretation.

- The segment of negation focus limits scope of an argument that makes the target statement be true: limit-agent, limit-object, limit-temporal, limit-locative, limit-quantity, and limit-others.
- The segment of negation focus is an adjunct for representing manner, frequency, aspect and so on: adjunct-continuous modification, adjunct-adnominal modification, and adjunct-aspect.

This type can be identified using a verb frame dictionary, a phrase thesaurus and other language resources. Such information may also be effective in detecting the focus of negation.

**Key expressions** expressions in the text used in judging a segment as the focus of negation.

**Focus particle** whether a focus particle such as “*shika* (only)” and “*mo* (emphasizing)” is in the segment.

**Comment** description of the reason why the segment was judged as the focus of negation.

### 3. Annotating Textdata in Two Corpora

We use the following two textdata for negation focus annotation:

- Rakuten Travel: User review data<sup>2</sup>
- Newspaper subcorpus of the “Balanced Corpus of Contemporary Written Japanese (BCCWJ)”<sup>3</sup>

Annotated instances are given in subsections of section 4.

#### 3.1. Rakuten Travel: User Review Data

This data contains about 4.7 million reviews for 82,458 facilities. We randomly selected 40 facilities each of which has 10 to 58 reviews for it. The reviews for the facilities contains 5,178 sentences. We call this text “Review.”

#### 3.2. Newspaper Subcorpus of BCCWJ

BCCWJ contains several types of text including newspaper, books, magazines, blogs and Q&A documents. We use Group “A” and “B” in the newspaper subcorpus of BCCWJ. This contains 5,582 sentences<sup>4</sup>. We call this text “Newspaper.”

#### 3.3. Annotation Process and Status of Annotated Corpora

We used a Japanese morphological analyzer MeCab<sup>5</sup> for POS tagging. Then, we marked morphemes as a negation cue candidate by simple use of POS information. The numbers of the candidates for Review and Newspaper are 1,246 and 901, respectively. We judged whether they are negation cues or pseudo-negation phrases. Table 1 shows that the numbers of the negation cues for Review and Newspaper are 1,023 and 762, respectively. The table indicates that auxiliary verbs are the most common cues in either of the textdata. The numbers of the negation foci that are not the whole scope of the negation are 300 and 190 for Review and Newspaper, respectively. Therefore, for 29% (300/1,023) and 25% (190/762) of the negation cues in Review and Newspaper, respectively, negation focus should be precisely identified for correct interpretation of negated statements.

We show classifications of syntactic types and semantic types of negation focus in Tables 2 and 3, respectively. As for Newspaper, we performed the semantic type classification for only Group “A” in the newspaper subcorpus of BCCWJ. Table 2 indicates that adverbs are the major segments as negation focus in Review. This is why adjunct-continuous modification which is mainly labeled to adverbs is dominant in Review at Table 3. These tables reveal that distributions of syntactic types and semantic types vary in different text genres.

<sup>2</sup>[http://rit.rakuten.co.jp/rdr/index\\_en.html](http://rit.rakuten.co.jp/rdr/index_en.html)

<sup>3</sup><http://www.ninjal.ac.jp/english/products/bccwj/>

<sup>4</sup>More precisely, the text contains 5,582 <SENTENCE> XML tags that indicate a sentence, or a sentence segment surrounded by quotations.

<sup>5</sup><http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>

Table 4 shows the numbers of focus particles in the segments judged as negation focus. From the table, we found out that 34% (165/490) of the negation foci includes focus particles. Particle “*wa* (contrast)” appears a useful clue for detecting the focus of negation, but it is commonly used in a negated sentence where the focus of negation is the whole scope of negation.

