Validating Contradiction in Texts Using Online Co-Mention Pattern Checking

CHENGWEI SHIH, Academia Sinica; National Tsinghua University
CHENGWEI LEE, Academia Sinica
RICHARD TZONGHAN TSAI, YuanZE University
WENLIAN HSU, Academia Sinica

Detecting contradictive statements is a foundational and challenging task for text understanding applications such as textual entailment. In this article, we aim to address the problem of the shortage of specific background knowledge in contradiction detection. A novel contradiction detecting approach based on the distribution of the query composed of critical mismatch combinations on the Internet is proposed to tackle the problem. By measuring the availability of mismatch conjunction phrases (MCPs), the background knowledge about two target statements can be implicitly obtained for identifying contradictions. