A Lexicon-based Investigation of Research Issues in Japanese Factuality Analysis

Kazuya Narita[*]**, Junta Mizuno**[†] **and Kentaro Inui**[*]
[*]Graduate School of Information Sciences, Tohoku University, Japan
[†]National Institute of Information and Communications Technology (NICT), Japan
{narita, junta-m, inui}@ecei.tohoku.ac.jp

Abstract

Event factuality is information about whether events mentioned in natural language correspond to either actual events that have occurred in the real world or events that are of uncertain interpretation. Factuality analysis is useful for information extraction and textual entailment recognition, among others, but sufficient performance has not yet been achieved by the machine learning-based approach. It is now important to take a closer look at the linguistics phenomena involved in factuality analysis and identify the technical research issues more precisely. In this paper, we discuss issues regarding lexical knowledge through error analysis of a Japanese factuality analyzer based on lexical knowledge and compositionality.

1 Introduction

Event factuality is information about whether events mentioned in natural language correspond to either actual events that have occurred in the real world or events that are of uncertain interpretation.

(1) a.
   kare-wa sakihodo heya-wo de-ta.
   (*He left the room a little while ago.*)

   b.
   mou osoi-kara, kare-wa saki-ni kaet-ta-no-daro-u.
   (*It's late now, so he may have gone home.*)

   c.
   mondai-ga hassei-suru-no-wo fusei-da.
   (*We prevented the occurrence of the problem.*)

For example, we can interpret that the event "*de*" (*leave*) in (1a) is factual in the real world, the event "*kaet*" (*go home*) in (1b) is possibly factual because of the modal auxiliary "*-ta-no-daro-u*" (*may have -ed*), and the event "*hassei-suru*"

(*occurrence*) in (1c) is counterfactual because of the implicative predicate "*fusei-da*" (*prevented*).

Factuality analysis is useful for a broad range of NLP applications such as information extraction, question answering, and textual entailment recognition. Prior work on factuality analysis has made considerable efforts for designing and creating corpora manually annotated with factuality-related information (Saurí and Pustejovsky, 2009; Matsuyoshi et al., 2010; Tanaka et al., 2013, etc.) and several empirical studies on those resources are reported revealing the difficulties of the task (Inui et al., 2008; Matsuyoshi et al., 2010; Morante and Blanco, 2012; Saurí and Pustejovsky, 2012). For Japanese, Matsuyoshi et al. (2010) report that their factuality classes are highly skewed and the minority classes are very difficult for their machine learning-based models to precisely identify. The minority classes include uncertain statements as in example (1b) and counterfactual statements as in (1c). Such "marked" statements are far less frequent than unmarked statements (i.e. certain factual statements) and thus are not as easy to collect as unmarked statements. While the label distribution is reported to be less skewed in English (Szarvas et al., 2008), still uncertain and counterfactual statements constitute minority classes. In addition, uncertain and counterfactual statements exhibit a very broad variety of linguistic devices for expressing uncertainty and negation. For those reasons, the whole task is not as easy as it appears and simple strategies based on supervised machine learning do not work well.

Given this background, rather than putting everything simply into a machine learning algorithm, it is now important to take a closer look at the linguistics phenomena involved in factuality analysis and identify the technical research issues more precisely. One promising way for it is to make use of existing lexical resources and divide the whole issues into those related to lexical knowledge and the rest. We take this approach in this

587

paper because (i) the factuality status is primarily expresses by lexical devices such as auxiliaries (e.g. "*-ta-no-darou*" (*may have -ed*)) and factual and counterfactual predicates (e.g. "*fusegu*" (*prevent*)), and (ii) there are existing Japanese lexicons of such factuality-related expressions (factuality markers, henceforth) available with a reasonably broad coverage. As a platform for computing factuality with factuality markers, we adopt Saurí and Pustejovsky's rule-based model for English factuality analysis (Saurí and Pustejovsky, 2012) and adapt it to the Japanese language. Saurí and Pustejovsky's model is suitable as it assumes the availability of a factuality lexicon and uses it to identify the factuality status of each subordinate event in a compositional manner from the factuality status of its superordinate event. For lexical resources, we use the dictionary of Japanese functional expressions (Matsuyoshi et al., 2007) and the dictionary of Japanese clue expressions for extended modality (Eguchi et al., 2010). This paper presents a *first* comprehensive investigation in Japanese factuality analysis, which is based on these sufficient lexicons.

