Expanded Dependency Structure based Textual Entailment Recognition System of NTTDATA for NTCIR10-RITE2

Megumi Ohki
NTT DATA Corporation
Research and Development Headquarters
ookimg@nttdata.co.jp

Takashi Suenaga
NTT DATA Corporation
Research and Development Headquarters
suenagatk@nttdata.co.jp

Daisuke Satoh
NTT DATA Corporation
Research and Development Headquarters
satodic@nttdata.co.jp

Yuji Nomura
NTT DATA Corporation
Research and Development Headquarters
nomurayu@nttdata.co.jp

Toru Takaki
NTTDATA Corporation
Research and Development Headquarters
takakit@nttdata.co.jp

ABSTRACT

This paper describes NTT DATA's recognizing textual entailment(RTE) systems for NTCIR10 RITE2. We participate in four Japanese tasks, BC Subtask, Unit Test, Exam BC and Exam Search[5].

Our approach uses a ratio with the same semantic relations between words. It is necessary to recognize two semantic viewpoints, which are the semantic relation and the meaning between words in a sentence, in order to recognize textual entailment. We divide the methods into **the semantic dependency relation between words** and **the meaning between words** for recognizing textual entailment. In this paper, we present our system using methods for recognizing semantic relations using expanded dependency structures.

Team Name / ID

NTTDRDH/NTTD

Languages / Subtasks

Japanese/Binary-class (BC), Unit Tests (Unit), ExamBC-class (Exam BC), Exam Search (Exam Search)

Keywords

semantic relations between words, meanings between words, expanded dependency structure

1. INTRODUCTION

It is necessary to discern various language phenomena to recognize textual entailment. But discerning a language phenomenon is not easy. In most of the existing research, the researchers have proposed methods of classifying language phenomena by extracting the feature of each sentence. They have been able to partially discern this language phenomenon. Although techniques of classification using this feature extraction have had a certain amount of success, we cannot understand why performance improves in many cases. The various features are used to classify types of textual entailment in the existing research. In order to classify the information to a high dimension, it is unknown which feature contributed to the classification in fact.

Our aims are to make clear the type of textual entailment that can be recognized even if it does not depend on the classification by feature extraction, and examine problems with RTE. We provide a technique of recognizing textual entailment using patterns of structure of a sentence instead of classification by feature extraction. The semantic relation between words makes the structure of a sentence. Our proposal is to divide such **semantic relation** into **meanings of words** and **semantic dependency relations**, and our system can recognize textual entailment by using entailment of meanings and semantic dependency relations between words.

An example of our approach in Japanese is ”りんごを買いに行った” (I went to buy an apple), shown in H1 in Figure 1. This sentence consists of the words, ”りんご” (apple), ”買う” (to buy), and ”行った” (went), and the relationship between these words is expressed by the postpositional particles in the dependency structure. ”買う” (to buy) is related to ”りんご” (apple), ”行った” (went) is related to ”買う” (to buy) in the case of H1. There is no relation between ”行った” (went) and ”りんご” (apple).

Entailing the meaning of words can be done by using means synonymous words and synonymous expressions, and entailing semantic dependency relations can be done by using modes of expression, such as swapping a subject with a predicate and the passive and the active voice. For example, in the entailment relation of H1 and T1 in Figure 1, **”りんご” (apple)** and **”果物” (fruit)** have a hyponymical relation. On the other hand, in the example of H2 and T2, there is difference dependency relation between ”私はりんごを買った” (I bought an apple) in T2 and ”私が買ったりんご” (an apple which I bought) in H2. However they have the same agent and patient word, and the two sentences have an entailment relation.

Thus, the semantic relation for the two sentences is considered to be a combination of the meaning and the semantic dependency between words in a sentence. We aim to establish a technique for recognizing textual entailment between sentences. In this paper, we aim to extract the semantic dependency between words especially. Although there are various types of semantic dependency, such as the relation between an agent and a patient or modifying relation, we examine first of all only the directivity of a semantic rela-