At the first step of annotation process, two annotators independently judged the focus of negation for Group “A” in the newspaper subcorpus of BCCWJ. A ratio of the number of segments for which the two annotators agree to the total number of negation cues is 66% (201/304). It took only two hours for them to discuss disagreement problems and solve all of them. The main causes of the disagreement are misunderstanding of the scope of negation and paying little attention to the broad context of a negated sentence. Then, one of the annotators judged the focus of negation for Review and Group “B” in the newspaper subcorpus of BCCWJ. Finally, the other annotator checked the annotated segments in the corpora, and by the way of discussion, about ten labels of negation foci were corrected.

## 4. System for Detecting the Focus of Negation in Japanese

We developed a system for detecting the focus of negation in Japanese using heuristic rules. Here, we describe an overview of the system. The following subsections explain 16 heuristic rules we constructed for detecting the focus of negation.

The proposed system receives a sentence where a negation cue is marked with a label “NEG,” and marks the focus of the negation with a label “FOC.” This system does not identify the scope of negation. If a negation cue in the sentence is a prefix such as “*hi* (un-)” and “*fu* (un-),” the system determines that its focus is the whole scope of the negation. Otherwise, the system uses heuristic rules one by one in order of priority. First, it applies the rule as described in subsection 4.1., whose priority is one, to the negated sentence. If the rule doesn’t detect the focus of negation, then the system applies the rule whose priority is two to the sentence. It applies heuristic rules to the sentence until the focus of negation is determined. When none of the 16 rules doesn’t detect negation focus, the system determines that the focus of the negation is the whole scope of the negation.

In most cases, a clause with a connective particle such as “*ga* (but),” “*ba* (if),” “*to* (when)” and “*keredomo* (although)” is not included in the scope of a negation cue for the main clause. Therefore, the system does not search such clause for the focus of negation.

In cases where a given sentence has more than one negation cue, our system deals with them independently.

### 4.1. Focus Particles

When the negated predicate has a segment with one of the focus particles “*shika* (only),” “*dake* (only),” “*made* (even)” and “*hodo* (contrast),” the segment is determined as the focus of the negation, as in Sentence (3), where a negation cue is written in a bold face and the focus is underlined.

	Review	Newspaper	Total
Auxiliary verb	637 (62%)	446 (58%)	1,083 (61%)
Suffix	116 (11%)	33 ( 5%)	149 ( 8%)
Prefix	19 ( 2%)	89 (11%)	108 ( 6%)
Adjective	211 (21%)	175 (23%)	386 (22%)
Noun	28 ( 3%)	14 ( 2%)	42 ( 2%)
Negative compound auxiliary verb	12 ( 1%)	5 ( 1%)	17 ( 1%)
(Subtotal)	(1,023) (100%)	(762) (100%)	(1,785) (100%)
Idiom	94	66	160
Other compound auxiliary verb	121	74	195
Error of a POS tagger	8	0	8
(Subtotal)	(223)	(139)	(362)
Total	1,246	901	2,147

Table 1: Numbers of negation cues and pseudo-negation phrases

	Review	Newspaper	Total
Adverb	140 (47%)	82 (43%)	222 (45%)
“ <i>ga</i> (subject)”-case	30 (10%)	7 ( 4%)	37 ( 8%)
“ <i>wo</i> (direct object)”-case	7 ( 2%)	8 ( 4%)	15 ( 3%)
“ <i>ni</i> (indirect object)”-case	49 (16%)	32 (16%)	81 (17%)
“ <i>de</i> (locative)”-case	17 ( 6%)	8 ( 4%)	25 ( 5%)
“ <i>made</i> (to)”-case	5 ( 2%)	11 ( 6%)	16 ( 3%)
“ <i>kara</i> (from)”-case	3 ( 1%)	2 ( 1%)	5 ( 1%)
“ <i>to</i> (with)”-case	3 ( 1%)	1 ( 1%)	4 ( 1%)
Other case	1 ( 0%)	3 ( 2%)	4 ( 1%)
Adnominal “ <i>no</i> (of)” phrase	20 ( 7%)	20 (11%)	40 ( 8%)
Adnominal verb phrase	8 ( 3%)	13 ( 6%)	21 ( 4%)
Prefix “ <i>zen</i> (all)”	1 ( 0%)	0 ( 0%)	1 ( 0%)
“ <i>te</i> (because)”-clause	1 ( 0%)	2 ( 1%)	3 ( 1%)
“ <i>to</i> (when)”-clause	1 ( 0%)	0 ( 0%)	1 ( 0%)
Aspect of the verb	14 ( 5%)	1 ( 1%)	15 ( 3%)
Total	300 (100%)	190 (100%)	490 (100%)