This paper is organized as follows. Section 2 describes related work. In Section 3, we construct a Japanese factuality analyzer based on compositional approach by Saurí and Pustejovsky (2012). In Section 4, we discuss issues regarding lexical knowledge through error analysis by applying our analyzer with Japanese text. Based on the analysis in Section 4, Section 5 discusses lexicon-based scope detection. Section 6 concludes this paper.

2 Related work

Previous work for an annotation schema of factuality and other associated information includes FactBank (Saurí and Pustejovsky, 2009), Japanese corpus with extended modality (Matsuyoshi et al., 2010), and so on. Saurí and Pustejovsky annotate event mentions with its source, epistemic modality (certainty) and polarity for representing the event factuality. Additionally, their FactBank is extended with pragmatically informed factuality judgments by de Marneffe et al. (2012). Matsuyoshi et al. mark up an event mention with seven components (*source, time, conditional, primary modality type, actuality, evaluation*, and *focus*). Our factuality corresponds to *actuality*. Tanaka et al. (2013) annotate the sense and usage of ambiguous expressions related to factuality.

For automatically analyzing factuality in text, there are approaches based on machine learning.

Inui et al. (2008) have proposed a method of analyzing modality and polarity of event mentions in Japanese text with an approach based on conditional random field. However, it is very difficult that their machine learning-based models precisely identify the minority classes.

There are also approaches based on rules. MacCartney and Manning (2009) have proposed a model of *natural logic*, which has focused on semantic containment and monotonicity. They also infers implicatives and factives based on *implication signatures* (Nairn et al., 2006) compositionally. But certainty is not considered in their approach. Saurí and Pustejovsky (2012) have proposed a rule-based method using information that can influence the factuality of events such as polarity particles, modality markers, and epistemic predicates. In their algorithm, factuality values of the event, consisting of certainty and polarity, are determined by the upper factuality values and rules, one by one, from the top of the dependency tree. Their model is suitable as it assumes the availability of a factuality lexicon and uses it to identify the factuality status of each subordinate event in a compositional manner from the factuality status of its superordinate event. So we adopt their model and adapt it to the Japanese language to discuss issues regarding lexical knowledge.

3 Japanese factuality analyzer

To discuss the problems about lexical knowledge, we construct a Japanese factuality analyzer based on the lexicon-based compositional approach proposed by Saurí and Pustejovsky (2012). Their analyzer is suitable for analyzing issues because it is based on the availability of a factuality-related simple lexicon and analogous lexicons for Japanese are also available. When we input a result of syntactic parsing to our factuality analyzer, it outputs the factuality of each event.

3.1 Factuality values

Saurí and Pustejovsky characterized a degree of event factuality as a pair of certainty (what is certain *vs* what is only possible) and polarity (positive *vs* negative). They divided the certainty axis into the values *certain* (CT), *probable* (PR), *possible* (PS) and *underspecified* (U), and the polarity axis into *positive* (+), *negative* (−) and *underspecified* (u). For example, an event "*de*" (*leave*) in (1a) is labeled with CT+. This means that it is certain that the event happened or will happen according to the author of the text. In the same way,

Table 1: Our Factuality values

certainty \ polarity	positive (+)	negative (−)
certain (CT)	fact (CT+)	counterfact (CT−)
probable (PR)	probable (PR+)	not probable (PR−)
underspecified (U)	unknown or uncommitted (U)	

Table 2: Example entries of the dictionary of Japanese functional expressions

Sense Category	Expressions	Effects on Factuality	
negation	-nai -nu	polarity:	+ → − − → +
speculation	-daro-u -kamo-shire-nai	certainty:	CT→PR
question	-ka -ka-na	certainty:	CT→U PR→U

Table 3: Example entries of the dictionary of Japanese clue expressions for extended modality

Expression	Tense of Embedded Event	Context Polarity	Factuality
fusegu (prevent)	non-perfective	+	CT−
		−	CT+
wasureru (forget)	non-pefective	+	CT−
		−	CT+
	perfective	+	CT+
		−	CT+

an event "*kaet*" (*go home*) in (1b) is labeled with PR+ and "*hassei-suru*" (*occurrence*) in (1c) is labeled with CT−. We use Saurí and Pustejovsky's factuality values; however, we make some changes to compensate for Japanese sentences.