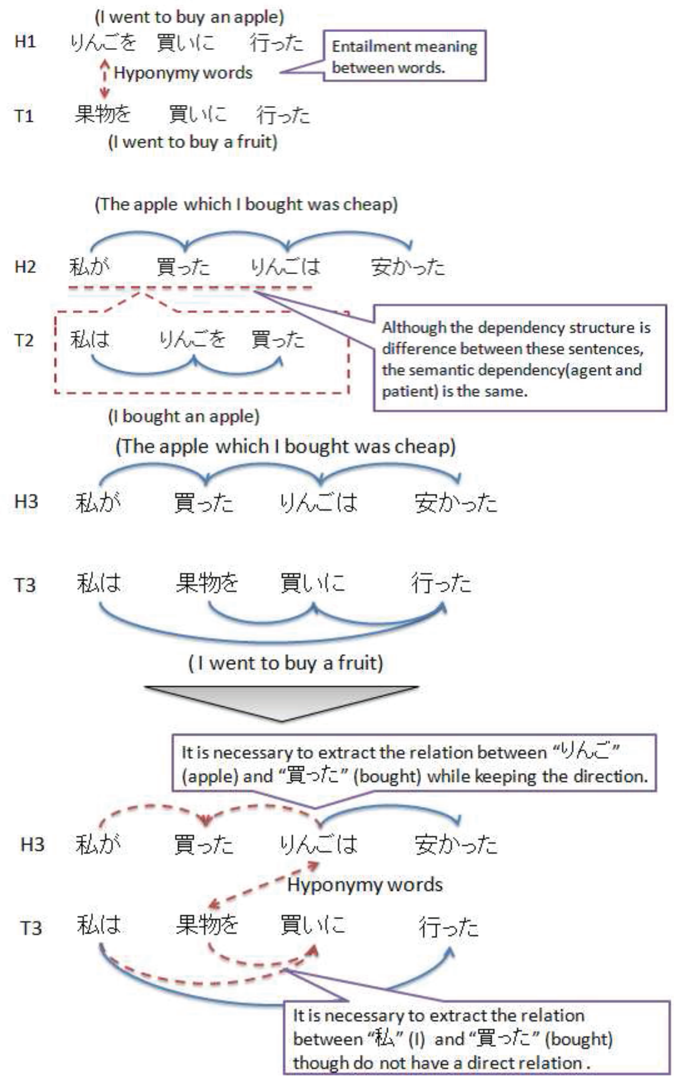

Figure 1: Example of textual entailment in consideration of the semantic relation between words

tion.

This means the relation of dependence, such as an agent and a patient, the thing to modify and the thing modified. Most of the semantic dependency between words can use a simple dependency structure as a relation between an agent and a patient. However, some sentences cannot extract a relation between an agent and a predicate by only simple dependency structure because Japanese has flexibility.

We describe these sentence with H3 and T3 in Figure 1.

H3 is a sentence containing two meanings that are "私がりんごを買った" (I bought an apple) and "りんごが安かった" (the apple was cheap), and entails the meaning of "私は果物を買いに行った" (I went to buy a fruit) in T3. Although the agent in H3 is "りんご" (apple) to "安い" (cheap), and "私" (I) to the act of "買う" (buy).

Moreover, "りんご" (apple) is a patient to "買う" (buy). It is unacquirable from a simple dependency structure in which "りんご" (apple) is a patient. There are some cases where it is necessary to extract the differing direction of dependency relation from a simple dependency structure.

On the other hand, "私" (I) is a subject and "買いに行く" (go to buy it) shows action. "私" (I) is regarded only an agent to "行く" (go) if a dependency relation is only gained. It is necessary to gain the relation which does not have a dependency relation directly at this time. If the dependency of the dotted line shown in H3 and T3 of Figure 1 can be extracted, this semantic dependency is the same. Therefore our system using this extracting semantic dependency can recognize that it is entailment.

In order to solve the semantic dependency between words which cannot use a simple dependency relation in this paper, we propose a method of acquiring the dependency between the words which have an extended dependency structure.

We analyze examples which should take into consideration the directivity between the words which do not have a dependency structure directly, and extract the dependency structure pattern for dependency extension. The dependency between words is built using this semantic dependency pattern.

2. CLASSIFICATION OF TEXTUAL ENTAILMENT

We analyzed data from the following three viewpoints, in order to grasp how many entailment relations can be recognized only by the meanings and semantic dependency between words.

1. The entailment relation only in consideration of the meaning of words
2. The entailment relation only in consideration of the semantic dependency between words
3. The entailment relation which needs other information

We use the Unit Test data[5]. Furthermore, in order to grasp what kind of language phenomenon there is in the meaning and dependency between words, we classify the language phenomenon into subclasses based on the information on Unit Test data. The result is shown in Table 1.

Of course, there are many entailment relations that must be recognized by combination of some language phenomena. We will aim to devise a solution which takes into consideration a combination of the meaning and semantic dependency between words in the future. However, our first aim is to investigate a method that can solve a single problem first of all, and so we analyze language phenomena in a textual entailment relation using the meaning and semantic dependency between Unit Test data.

The normal semantic dependency between words can be solved by looking at the corresponding dependency structures. Examples are shown in Table 1 A1.

On the other hand, a sentence pair with a contradict relation has completely different semantic dependencies which are gained from a simple dependency structure (such as A2) or there is too much semantic dependency in t2 than t1(such as A3). T2 entails T1 in A3, and it is called "retrorse entailment." Some of the dependency between the words cannot be solved only by a simple dependency structure because there are various types of Japanese grammar.

Examples are shown in B4 from B1 of Table 1. A sentence may express the same meaning as another sentence but use different expressions that rephrase the passivity and activity voices and swap the subject and other words. As a result of analyzing data, we find the following four types which cannot be shown only according to a simple dependency structure.

1. is-a relations
2. modifying and modified relations
3. verb pair relations in complex sentences
4. parallel entry relations

When a word pair has an is-a relation, the relation between a subject and a predicate is swapped. Since this subject and predicate have an is-a relation, this sentence has the same meaning even if both sides are swapped.

In a sentence which has a modifying and modified relation, the modifying clause and modified clause correspond. For instance, t1 says "Alfred Bernhard Nobel was the inventor of dynamite" as a modifying clause, and t2 explains it as a modified clause in B2. Usually modifying clauses contain verbs and modified clauses are noun clauses. It can be said that these have the same meaning but use different expressions.

Moreover, the relation between the verbs in a complex sentence cannot be used to extract the relation between subject and verbs by only simple dependency structures since two or more verbs in a sentence exist.

A parallel entry relation exists in a sentence which indicates two or more matters by illustration. When a sentence uses illustrating markers which are "と" (and) and "および" (both A and), a comma, the words before and after the markers have the same implications. However, these relations are also unacquirable by a simple dependency structure. We aimed to extract these unacquirable semantic dependencies.

Moreover, we divided instances of meanings relation between words into three types. One type can recognize textual entailment by the same and different semantic relations between words such as C1-C5 in Table 1. The other type can do so by examining the existence of particular words such as D1-D2, and another type needs filling pronouns such as E1-E2. Our system recognizes synonymous words, entailment words and the existence of negative expressions. With entailment words, the meaning of one word entails the meaning of another word.

Other entailment relations need more information, which are implicit information, paraphrasing of expression and inference. These need to grasp the various modes of Japanese expression and require background knowledge. In order to recognize these, it is necessary to examine another solution. But that lies outside the scope of this paper and we do not perform special correspondence here.

3. OUR PROPOSING METHODS FOR RECOGNIZING TEXTUAL ENTAILMENT

In this section, we explain the method of acquiring the semantic dependency and meanings between words, and we also explain the method of judging entailment. Henceforth, the words in a semantic dependency are expressed as [**the depending word, the depended word**].

3.1 Entailment decision method

Before explaining our method, we explain the method of judging entailment in our recognition technique. Our data analysis showed that the probability of entailment is high, if t1 has more of the same semantic dependencies of each word of t2. Our method uses the ratio of correspondence of the semantic dependency between words as a certain factor for the technique of entailment.

3.1.1 Method of acquiring the dependency between words

We showed that the dependency between words is generally acquirable by using simple dependency structures in section 2. On the other hand, various expressions may be used in sentences which express the same meaning since Japanese is flexible. Therefore, in some cases we cannot correctly extract the semantic dependency only using a simple dependency relation.

In order to solve this problem, we devised the method to acquire the semantic dependency by extended dependency structures. This section shows both the method of acquiring the semantic dependency by simple dependency structures and expanded dependency structures.

3.1.2 System of acquiring dependency by simple dependency structures

Two words in a clause which has a dependency structure are extracted as a pair of semantic dependencies. The procedure is shown below.

1. Analyze simple dependencies in a sentence. We use CaboCha[4].

2. Acquire dependency structures each clause, the original form and part-of-speech information.

3. Delete the unnecessary part-of-speech information within the dependency structur in each clause.

   We set up "助詞" (particle), "助動詞" (auxiliary verb), "動詞-非自立" (verb-not independent), "動詞-接尾" (verb-suffix), "名詞-非自立" (noun-not independent) and "記号"(sign) as unnecessary part-of-speech information based on the result of analyzing data.

   However, there are expressional non-delete words, one word is "ない" (not) which is useful for extracting negative information; another word is "だ" (be) which is

Kind of RTE	Number	type	Y	N	t1	t2
semantic dependency	A1	simple comparison of dependency relations	70	0	親の投資とは、進化生物学では、一人の子の利益のために親が支払うあらゆる資源を指す。	進化生物学では、親の投資とは、一人の子の利益のために親が支払うあらゆる資源を指す。
	A2	different semantic dependencies	0	17	三遊亭花圓遊は佐々木安太郎という本名である。	三遊亭 花圓遊は、落語家である。
	A3	excess of semantic dependency	0	5	入谷明は近畿大学生物理工学研究科教授である。	入谷 明は日本の畜産学者で、京都大学名誉教授、近畿大学生物理工学研究科教授である。
	B1	is-a relation	11	0	国際チェス連盟の会長はキルサン・イリュムジーノフである。	キルサン・イリュムジーノフは、国際チェス連盟の会長である。
	B2	modifying and modified relations	21	0	ダイナマイトの発明者として知られるアルフレッド・ベルンハルド・ノーベルは、ノーベル賞の提唱者である。	アルフレッド・ベルンハルド・ノーベルはダイナマイトの発明者として知られ、ノーベル賞の提唱者である。
	B3	verb pair relations in complex sentences	16	0	カップめんの容器に発泡スチロールを使うのは、保温性が良く、持ったときに熱くないという理由からである。	カップめんの容器に発泡スチロールを使うのは、保温性が良いという理由からである。
	B4	parallel entry relations	19	0	PL 学園高等学校のほか、中学校、小学校、幼稚園、専門学校もパーフェクト リバティー教団教団本部敷地内にある。	PL 学園高等学校のほか、中学校もパーフェクト リバティー教団教団本部敷地内にある。
meaning between words	C1	synonymous word	5	0	コニャックは、フランスのコニャック周辺の特定の地域で生産されなければならないブランデーである。	コニャックは、フランスのコニャック周辺の特定の地域で産出されなければならないブランデーである。
	C2	entailment word	8	0	PL 学園中学校・高等学校はパーフェクト リバティー教団教団本部敷地内にある。	PL 学園中学校・高等学校はパーフェクト リバティー教団教団本部敷地内に立地している。
	C3	hyponymy word	7	1	菱川師宣の代表作として有名な『見返り美人図』は、肉筆浮世絵である。	菱川師宣の代表作として有名な『見返り美人図』は肉筆画である。
	C4	partitive word	1	0	明治憲法において、皇室と国との区別は認められていた。	明治憲法において、天皇と国との区別は認められていた。
	C5	difference of word meaning	0	5	スコティッシュ・テリアは犬の一品種である。	スコティッシュ・フォールドは、猫の一品種である。
	D1	existence of negative expression	0	4	鉄芽球性貧血とは、十分な鉄を摂取している場合に起こる貧血だ。	鉄芽球性貧血とは、十分な鉄を摂取していない場合に起こる貧血だ。
	D2	existence of particular information	0	1	コニャックは、フランスのコニャック周辺の特定の地域で産出されなければならないブランデーである。	コニャック は、フランスのコニャック周辺で産出されるブランデーである。
other	E1	pronouns	3	0	大衆向け飲み物の種類には、アイソトニック（栄養学で、浸透圧が血液のそれに近い）がある。	大衆向け飲み物の種類には、アイソトニック（栄養学で、飲み物の浸透圧が血液の浸透圧に近い）がある。
	E2	zero-pronouns	0	3	コニャックは、コニャック周辺の特定の地域で生産されなければならない。	コニャックは、フランスのコニャック周辺の特定の地域で生産されなければならない。
	F1	implicit information	20	0	『二百三高地』は、日露戦争における、203 高地の日露両軍の攻防戦を描いた。	『二百三高地』は、日露戦争における、203 高地の日露両軍の攻防戦を描いた作品。
	F2	rephrase	39	0	玄翁和尚が殺生石を破壊し、殺生石のかけらは全国 3ヶ所の高田と呼ばれる地に飛散したと伝わる。	玄翁和尚が殺生石を破壊し、殺生石のかけらは各地へと飛散したと伝わる。
	F3	inference	7	0	佐賀藩は、蒸気軍艦「電流丸」を発注するなどしていた。	佐賀藩は、蒸気軍艦「電流丸」を所有していた。