Table 2: Classification of the negation foci based on syntactic types

	Review	Newspaper	Total
Limit-agent	13 ( 4%)	5 ( 7%)	18 ( 5%)
Limit-object	27 ( 9%)	12 (17%)	39 (10%)
Limit-temporal	10 ( 3%)	9 (12%)	19 ( 5%)
Limit-locative	40 (14%)	3 ( 4%)	43 (12%)
Limit-quantity	10 ( 3%)	5 ( 7%)	15 ( 4%)
Limit-others	43 (14%)	12 (17%)	55 (15%)
Adjunct-continuous modification	124 (42%)	15 (21%)	139 (37%)
Adjunct-adnominal modification	19 ( 6%)	11 (15%)	30 ( 8%)
Adjunct-aspect	14 ( 5%)	0 ( 0%)	14 ( 4%)
Total	300 (100%)	72 (100%)	372 (100%)

Table 3: Classification of the negation foci based on semantic types

	Review	Newspaper	Total
“ <i>wa</i> (contrast)”	66 (62%)	34 (58%)	99 (60%)
“ <i>shika</i> (only)”	34 (32%)	20 (34%)	54 (33%)
“ <i>mo</i> (emphasizing)”	7 ( 6%)	2 ( 3%)	9 ( 5%)
“ <i>dake</i> (only)”	0 ( 0%)	3 ( 5%)	3 ( 2%)
Total	106 (100%)	59 (100%)	165 (100%)

Table 4: Focus Particles in the segments judged as negation focus

(3) *Fudan wa kimatta mono shika ryori shi nai.*

(I normally make nothing but regular dishes.)

#### 4.2. Adverbial Phrases Representing Degree

When an adverbial phrase representing degree such as “*amari* (much)”, “*nakanaka* (hardly)” and “*hobo* (little)” is in the sentence, the phrase is determined as the focus of the negation, as in Sentence (4).

(4) *shigaichi dewa kogai de yozora no hoshi ga nakanaka mie nai.*

(You can hardly see stars in the sky of an urban area because of light pollution.)

#### 4.3. Adverbial Phrases Representing Time

An adverbial phrases representing time such as “*konkai* (this time)”, “*jikai* (next time)” and “*... irai* (since ...)” is determined as the focus of the negation, as in Sentence (5), as in the case of the previous subsection.

(5) *sono senshu wa ganjitsu irai shiai ni de te i nai.*

(The player has been absent from games since New Year’s Day.)

#### 4.4. Adverbial Phrases Representing Frequency

An adverbial phrases representing frequency such as “*himpan ni* (frequently)”, “*shocchu* (very often)” and “*yoku* (often)” is determined as the focus of the negation, as in Sentence (6).

(6) *himpan niwa riyo deki nai o-nedan deshita.*

(The price was too expensive for me to use this hotel frequently.)

#### 4.5. Adverbial Phrases Representing Manner

An adverbial phrases representing manner such as “*umaku* (well)”, “*yukkuri* (slowly)” and “*heizen to* (calmly)” is determined as the focus of the negation, as in Sentence (7).

(7) *kono hoteru wa kanki ga umaku deki te i nai.*

(I stayed at an inadequately-ventilated room in this hotel.)

#### 4.6. Aspect of Perfection

When the negated predicate takes a perfective aspect form “*kire*,” the word is determined as the focus of the negation, as in Sentence (8).