The first is the distinction between PR and PS. In English, event factuality can be interpreted by specific expressions. For instance, PR is interpreted by *probable* and PS is interpreted by *possible*. However, in Japanese, it is not so straightforward to distinguish between PR and PS due to a diverse variety of modality expressions. Furthermore, PR and PS are minority classes. We therefore combine PR and PS into PR in order to focus on the distinction between certain and uncertain.

The second is underspecified values. Saurí and Pustejovsky used two underspecified values: the partially underspecified CTu and the fully underspecified Uu. For simplification, we do not distinguish two underspecified values. Instead we use U as the underspecified value.

Furthermore, in the present study, we start with focusing only on event factuality attributed to the author of the text. Analyzing factuality for other discourse participants is left for our future work.

We use Saurí and Pustejovsky's factuality values except for these changes. In other words, we divide the certainty axis into the values *certain* (CT), *probable* (PR) and *underspecified* (U), and we also divide the polarity axis into *positive* (+) and *negative* (−). Table 1 shows factuality values by a combination of certainty and polarity.

3.2 Lexical knowledge

In Saurí and Pustejovsky's model, the factuality is analyzed based on lexical knowledge, expressions (called *factuality markers*) that can influence the event factuality. For example, polarity particles of negation, such as the adverb *not*, switch the original polarity of its context, and particles of certainty, such as the auxiliary *may*, change the original certainty of its context. Saurí and Pustejovsky consider not only particles but also predicates. For instance, in the case of the expression *know that*, it presupposes that the event in *that*-clause is fac-

tual. Therefore, the predicate *know* is a factuality marker which changes the factuality of the event in *that*-clause into CT+.

Similarly, in Japanese, some expressions correspond to English factuality markers. We use the dictionary of Japanese functional expressions (Matsuyoshi et al., 2007) and the dictionary of Japanese clue expressions for extended modality (Eguchi et al., 2010) as factuality markers.

The dictionary of Japanese functional expressions is semantically categorized and contains a lot of functional expressions using a hierarchy with nine abstraction levels such as sense and grammatical function. This dictionary includes 341 direction words (16,711 expressions). We can use some categories as factuality markers. Table 2 shows example entries of this dictionary and corresponding effects on factuality. For instance, expressions categorized as speculation, such as "*-daro-u*" (*may*) seen in (1b), change the original certainty of its context. We use 5,345 expressions selected according to categories as factuality markers.

The dictionary of Japanese clue expressions for extended modality contains how predicates influence extended modality of surrounding events. This dictionary includes 8,122 predicates selected from Bunrui Goihyo (National Institute for Japanese Language and Linguistics, 2004). These predicates also relate to the factuality. Therefore, we can use these predicates as factuality markers. Table 3 shows example entries of this dictionary and corresponding factuality. For example, the predicate "*fusei-da*" (*prevented*), seen in (1c), is regarded as the factuality marker that switches the polarity of the preceding event "*hassei-suru*" (*occurrence*).

Dependency tree	Factuality markers	Contextual factuality	**Event factuality**
		CT+ (initial value)	
知らない *shira-nai* (*does not know*)	*-nai* (*not*) polarity: + → −	CT−	*shira* (*know*): CT−
	shira (*know*) certainty: CT polarity: + → −	CT+	
相手は *aite-wa* (*the opponent*)　ことを *koto-wo* (*that*)		CT+	
断念した *dannen-shi-ta* (*had abondoned*)	*dannen-shi* (*abondon*) polarity: + → −	CT+	*dannen-shi* (*abondon*): CT+
		CT−	
彼が *kare-ga* (*he*)　出場を *shutsujou-wo* (*the participation*)		CT−	*shutsujou* (*participation*): CT−

Figure 1: Computing event factuality in (2)

3.3 Algorithm

The factuality analyzer determines an event factuality by propagating a pair of certainty and polarity along a dependency tree from the root of the sentence. The algorithm can reflect dependency between events by the propagation of the factuality. The algorithm determines the factuality of an event based on following components:

Predicates
    The factuality is updated by predicates of its context.