Table 1: Classification of textual entailment recognition

expression of assertion; and yet another word is "いる" (exist) which express condition.

We describe this procedure in T3 in Figure 1. In T3, "私は" (I) and "買いに" (to buy) depend on "行った" (went), and "果物を" (fruit) depends on "買いに" (to buy). At this time, pairs of original forms are created only by the target which is the content word. Then, three semantic dependency pairs are extracted, which are [私 (me), 行く (go)], [買う (buy), 行く (go)] and [果物 (fruit), 買う (buy)]. These pairs keep the order of starts and ends for dependency since the starting word of dependency has an influence on the ending word of dependency.

Moreover, the word in negative expression, which changes the meaning of the sentence by being used simultaneously with other words, is given to semantic dependency pairs. For example, if T3 is "私が果物を買いに行かなかった (I did not go to buy fruit)", semantic dependency pairs are created such as [私 (I), 行く-ない (go-not)],[買う (buy), 行く-ない (go-not)] and [果物 (fruit), 買う (buy)].

3.1.3 Methods of acquiring dependency by dependency expansion

The previous section explained the method of acquiring the semantic dependency by using simple dependency structures.

However, since the word order in a Japanese sentence can be swapped easily, there are some cases where the order of semantic dependency changes. Moreover, although two or more verbs exist for a subject in a complex sentence, the end of dependency is decided to be only one with a dependency analysis system. Therefore, some pairs of semantic dependency cannot be created.

We consider whether an expended dependency structure can extract an item which cannot be shown only by simple dependency structures, and devise the method and patterns of acquiring the semantic dependency based on the result of analysis instances which have semantic dependencies between words. These patterns are shown in Table 2.

Is-a relation

When a noun clause containing the particle "が" (ga) or "は" (ha) which shows a subject depends on a predicate clause, the relation between the word of the noun clause and the word of a predicate clause has an is-a relation; that is, there is a relation which depends on both sides. For this reason, the is-a relation does not arise directivity in a semantic dependency.

Although the semantic dependency acquirable by simple dependency structures serves as [沢田研二 (Sawada Kenji), 歌手 (a singer)] in G1, the semantic dependency which transposes the depending and depended word such as [歌手 (a singer), 沢田研二 (Sawada Kenji)] by an extended is-a relation is newly acquirable.

Modify and modified relation

There is a problem with modification with a phrase: being unable to take the right order of the semantic dependency between words. When a verb phrase modifies a noun especially and only a dependency relation is used, a semantic dependency is created in the order of [動詞 (verb), 名詞 (noun)].

However, it is desirable to structure a dependency in the order of [名詞 (noun), 動詞 (verb)] because the noun performs a subject essentially. For example, in "歌手である沢田研二" (Sawada Kenji who is a singer) in G2, when a dependency relation is acquired simply, this semantic dependency is created [歌手 (a singer), 沢田研二 (Sawada Kenji)].

However, as extension by modifying and modified relations, when a verb phrase depends on a noun, the dependency in the opposite direction is also created. Then, the dependency which was not able to be acquired by the simple dependency structure of [沢田研二 (Sawada Kenji), 歌手 (a singer)] can be structured.

Verb pair relation in complex sentence

In the case of a complex sentence, two or more verbs exist for one certain subject. However, two or more ends of dependency cannot be specified with the present dependency analysis system.

In a complex sentence, the verb in the previous clause usually depends on the verb in the following clause. Semantic dependencies in a complex sentence are newly acquired by these simple dependencies between verbs.