(8) *shokuji wa tabe kire nai hodo no ryo de, oishikatadesu.*

(The meal that this hotel provided was good and the quantity of the meal was so adequate that we couldn’t eat the whole thing.)

#### 4.7. Partial Negation

For partial negation, we determine a word representing “all” or “every” such as “*subete*” and “*zen*” as the focus of the negation, as in Sentence (9).

(9) *keki ya pan ga subete tezukuri to iu wake dewa nasa sodesu.*

(I guess that not every cake and every bread are hand-made.)

#### 4.8. Syntactic Pattern: “As for”

When the negated sentence includes a topic segment marked with “*dewa* (as for)” or “*deno* (as for)” and an emphasis for contrast with Particle “*wa*,” the topic segment is determined as the focus of the negation, as in Sentence (10).

(10) *saibansho ni makaseru to iu youna houhou dewa kono kiki wa norikire nai.*

(A way of leaving the problem to a law court can’t overcome this crisis.)

#### 4.9. Immediate “*ni*”-Case

We determine a segment marked with Particle “*ni*” that is immediate before the negated predicate as the focus of the negation, as in Sentence (11).

(11) *sude ni gohan ga ja ni naku, kakari no kata ga renji de chin shi te dashi te kure mashita.*

(The jar rice cooker didn’t have rice, so a waiter served rice warmed up in the microwave to me.)

#### 4.10. Numeral with Particle “*mo*”

A numeral with Particle “*mo*” is determined as the focus of the negation, as in Sentence (12).

(12) *kosoku de 20 pun mo kakara zu bijinesu demo riyo dekiru.*

(Because it takes less than 20 minutes to get to the hotel with the expressway, it is also suitable for business use.)

#### 4.11. Syntactic Pattern: “while”

An expression “*no hou no*” compares two similar entities. We determine a segment just before the expression as the focus of the negation, as in Sentence (13).

(13) *shawa busu no hou no toire ga danbo benza dewa nai no ga tsurai.*

(It is not good that the toilet in the shower booth didn’t have a heated seat while the other did.)

#### 4.12. Locative with Particle “*wa*”

When the negated sentence has a locative with Particle “*wa*,” the phrase is determined as the focus of the negation, as in Sentence (14).

(14) *onsen igai niwa tomara nai.*

(I won’t stay at any hotels but ones with a hot spring.)

#### 4.13. Immediate Particle “*wa*” for Noun Negation Cue

When a negation cue is Noun “*nashi* (no existence)” and a segment with Particle “*wa*” is immediate before the negated noun phrase, the segment is determined as the focus of the negation, as in Sentence (15).

(15) *eki karawa taihen chikaku ricchi wa monku nashi.*

(I have no complaints for location because this hotel is very near the station.)

	Review		Newspaper		Total	
	# of Sent.	# of Neg.	# of Sent.	# of Neg.	# of Sent.	# of Neg.
Training data	2,209	437	2,708	304	4,917	741
Test data	2,969	586	2,874	458	5,843	1,044
Total	5,178	1,023	5,582	762	10,760	1,785

Table 5: Numbers of sentences and negation cues in our corpora

#### 4.14. Syntactic Pattern for Adnominal “no” Phrase

When an adnominal “no (of)” phrase modifies a word marked with Particle “wa,” the adnominal phrase is determined as the focus of the negation, as in Sentence (16).

(16) *kore ijou no enmei wa shinobi nai.*

(We can’t bear another life-prolonging care for him.)

#### 4.15. Syntactic Pattern for Immediate Adnominal “na” Phrase

As in the case of the previous subsection, an immediate adnominal “na” phrase is determined as the focus of the negation, as in Sentence (17).

(17) *goka na puran nashi.*

(This hotel provides no plan of a luxury room.)

#### 4.16. “ni”-Case at the Beginning of the Sentence

When there is a segment with Particle “ni” at the beginning of the negated sentence, the segment is determined as the focus of the negation, as in Sentence (18).