Functional Expressions
    The factuality is updated by functional expressions attached to the event.

Propagated Factuality
    The factuality is determined based on the original factuality of the preceding event.

Figure 1 shows the analysis process when our algorithm is applied to (2). The input is the dependency tree of the sentence (2) (the left side of Figure 1) and the output is the factuality of each event (the right side of Figure 1).

(2)
　　　kare-ga shutsujou-wo dannen-shi-ta-koto-wo aite-wa shira-nai.
　　　(*The opponent does not know that he had abandoned the participation.*)

First of all, the factuality at the top level is set to CT+ as initial value (by the naïve assumption), and the factuality is propagated along a dependency tree from the root of the sentence. The process at each phrase consists of 3 steps.

As a first step, the analyzer updates the contextual factuality if the functional expression is found in the dictionary of Japanese functional expressions. For the first phrase "*shira-nai*" (*does not know*) in this example, the contextual factuality is updated to CT− by the negation "*-nai*" (*not*). As a second step, the factuality value is assigned to every found event. The factuality value CT− is assigned to the event "*shira*" (*know*) in the example. As a third step, the analyzer updates if the predicate is found in the dictionary of Japanese clue expressions for extended modality. In the example, the contextual factuality is updated to CT+ by the factive predicate "*shira*" (*know*). In referring to dictionaries in first and third steps, we adopt simple longest match for the surface. The third step needs to be performed after the second step due to the double nature of predicates, which are both event-denoting expressions and, at the same time, factuality markers.

Similarly, for the phrase "*dannen-shi-ta*" (*had abandoned*), the algorithm outputs CT+ as the factuality of the event "*dannen-shi*" (*abondon*), because of *Propagated Factuality* CT− (the factuality of the preceding event "*shira*" (*know*)), *Predicates* "*shira*" (*know*) (CT− → CT+) and *Functional Expressions* (empty for this case). The analyzer iterates the propagation and updates the con-

Table 4: Correspondence of *actuality* to factuality

certainty \ polarity	+	−
CT	certain+ certain− → certain+	certain− certain+ → certain−
PR	probable+ probable− → probable+	probable− probable+ → probable−
U	unknown	

textual factuality. As a result, CT− as the factuality of the event "*shira*" (*know*), CT+ as the factuality of the event "*dannen-shi*" (*abandon*), and CT− as the factuality of the event "*shutsujou*" (*participation*) are obtained.

4 Findings from empirical evaluation

4.1 Data and experimental setup

We apply our algorithm to 6,404 sentences on the Yahoo! Japan Q&A section for the Japanese corpus with extended modality (Matsuyoshi et al., 2010). These sentences are included in the *Balanced Corpus of Contemporary Written Japanese* (BCCWJ)[1], and each event mention is labeled with extended modality (*source, time, conditional, primary modality type, actuality, evaluation*, and *focus*). *Actuality* denotes the degree of certainty and corresponds to our factuality. Table 2 shows the correspondence of *actuality* to our factuality.

In this experiment, we apply our algorithm to 11,395 event mentions, where *source* is "wr" (writer of the sentence). These event mentions are also selected by part-of-speech, such as verb and adjective. For the identification of the event mention, we give the gold data to the analyzer because we discuss only about lexical knowledge.

If the analyzer makes an error in regards to the factuality of an event, then this error will have an influence on the factuality of the next event, because the analyzer propagates the updated factuality to the next event. Our intent for this experiment is not to analyze this kind of error. Therefore, we use the gold label as *Propagated Factuality* in order to prevent the error propagation.