For example, our system can acquire [ムネーモシュネー (MnĂŞmosynĂŞ), 添う-臥す (accompany-lie)] and [添う-臥す (accompany), 産む-だ (bear-done)] if simple dependencies are used in G3, but it cannot grasp a relation with the subject of "産む-だ" (bear-done). Verbs with a semantic dependency have the same subject; therefore, the noun in [名詞 (noun), 依存する動詞 (the depending verb)] is treated as a subject. In this way, [ムネーモシュネー (MnĂŞmosynĂŞ), 産む-だ (bear-done)] are acquirable.

Parallel entry relation

Two patterns exist in the relation of a parallel entry. They are a pattern using a word showing a parallel structure and having commas showing a parallel structure.

Present dependency analysis system analyze the dependency structure between the words currently written together, and it cannot specify correctly two or more starts of dependency. However, a usual parallel entry has just one verb relating to words written in a parallel manner. So, the dependency between words is acquired using the dependency structure in a parallel entry.

For example, if a simple dependency structure in G4 is acquired, semantic dependencies acquirable to a verb "作る" (make)" are [鹿児島県 (Kagoshima Prefecture), 作る (make)] and [つのまき (Tsunomaki), 作る (make)]. Since "あくまき" (Akumaki) is an instance the same as "つのまき" (Tsunomaki).

Therefore, the predicate for "あくまき" (Akumaki) is also "作る" (make). Then, in the instance written together by the illustrating maker, an end of dependency which is last instance is regarded as a predicate clause. In this way, the semantic dependency, which is [あくまき (Akumaki), 作る (make)], can be structured.

G5 is the same pattern as G4. Not the word how to write a parallel entry by illustrating makers to be but the comma is used.

3.2 Method of extracting meanings between words

Type of semantic relation	Number	Pattern	Example sentence	Simple dependency relation	Semantic relation by expanded dependency structure
is-a relation	G1	A noun clause which includes particle "が" (ga) or "は" (ha) indicating subject depends on a predicate clause	沢田研二は、日本の歌手である	【沢田-研二、歌手-だ】、【日本、歌手-だ】	【歌手-だ、沢田-研二】
modifying and modified relation	G2	A predicate clause depends on a noun clause	胸いっぱいの悲しみ（むねいっぱいのかなしみ）は、日本の歌手である沢田研二の７枚目のシングルである	【むねいっぱい、かなしみ】、【胸いっぱい、悲しみ】、【悲しみ、シングル-だ】、【日本、歌手-だ】、【歌手-だ、沢田-研二】、【沢田-研二、７枚目】、【７枚目、シングル-だ】	【沢田研二、歌手-だ】、【シングル-だ、悲しみ】
verb pair relation in complex sentences	G3	Two or more predicate clauses depend on another predicate clauses	ヘーシオドスの『神統記』によると、ムネーモシュネーはゼウスと添い臥し、９人のムーサたちを産んだという	【ヘーシオドス、神-統-記】、【神-統-記、よる】、【ムネーモシュネー、添う-臥す】、【ゼウス、添う-臥す】、【9人、ムーサ-たち】、【ムーサ-たち、産む-だ】、【添う-臥す、産む-だ】	【ムネーモシュネー、産む-だ】
parallel entry relations	G4	A clause which includes illustrating makers depends on another clause	鹿児島県では「あくまき」や「つのまき」を作る	【鹿児島県、作る】、【あくまき、つのまき】、【つのまき、作る】	【あくまき、作る】
	G5	A clause which includes commas depends on another noun clause	生物学の一分野である進化生物学では、共通祖先からの種の起源や進化、繁殖、生物多様性などについて研究を行う	【生物-学、一分野-だ】、【一分野-だ、進化-生物-学】、【進化-生物-学、行う】、【共通-祖先、種】、【種、生物-多様-性】、【生物-多様-性、行う】、【起源、進化】、【進化、繁殖】、【繁殖、生物-多様-性】、【研究、行う】	【進化、行う】、【繁殖、行う】

Table 2: Pattern table of semantic relation by expanded dependency structure

As meanings between words, the information on synonyms and negative expression are acquired.

3.2.1 Expansion of synonymous words

Expansion of synonymous words is performed on words in t2 for word pairs in which a semantic dependency does not exist. As synonymous words dictionary, we used the context-similar word database and the entailment verb database of ALAGIN[1], and the Japanese WordNet synonymous pairs database of NiCt[3]. At this time, words with a reliability 60% or more were used in the sentence similar word database.

Moreover, only three dictionaries were used in the entailment verb database. One of dictionary has entailment words which have a including relation in characters/hyponymy words, the second one is entailment words/hyponymy words, and the third one is a relation of action and reaction.

1. Semantic dependencies between words are structured by each of the simple dependency structures in t2.

   t2 [フィラデルフィア (Philadelphia), 場所-だ (a place-is)], [独立宣言 (the Declaration of Independence), 立案する (plan)], [立案する (plan), 場所-だ (a place-is)]

2. The synonymous words in semantic dependency pairs of t2 are looked up in dictionaries.
   For the semantic dependencies in t2, the semantic dependency pairs which replaced each word by the synonymous words registered in dictionaries are created.

   For example, when "起草" (drafting) and "立案" (planning) are registered in dictionaries as synonymous words, the semantic dependency pairs using synonymous words are given to t2. At this time, the semantic dependency pairs of each word in t2 are shown in the following.

   t2 [フィラデルフィア (Philadeiphia), 場所-だ (a place-is)], [独立宣言 (the Declaration of Independence), 立案する (plan)], [立案する (plan), 場所-だ (a place-is)], [起草する (draft), 場所-だ (a place-is)], [独立宣言 (the Declaration of Independence), 起草する (draft)]

   All synonyms of target words retrieve in dictionaries through above processing.