(18) *heya ni reizoko ga naku robi ni aru kyodo reizoko wo tsukatta.*

(The room where I stayed had no fridge, so I used common one in the lobby.)

## 5. Experiments

We conduct experiments for evaluating performance of our system described in the previous section.

### 5.1. Experimental Settings

We perform the following three types of experiments:

**Overall** Whether a system can detect the focus of negation for all negation cues in the corpora, including ones whose foci are the whole scopes of negation.

**Binary** Whether a system can classify a negation cue into the following two classes: its focus is the part of the scope; and its focus is the whole scope.

**SegSelect** Whether a system can select a segment in the negated sentence for the only negation cues whose foci are the parts of the scope.

As described in section 3., about 70% of negation cues in our corpora are ones whose foci are the whole scopes of negation. This leads a system which outputs “the whole scope” as negation focus for every negation cue to about 70% of accuracy. Therefore, we conduct the above Binary

	Review	Newspaper
Baseline	69% (301/437)	75% (227/304)
Proposed	84% (366/437)	86% (262/304)

Table 6: Results of Overall experiments using the training data

	Review	Newspaper
Baseline	68% (397/586)	73% (334/458)
Proposed	73% (430/586)	80% (369/458)

Table 7: Results of Overall experiments using the test data

classification and SegSelect experiment in addition to Overall experiment.

We use our corpora described in section 3. for evaluating performance of the system. We employ a half part of our corpora as a training data for developing the rules as described in section 4. The remaining half part of the corpora is used as a test data for an open test. Table 5 shows the numbers of sentences and negation cues in the training data and test data. We adopt accuracy as evaluation measure.

We implemented a baseline system where an adverbial phrase near the negated predicate is determined as the focus of negation if the phrase exists. This is because Table 2 shows that adverbs are the major segments as the focus of negation. If no adverbial phrase exists in a given sentence, this system determines that its focus is the whole scope of negation. The baseline system also does not search a clause with a connective particle such as “ga (but)” and “ba (if)” for the focus of negation. In the experiments, we compare the proposed system with the baseline system.

### 5.2. Results

Tables 6 and 7 show the results of Overall experiments using the training data and the test data, respectively. Evaluation measure in the tables is accuracy that is a ratio of the number of correctly identified foci to the total number of negation cues in the corpus. In Overall experiments, the proposed system achieved 73% (430/586) and 80% (369/458) in the measure for the test data of Review and Newspaper, respectively, as shown in Table 7. We think that these figures are sufficient for a first step of negation focus detection.

Tables 8 and 9 show the results of Binary classification experiments using the training data of Review and Newspaper, respectively. In these tables, performance of the proposed system for Part-Part cells is more than twice that of the baseline. On the other hand, there is no signifi-

		Total	Baseline		Proposed	
			Part	Whole	Part	Whole
Gold	Part	124	44	80	100	24
Standard	Whole	313	43	270	42	271
Total		437	87	350	142	295

Table 8: Results of Binary classification (training data, Review)

		Total	Baseline		Proposed	
			Part	Whole	Part	Whole
Gold	Part	73	12	61	53	20
Standard	Whole	231	13	218	14	217
Total		304	25	279	67	237

Table 9: Results of Binary classification (training data, Newspaper)

cant difference between performance of the two systems for Whole-Whole cells. This is because about 70% of negation cues in our corpora are ones whose foci are the whole scopes of negation, and because both of the two systems determine that the focus of negation is the whole scope when none of rules doesn't detect negation focus.

Tables 10 and 11 show the results of Binary classification experiments using the test data of Review and Newspaper, respectively. In these tables, performance of the proposed system for Part-Part cells is about twice as much as that of the baseline. On the other hand, there is no significant difference between performance of the two systems for Whole-Whole cells, as in the case of the experiments with the training data. The proposed system has no aggressive rule for determining that the focus of negation is the whole scope of negation. We need to develop such rules for improvement of the proposed system.