4.2 Discussion

We discuss issues about lexical knowledge through the error analysis of the analyzer based on lexical knowledge and compositionality. Our algorithm computes the event factuality based on *Predicates, Functional Expressions* and *Propagated Factuality*, but for matrix clauses, it determines the factuality based only on *Functional Expressions*. We expect issues to arise for func-

Table 5: Accuracy for each case

	Matrix clauses	Subordinate clauses	Total
Correct	3,529	3,652	7,181
Wrong	693	3,521	4,214
Accuracy	0.836	0.509	0.630

Table 6: Confusion matrix for the certainty axis at matrix clauses

gold \ system	CT	PR	U	Total	Recall
CT	**2,478**	47	230	2,755	0.899
PR	145	**63**	50	258	0.244
U	151	11	**1,047**	1,209	0.866
Total	2,774	121	1,327	4,222	
Precision	0.893	0.521	0.789		

Table 7: Confusion matrix for the polarity axis at matrix clauses

gold \ system	+	−	Total	Recall
+	**2,374**	59	2,433	0.976
−	7	**293**	300	0.977
Total	2,381	352	2,733	
Precision	0.997	0.832		

Table 8: Confusion matrix for the certainty axis at subordinate clauses

gold \ system	CT	PR	U	Total	Recall
CT	**3,335**	330	1,997	5,662	0.589
PR	245	**104**	175	524	0.198
U	329	41	**617**	987	0.625
Total	3,909	475	2,789	7,173	
Precision	0.853	0.219	0.221		

Table 9: Confusion matrix for the certainty axis at subordinate clauses

gold \ system	+	−	Total	Recall
+	**3,224**	434	3,658	0.881
−	55	**301**	356	0.846
Total	3,279	735	4,014	
Precision	0.983	0.410		

tional expressions at matrix clauses. At subordinate clauses, on the other hand, we expect complex issues involving multiple components. We therefore analyze both the issues at matrix clauses and the issues at subordinate clauses, respectively.

Table 5 shows accuracy and Tables 6-9 show each confusion matrices for the certainty axis and the polarity axis for each case. These tables show that minority classes PR and U are difficult on the certainty axis. On the polarity axis, we obtain relatively high accuracy. Comparing matrix clauses to subordinate clauses, accuracy at subordinate clauses, which is based on some components, is lower than the accuracy at matrix clauses, which is based only on functional expressions. For each minority label (PR and U on the certainty axis, and − on the polarity axis), subordinate clauses have lower precision relative to matrix clauses. One reason for this is that we do not consider the scope of negation and speculation.

Table 10 shows the error type distribution. At

Table 10: Error type distribution

	Analyzed errors	Error type		Errors
Matrix clauses	108	functional expressions	semantic ambiguity	102
			insufficient coverage	4
		others		2
Subordinate clauses	1,041	functional expressions	semantic ambiguity	412
			insufficient coverage	16
		predicates	semantic ambiguity	4
			insufficient coverage	34
		scope		656

matrix clauses, the issue regarding functional expressions is found for 106 errors when analyzing 108 errors, and the rest of errors are due to an adverb and the parsing error. At subordinate clauses, we analyze 1,041 errors. Issues regarding the functional expressions (428 errors), predicates (38 errors), and the scope (656 errors) are found. Some errors are due to multiple issues. In the following paragraphs, we describe these issues in detail.

4.2.1 Functional expressions

Out of the 106 errors for functional expressions, 53 false-positive errors regarding U were most common. Almost all of these errors are due to semantic ambiguity for functional expressions.

(3)
  shira-nai-no-mo *fushigi-de-wa-nai-desu.*
  (*It is no wonder that he doesn't know.*)
  (Gold: CT−, System output: U)

(3) is an example for semantic ambiguity of the functional expressions. Our analyzer refers to dictionaries by simple longest match. Therefore, the factuality of the event "*fushigi*" (*wonder*) is wrongly assigned as U because "*-de-wa*" is recognized as a recommendation (*how about*). In this context, the expression "*-de-wa*" is a part of inflection. So it has no special meaning.

As seen above, semantic ambiguity for functional expressions is a critical problem for Japanese factuality analysis. But disambiguation of Japanese functional expressions is not simple. Some previous work is engaged on this task, such as Tanaka et al. (2013). They construct MCN corpus for the disambiguation of expressions related to factuality. It is important to import this line of prior work to our analyzer.

Coverage for the dictionary of Japanese functional expressions also becomes a problem. However, the number of problems contains only 4 errors. We find that coverage for the dictionary of Japanese functional expressions is sufficient.

4.2.2 Predicates

At subordinate clauses, 38 errors arise which are caused by predicate issues. 34 of the 38 errors are due to insufficient coverage for predicates and the other 4 errors are due to semantic ambiguity for predicates.