3.2.2 Existence of negative expression

Determining the existence of a negative expression is an important pointer to grasp the different meanings of sentences which have almost the same word structure dependencies. Our system examines the existence of negative expressions in semantic dependency pairs to recognize textual entailment in the following processing.

1. Acquire semantic dependency pairs in sentence pairs is acquired.

2. Acquire the difference in the semantic dependency pairs of sentence pairs.

3. Judge the existence of negative expressions based on the difference acquired in process of 2.
   We set two auxiliary verbs for negative expressions, whose each original form is "ぬ" and "ない".

4. PROCESSES OF OUR SYSTEM

4.1 Proceeding process

1. Dividing round parentheses and sentences

When a round bracket exists in a sentence, there is a problem which cannot do analysis of a dependency structure correctly.

In order to avoid this problem, our system divides the sentence into the phrase in the round bracket and the other main phrases before conducting dependency analysis.

For example, ex2 is classified "旅順攻囲戦は，日露戦争における戦闘の一つ" (The Lushun siege warfare is one of the battles in the Russo-Japanese War) and "1904 年 8 月 19 日-1905 年 1 月 1 日"(from August 19, 1904 to January 1, 1905).

ex2 旅順攻囲戦（1904 年 8 月 19 日-1905 年 1 月 1 日）は，日露戦争における戦闘の一つ (The Lushun siege warfare (August 19, 1904 to January 1, 1905) is one of the battle in the Russo-Japanese War)

2. Unification of full size and half size alphanumeric characters and commas

In order to prevent dependency analysis from failing because of differing from part-of-speech grant, character strings are unified, such as full size or half size alphanumeric characters and punctuation marks.

3. Replacement of a square bracket

The words inside a square bracket are treated as a proper noun in many cases. However, when a dependency analysis is conducted, words in parentheses is divided into a morpheme and in some cases dependency analysis fails.

Then, the word in the square brackets is replaced by an ID so that it is not divided into a morpheme. The "word" described here is a continuation of words or a noun word, and since it did not become a proper noun when a particle and a verb are included in the part-of-speech of each word in parenthesis, it is carried out the outside of an object.

4. Sentence division

Some dependency structures are not in agreement, when one side is a complex sentence and the other side is a simple sentence. Then, the complex sentence is divided into simple sentences, and pretreated so that a more suitable dependency relation can be acquired. The-part-of-speech of the word in front of a comma connects, and when it is a conjunctive particle, a verb, or an auxiliary verb, it is considered to be a complex sentence and our system divides it with commas.

4.2 Weighting of semantic dependency

The semantic dependency used as the frame of a sentence is important; for example, the relation between a subject and a predicate, the relation between a subject and an object, and the relation between an object and a predicate. For textual entailment recognition, the semantic dependency between words used as the frame of a sentence should be treated as more important.

Then, the importance of the semantic dependency between words is given to each semantic dependency pair as dignity. Since the reliability differs in terms of semantic dependency types, one is created by the simple dependency and the other is created by the expanded dependency, and the dignity of the dependency pair using expanded dependency is also set up.

Moreover, since reliability becomes low also when synonymous words are expanded, dignity for this case is set up.

When the start of dependency and the particles immediately after a noun are "が" (ga) or "は" (ha), it is judged as being the "subject part", and it is judged that the clause in which it is the end of dependency, the end of sentence and also a verb are contained is a "predicate part."

Semantic dependencies which performs weighting are following six.

1. a pair only with a subject
2. a pair only with a predicate
3. couple of pairs with a predicate and subject
4. a pair created by semantic relation using expanded dependency structures
5. a pair created by expansion of synonyms

4.3 All processes of our system

This section explains the textual entailment recognition method using each technique shown above.

1. The preparing process explained in section 4.1 to t1 and t2 is performed.
2. Simple dependency structures in t1 and t2 are acquired.
3. In order to perform weighting of semantic dependencies explained by section 4.2, a subject part and a predicate part is extracted.
4. Semantic dependencies are created based on the simple dependency acquired by process 2.
5. Expanded dependency structure which is explained in section 3.1.3 is performed using the dependency acquired by process 4, and created semantic dependencies are added.
6. The non corresponding group and the corresponding group of semantic dependency pairs re-created in section 5 is acquired.
   At this time, the degree of corresponding semantic dependencies is computed using words except "sign", "particle" and "auxiliary verb" which are unnecessary in many cases.
7. Expansion of synonyms explained by section 3.2.1 is performed for the dependency conflicting between sentence pairs acquired in process 6.
8. Corresponding of the semantic dependency which acquired in process 7 in t2 and the semantic dependency in t1 is judged.
9. It is recognized whether there is any negative expression explained in section 3.2.2 to clauses conflicting between t1 and t2 process 8.

10. The rate of the number of semantic dependencies which is in agreement with t1 from all the semantic dependencies in t2.
    At this time, dignity of the semantic dependency explained by section 4.2 is used.
    When the clause weighted in t2 is in agreement between sentence pairs, dignity is added to congruous numbers of semantic dependencies. If negative expression do not exist, our system judges "entailment" when the rate of corresponding semantic dependencies between sentence pairs is more than a threshold, or it judges "non entailment" if lower than a threshold.