Tables 12 and 13 show the results of SegSelect experiments using the training data and the test data, respectively. Evaluation measure in the tables is accuracy that is a ratio of the number of correctly identified foci to the number of only negation cues whose foci are the parts of the scope. In SegSelect experiments with the training data, the proposed system achieved 76% (95/125) and 62% (45/72) in the measure for Review and Newspaper, respectively, as shown in Table 12. This indicates that the 16 heuristic rules are effective for detecting the focus of negation in Japanese. The figure for Newspaper is less than that for Review by 14%. As shown in Table 2, Review has more adverbs as negation foci than Newspaper. We found out that the rules of the proposed system were robust over adverbs as negation foci while they were insufficient for segments characteristic of Newspaper.

In SegSelect experiments with the test data, the proposed system achieved 53% (93/175) and 51% (60/118) in the measure for Review and Newspaper, respectively, as shown in Table 13. Although these figures far exceed the ones for the baseline, they are about 50%. We conducted error analysis for SegSelect experiments. A major cause is that it is too hard to judge whether an argument without a outstand-

		Total	Baseline		Proposed	
			Part	Whole	Part	Whole
Gold	Part	173	63	110	114	59
Standard	Whole	413	67	346	76	337
Total		586	130	456	190	396

Table 10: Results of Binary classification (test data, Review)

		Total	Baseline		Proposed	
			Part	Whole	Part	Whole
Gold	Part	118	24	94	72	46
Standard	Whole	340	26	314	31	309
Total		458	50	408	103	355

Table 11: Results of Binary classification (test data, Newspaper)

ing characteristic is the focus of negation, as in Sentence (19).

- (19) *furo wa arai<sub>ba</sub> ni jaguchi ga nakatta node sukoshi fuben deshita.*

(It was a little inconvenient because the washing place in the bathroom has no faucet.)

It is important to observe instances in our corpora for developing additional rules. A type, a conjugated form and tense of a negation cue may be available for conditions of rules.

Another major cause is that candidate selection requires information obtainable from context. Each of Examples (20) and (21) contains two sentences. The first sentences of the two are exactly the same string and include a negation cue “*nai*.”

- (20) *arukoru wa kombini ni oi te i nai. chikaku no sakaya de kaeru ga, hoteru nai no mise no hou ga yasui.*

(A convenience store near the hotel has no alcoholic drinks. A liquor shop near the hotel sells them, but a souvenir shop in the hotel provides them more cheaply than the liquor shop.)

- (21) *arukoru wa kombini ni oi te i nai. dakara kyo wa ocha wo katta.*

(A convenience store near the hotel has no alcoholic drinks. So, I bought a bottle of Japanese tea in the store today.)

The first sentence of Example (20) has the following two candidate segments as negation focus: “*arukoru wa*” and “*kombini ni*.” The second sentence disambiguates the problem because a liquor shop and a souvenir shop has alcoholic drinks, i.e., “*arukoru wa oi te iru* (Some shop near the hotel has alcoholic drinks)” is true.

The second sentence of Example (21) also disambiguates the above problem because it is found that the convenience store has some drinks other than alcoholic drinks.

We need a framework for using context and a method of dealing with information obtainable from the context.

	Review	Newspaper
Baseline	25% (31/125)	12% ( 9/72)
Proposed	76% (95/125)	62% (45/72)

Table 12: Results of SegSelect experiments using the training data

	Review	Newspaper
Baseline	29% (51/175)	17% (20/118)
Proposed	53% (93/175)	51% (60/118)

Table 13: Results of SegSelect experiments using the test data

## 6. Conclusion

In this paper, we proposed an annotation scheme for the focus of negation in Japanese text. We reported our annotated corpora of 1,327 negation cues and the foci<sup>6</sup>. We implemented a system for detecting the focus of negation in Japanese using 16 heuristic rules. This system achieved 73% and 80% in accuracy for the test data of Review and Newspaper, respectively.