(4)
  tadashii koto-wo kakunin-shi-te-kudasai.
  (*Please check that it is correct.*)
  (Gold: CT+, System output: U)

(4) is an example of insufficient coverage for predicates. In (4), our algorithm assigns U as the factuality of the event "*tadashii*" (*correct*) because U (the factuality of the event "*kakunin-shi*" (*check*), which is influenced by the request expression "*kudasai*" (*please*)) is propagated without any update. However, the predicate "*kakunin-shi*" (*check*) presupposes that the preceding context is factual, so it should be assigned CT+ as the factuality of the event "*tadashii*" (*correct*). This incorrect assignment occurs because that predicate does not exist in the dictionary of Japanese clue expressions for extended modality.

Out of 1,041 errors at subordinate clauses, 417 events are that predicates in the dictionary of Japanese clue expressions for extended modality are used. Only 4 errors, however, are due to semantic ambiguity for predicates. We therefore find that semantic ambiguity for predicates poses little problem. Furthermore, we focus on correct events by predicates. Out of the 1,128 correct instances in the area analyzed by the corpus, 351 are correct by predicates in the dictionary. In contrast to this, only 34 errors are due to insufficient coverage for predicates. For this reason, we find that insufficient coverage for predicates is a small issue.

4.2.3 Scope

In Section 3, we described that our analyzer determines the event factuality based on three components: *Predicates, Functional Expressions*, and *Propagated Factuality*. However, we find that it

is crucial to determine boundaries whether the analyzer should propagate the factuality. In other words, it should resolve the scope of negation and speculation though the actual analyzer regards all embedded contexts as the scope. The errors due to the scope, in fact, are the majority of errors at subordinate clauses (656/1,041).

(5)

    sukoshi kougai-ni deru-to onsei-ga kikitore-mase-n.
    (*I cannot hear the voice if I leave the suburbs.*)
    (Gold: CT+, System output: CT−)

Our algorithm wrongly assigns − as the polarity of the event "*deru*" (*leave*) in (5). This is because − (the polarity of the event "*kikitore*" (*hear*), which is influenced by the negation "*-n*" (*cannot*)) is propagated with no update. The negation "*-n*" (*cannot*) denies only the event "*kikitore*" (*hear*) but not the event "*deru*" (*leave*). As exemplified, the issue regarding the scope of negation and speculation is very crucial.

Of the 233 events where the analyzer outputs − as the polarity and the gold *Propagated Factuality* is −, 28 events are correct for the polarity, whereas 112 events are errors due to the scope. As shown, there are many cases where the analyzer should not propagate due to scope, and there are also many cases where the analyzer should propagate as −. We find that resolving the scope is a significant, but difficult challenge.

Next, we focus on the conjunction particles, such as "*-to*" in (5), as the key to detect scope in practice. Out of 656 errors due to the scope, the conjunction particle "*-to*" follows 126 events, "*-ga*" follows 78 events, "*-te*" follows 70 events, and so on. Therefore, when we detect scope in practice, we assume to use conjunctive particles as the key to determine propagation boundaries. In the next section, we investigate scope detection based on such expressions.

5 Lexicon-based scope detection

In the previous section, we found that detecting a scope is very crucial. In this section, we investigate the limitation of the lexical knowledge for a scope and identify the technical research issues more precisely through experiments for rule-based scope detection.

5.1 Related work for scope detection

In recent years, the detection of negation and speculation scopes is intensively being research for English (Szarvas et al., 2008; Apostolova et al., 2011), such as Shared Task in CoNLL-2010 (Farkas et al., 2010) and *SEM 2012 (Morante and Blanco, 2012). For example, the BioScope corpus (Szarvas et al., 2008) is annotated with negation and modality expressions with their scope, and it is extensively used for resolution of the scope. However, studies for the detection for scope are insufficient for Japanese. Detection of scope in Japanese is a significant challenge, and will be highly beneficial for Japanese factuality analysis.

5.2 Knowledge-based scope detection

We take a rule-based scope detection approach to block propagating a contextual factuality. Before the first step on each phrase as described in Section 3, this approach blocks the propagation when the specific expressions are found in the event. The approach then assigns the contextual factuality as initial value CT+ and restarts the propagation. When such expressions are not found, the propagation is not blocked.