5. STRUCTURE OF OUR SYSTEM

The developed system consists of a BC system which recognizes textual entailment between two sentences, and a search system used by an examination subtask. Moreover, we use the BC system for the BC subtask and combination of the search system and the BC system for the examination subtask.

5.1 Composition of BC system

An outline of the BC system built based on the system explained in section 4.3 is shown in Figure 2. The processing until computing the coincidence rate of the semantic dependency between sentence pairs in t2 from a preparing process is called "system for calculating coincidence rate of a semantic dependency." If in charge of a textual entailment recognition, it considered that the coincidence rate of the semantic dependency was reliability, and our system judges the entailment relation using the threshold.

5.2 Structures of Exam Search System

An outline of the search system used for an examination subtask is shown in Figure 3. In our system, the keyword used for search is acquired from t2, and the sentence equivalent to t1 is searched and acquired from Wikipedia and textbook data. We use Solr[2] as a search engine. The processing which selects the keyword used for search is as follows.

1. A morphological analysis performs word division and part-of-speech grant for the sentence of t2.

2. Four part-of-speech types are carried out the outside of object, which are "非自立語" (the non-independent word), "助詞" (particle), "助動詞" (auxiliary verb) and "記号" (sign). However, "ない" (not) which is negative expression is treated as a keyword.

Text information was retrieved based on this keyword, and after acquiring a sentence before and after a searching keyword, it was considered as the sentence which is equivalent to t1 by dividing with a period.

Although the BC system shown in Figure 2 is used for recognizing textual entailment, in the examination task, since two or more sentences equivalent to t1 obtained, the processing for corresponding to recognize textual entailment using two or more sentences of t1 was added.

Specifically, coincidence existence of the semantic dependency of t2 was not judged for every sentence of t1, but the semantic dependency included in a document set of t1 was acquired, and our system calculated the number of the semantic dependencies of t2 contained in these sets. It then

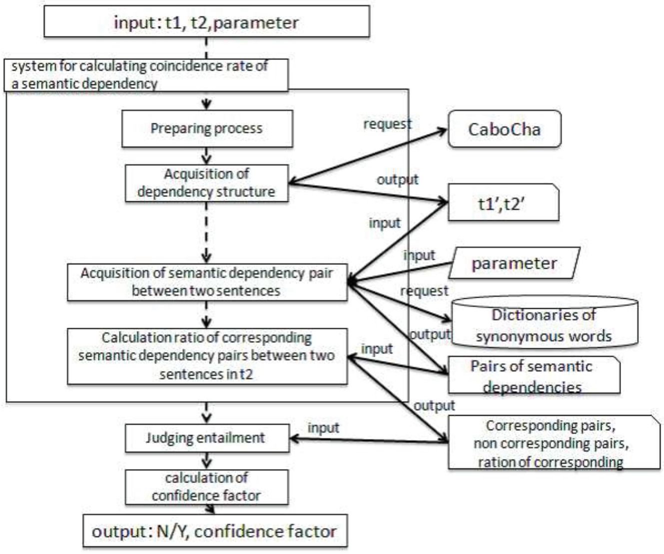

Figure 2: System for BC subtask

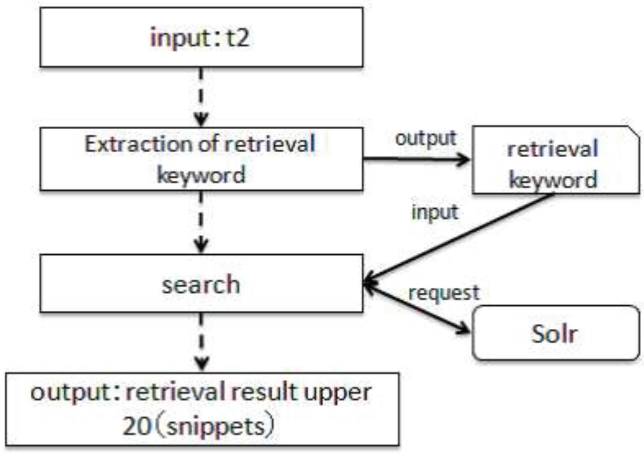

Figure 3: Text searching system

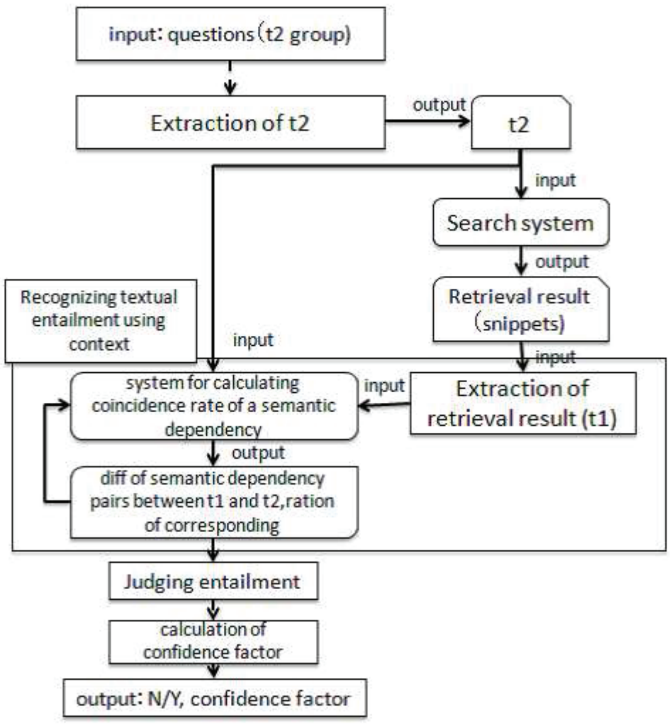

Figure 4: System for Exam Search subtask

computed the coincidence rate based on the number of semantic dependencies of t2. An outline of the system used by the examination subtask is as Figure 4.