Our future work contains the following two tasks: one is to apply our annotation scheme to different kinds of resources for assessing the scheme and extending our corpora; the other is to improve a detector of the focus of negation using machine learning approaches, such as support vector machines and conditional random fields, and the corpora as the training data.

## 7. Acknowledgements

This work was partly supported by JSPS Grant-in-Aid for Young Scientists (B) (KAKENHI Grant Number: 25870278).

## 8. References

- Babko-Malaya, O., (2005). *PropBank Annotation Guidelines*. ACE (Automatic Content Extraction) Program. <http://verbs.colorado.edu/~mpalmer/projects/ace/PBguidelines.pdf>.
- Blanco, E. and Moldovan, D. (2011a). Semantic Representation of Negation Using Focus Detection. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pages 581–589.
- Blanco, E. and Moldovan, D. (2011b). Some Issues on Detecting Negation from Text. In *Proceedings of the 24th International Florida Artificial Intelligence Research Society Conference*, pages 228–233.
- Huddleston, R. and Pullum, G. K., editors. (2002). *The Cambridge Grammar of the English Language*. Cambridge University Press.
- Jamasi, G., editor. (1998). *A Dictionary of Japanese Sentence Patterns for Teachers and Learners*. Kurosio Publishers. (in Japanese).

Kawazoe, A., Saito, M., Kataoka, K., Choi, S.-Y., and Bekki, D., (2011). *A Guideline of Annotation for Japanese Expressions and their scope Affecting Certainty Version*. 2.4. Technical Report of Department of Information Science, Ochanomizu University. (in Japanese).

Li, J., Zhou, G., Wang, H., and Zhu, Q. (2010). Learning the Scope of Negation via Shallow Semantic Parsing. In *Proceedings of the 23rd International Conference on Computational Linguistics (Coling 2010)*, pages 671–679.

Matsumura, A. and Editors, D., editors. (1998). *Daijisen Dictionary*. Shogakukan Inc. (in Japanese).

Matsuyoshi, S., Eguchi, M., Sao, C., Murakami, K., Inui, K., and Matsumoto, Y. (2010). Annotating event mentions in text with modality, focus, and source information. In *The seventh international conference on Language Resources and Evaluation (LREC)*.

Morante, R. and Blanco, E. (2012). \*sem 2012 shared task: Resolving the scope and focus of negation. In *\*SEM 2012: The First Joint Conference on Lexical and Computational Semantics – Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (SemEval 2012)*, pages 265–274.

Morante, R. and Sporleder, C. (2012). Modality and negation: An introduction to the special issue. *Computational Linguistics*, 38(2):223–260.

Morante, R., Liekens, A., and Daelemans, W. (2008). Learning the Scope of Negation in Biomedical Texts. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 715–724.

Morita, Y. and Matsuki, M. (1989). *Japanese Sentence Patterns: Mean and usage of Compound Particles and Compound Auxiliaries*. ALC Press Inc. (in Japanese).

Nishio, M., Iwabuchi, E., and Mizutani, S., editors. (2000). *Iwanami Japanese Dictionary 6th Edition*. Iwanami Shoten, Publishers. (in Japanese).

Rosenberg, S. and Bergler, S. (2012). UConcordia: CLaC Negation Focus Detection at \*Sem 2012. In *Proceedings of the First Joint Conference on Lexical and Computational Semantics: SemEval’12*, pages 294–300.

Saurí, R. and Pustejovsky, J. (2009). FactBank: A Corpus Annotated with Event Factuality. *Language Resources and Evaluation*, 43(3):227–268.

Vincze, V., Szarvas, G., Farkas, R., Móra, G., and Csirik, J. (2008). The BioScope Corpus: Biomedical Texts Annotated for Uncertainty, Negation and their Scopes. In *BMC Bioinformatics*, pages 1–9.

<sup>6</sup>We plan to distribute them freely at our web site.

<http://cl.cs.yamanashi.ac.jp/nldata/negation/>