We used the terms shown in Table 11 to detect such expressions. When one of the terms appears at the end of an event, the event blocks the propagation. The terms are categorized by Minami (1974) according to the intensities of the constructing subordinate clauses: A is high, C is low and B is intermediate. These intensities would be used as a tendency of blocking the propagation. However, because there are some ambiguities such as " " (*-te*) which belongs to all categories, we used all terms to detect scope and block the propagation.

5.3 Results

Table 12 shows the experimental results with/without lexical knowledge for scope. In the previous experiments as described in Section 4, in order to avoid the propagation error, we used gold contextual factuality. However, in our experiment, we focus on the propagation, so we do not use gold contextual factuality.

Table 12 shows that F_1-score increases 19.2% (0.112) by adding lexical knowledge. Focusing on each labels, our approach had no negative effect except recall of U. This means that our approach based on lexical knowledge works well, especially for minor labels. However, some errors still remain.

5.4 Remaining issues

We identify the remaining issues through the error analysis of the result. We focus on the

Table 11: Expressions to prevent propagating a contextual factuality

Category	Expressions						
A	(*-nagara*),	(*-tsutsu*),	(*-te*),	(*-de*)			
B	(*-te*), (*-to*),	(*-nagara*),	(*-no-de*),	(*-no-ni*),	(*-ba*),	(*-tara*),	
	(*-nara*),	(*-te-mo*),	(*-te*),	(*-zu*),	(*-zu-ni*),	(*-nai-de*)	
C	(*-ga*),	(*-kara*),	(*-keredo*),	(*-keredo-mo*),	(*-kedo-mo*),		
		(*-kedo*),	(*-shi*),	(*-te*)			

Table 12: Performance with/without lexical knowledge for scope

		CT+	PR+	PR−	CT−	U	Micro-Average	Macro-Average
The number of events		7,569	678	104	848	2,196	11,395	
With lexical knowledge	Precision	0.850	0.372	0.123	0.605	0.455	0.696	0.481
	Recall	0.753	0.178	0.067	0.672	0.697	0.696	0.474
	F_1	0.799	0.241	0.087	0.637	0.551	0.696	0.463
Without lexical knowledge	Precision	0.850	0.321	0.060	0.451	0.348	0.584	0.406
	Recall	0.584	0.156	0.048	0.542	0.756	0.584	0.417
	F_1	0.692	0.210	0.053	0.492	0.477	0.584	0.385

events which have a propagated factuality; in other words, it is not the last event of the sentence. In addition, the events whose propagated factuality is CT+ are also excluded from the analysis target, because when a CT+ is propagated to an event, even if the event blocks or doesn't block the CT+, the propagated factuality to the first step is CT+.

There are 1,739 events which satisfy the above conditions and we apply the block rule to 925 of them (i.e. some terms in Table 11 are found in the events). Table 13 shows the changes in the number of correct and incorrect results by adding the lexical knowledge. When using the block rule, 553 out of 925 incorrect events become correct. On the other hand, 100 of the correct events became incorrect. This suggests some ambiguities of expressions caused too much blocking.

(6) a.
　　shikaku-wo umaku ikashi-te hataraku koto-ga deki-nakat-ta.
　　(*I could not work by making best of my qualification.*)

　　b.
　　ima-wa shojijou-ga at-te rikon-deki-nai.
　　(*I cannot get a divorce because I have various reasons.*)

For example, "　" (*-te*) in (6a) causes blocking but in (6b) should not cause blocking.

Focusing on the coverage of the lexical knowledge, as described in Section 4, there are 656 errors due to the error of scope detection. 402 of them do not have CT+ as the propagated factuality and all of them should block the factuality propagation. However, only 229 of 402 blocked the propagation. This shows that the coverage of the lexical knowledge is still limited.

(7)
　　hantoshi-mae-no tenken-de-wa ijou-ga mi-rare-nakat-

Table 13: Result changes by adding lexical knowledge

		with	
		correct	wrong
without	correct	49	100
	wrong	553	223

ta.
(*There are no defect in checking half a year ago.*)

For example, "　" (*-de-wa*) in (7) is not covered in this lexical knowledge.