6. EXPERIMENTAL RESULT

In order to check the validity of the technique proposed in section 3, we participated in the Unit Test subtask, BC subtask, Exam BC subtask, and Exam Search subtask[5]. Each execution result of these subtasks is shown in Table 3.

We used the threshold of degree of coincidence of a dependency as a parameter, and prepared two or more systems.

Since our system is not aimed at an entailment type of implication, it seldom fits the task which recognizes sentence pairs of paraphrasing like an examination subtask (Exam BC, Exam Search.) However, we also challenged the examination subtask in order to verify that our system has effects to recognize textual entailment.

In the case of our system, the reliability is set to ”0” in many cases. In this case, when performing an actual examination task, the answer of the same question will be determined without using reliability.

Then, our system uses a weight of each word since the difference is attached, In order not to change the output result of Y and N, the weight used 1/1000. The Parameter column in Table 3 is the decision criteria of the parameter of each system.

When a word is given a high weighting, it means our system highly evaluates the coincidence rate of a word between two sentences.

Since the F value is high when the BC subtask estimates the coincidence rate of a word highly, although the coincidence rate of a word is large, there are many instances in which our system cannot correctly grasp correctly the semantic dependency between sentences.

However, our system succeeded partly for the Unit Test subtask and Exam Search subtask. Our system can recognize a textual entailment type which has one language phenomenon. In the Exam Search subtask, our system used all document sets of t1, it means our system use partial context, so our system got a fairly good score.

7. DISCUSSION

We proposed using meanings and semantic dependencies between words in sentences. Moreover, we devised a method to recognize semantic dependency by using an expanded simple dependency structure. The result of a formal run showed that our system has partial effects to recognize the entailment which has one language phenomenon.

In the future, we will investigate an approach to textual entailment types of ”rephrase” and ”inference”, and aim to devise a technique that can comprehensively recognize textual entailment.

8. REFERENCES

[1] Alagin. https://alaginrc.nict.go.jp/resources/nictmastar/resource-info/abstract.html.

[2] Solr. http://lucene.apache.org/solr/.

[3] F. Bond, T. Baldwin, R. Fothergill, and K. Uchimoto. Japanese semcor: A sense-tagged corpus of japanese. In *The 6th International Conference of the Global WordNet Association*.

[4] T. Kudo and Y. Matsumoto. Japanese dependency analysis using cascaded chunking. In *CoNLL 2002: Proceedings of the 6th Conference on Natural Language Learning 2002*, pages 63–69, 2002.

[5] Y. Watanabe, Y. Miyao, J. Mizuno, T. Shibata, H. Kanayama, C.-W. Lee, C.-J. Lin, S. Shi, T. Mitamura, N. Kando, H. Shima, and K. Takeda. Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10. In *Proceedings of the 10th NTCIR Conference*, 2013.

Subtask	System	Parameter	Value	F1 in development data	McroF1	Accuracy	Correct Answer Ratio	Y-F1	Y-Prec	Y-Rec	N-F1	N-Prec	N-Rec
BC	01	F value is highest	20.00	55.65	57.59	64.10	-	40.97	66.09	29.69	74.20	63.64	88.98
	02	F value is highest in Unit Test sub-task	50.00	50.11	50.38	62.46	-	25.89	75.47	15.63	74.86	61.22	96.33
	03	F value is highest, when 5.0 points is added to the weight of each word	20.00	61.36	61.90	62.30	-	58.03	54.45	62.11	65.77	69.50	62.43
Unit Test	01	F value is highest	50.00	67.80	61.99	74.27	-	83.60	95.18	74.53	40.38	28.00	72.41
	02	Well-balanced trade-off between precision and recall	20.00	64.18	68.98	84.23	-	90.73	93.94	87.74	47.22	39.53	58.62
	03	Well-balanced number between N and Y answer	85.00	50.96	47.91	53.94	-	65.63	95.50	50.00	30.19	18.46	82.76
Exam BC	01	F value is high and score is also high	5.00	56.44	52.02	58.93	31.48	33.81	44.76	27.17	70.23	63.27	78.91
	02	Score is highest, when 5.0 points is added to the weight of each word	15.00	56.86	55.57	55.58	34.26	54.88	45.15	69.94	56.26	71.11	46.55
	03	F value is highest, when 5.0 points is added to the weight of each word	20.00	58.56	53.12	54.02	34.26	46.63	42.25	52.02	59.61	64.68	55.27
Exam Search	01	F value is highest	30.00	60.17	55.02	58.04	25.93	43.37	45.28	41.62	66.67	65.05	68.36
	02	Score is highest	10.00	56.30	49.15	49.33	25.93	52.21	41.06	71.68	46.08	66.44	35.27

Table 3: Result of formal run