6 Conclusion

We described Japanese factuality analysis, which is useful for information extraction and textual entailment recognition, among others. We discussed issues regarding lexical knowledge through error analysis by using a Japanese factuality analyzer based on lexical knowledge and compositionality. As a result, coverage of existing lexical resources is sufficient but issues regarding the semantic ambiguity of functional expressions and issues regarding scope were found. In particular, it was revealed that the problem regarding scope is most significant. We therefore performed an additional experiment with lexical knowledge for scope and discussed its helpfulness. However, the issue regarding scope includes the issue by profound meaning and context. Therefore, we consider that this issue is high-priority challenge.

In the future, we will address these challenges toward a high-performance Japanese factuality analyzer with other lexical knowledge and linguistic phenomena. Furthermore, we aim to construct a Japanese modality analyzer through the extension of the framework for factuality.

594

Acknowledgement

References

Emilia Apostolova, Noriko Tomuro, and Dina Demner-Fushman. 2011. Automatic extraction of lexico-syntactic patterns for detection of negation and speculation scopes. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers-Volume 2*, pages 283–287.

Marie-Catherine de Marneffe, Christopher D. Manning, and Christopher Potts. 2012. Did it happen? the pragmatic complexity of veridicality assessment. *Computational Linguistics*, 38(2):301–333.

Megumi Eguchi, Suguru Matsuyoshi, Chitose Sao, Kentaro Inui, and Yuji Matumoto. 2010. An analyzer of modality, actuality and valuation of events in japanese. In *Proceedings of the 16th Annual Meeting of Natural Language Processing (in Japanese)*, pages 852–855.

Richárd Farkas, Veronika Vincze, György Móra, János. Csirik, and György Szarvas. 2010. The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In *Proceedings of the Fourteenth Conference on Computational Natural Language Learning — Shared Task*, pages 1–12.

Kentaro Inui, Shuya Abe, Kazuo Hara, Hiraku Morita, Chitose Sao, Megumi Eguchi, Asuka Sumida, Koji Murakami, and Suguru Matsuyoshi. 2008. Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents. In *the 2008 IEEE/WIC/ACM International Conference on Web Intelligence*, pages 314–321.

Bill MacCartney and Christopher D. Manning. 2009. An extended model of natural logic. In *Proceedings of the Eighth International Conference on Computational Semantics (IWCS-8)*.

Suguru Matsuyoshi, Satoshi Sato, and Takehito Utsuro. 2007. A dictionary of japanese functional expressions with hierarchical organization. *Journal of Natural Language Processing (in Japanese)*, 14:123–146.

Suguru Matsuyoshi, Megumi Eguchi, Chitose Sao, Koji Murakami, Kentaro Inui, and Yuji Matsumoto. 2010. Annotating event mentions in text with modality, focus, and source information. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, pages 1456–1463.

Fujio Minami. 1974. *Structure of modern Japanse (in Japanese)*. Taishukan Shoten.

Roser Morante and Eduardo Blanco. 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*, pages 265–274.

Rowan Nairn, Cleo Condoravdi, and Lauri Karttunen. 2006. Computing relative polarity for textual inference. In *Proceedings of Inference in Computational Semantics (ICoS-5)*.

National Institute for Japanese Language and Linguistics. 2004. *Bunrui Goihyo (Word List by Semantic Principles)*. Dainippon-tosho.

Roser Saurí and James Pustejovsky. 2009. FactBank: a corpus annotated with event factuality. *Language resources and evaluation*, 43(3):227–268.

Roser Saurí and James Pustejovsky. 2012. Are you sure that this happened? assessing the factuality degree of events in text. *Computational Linguistics*, 38(2):261–299.

György Szarvas, Veronika Vincze, Richárd Farkas, and J. Csirik. 2008. The BioScope corpus: annotation for negation, uncertainty and their scope in biomedical texts. In *Proceedings of the Workshop on Current Trends in Biomedical Natural Language Processing*, pages 38–45.

Ribeka Tanaka, Daisuke Bekki, and Ai Kawazoe. 2013. MCN corpus: The design and operation of the guidelines. In *Proceedings of the 19th Annual Meeting of Natural Language Processing (in Japanese)*, pages 77–80.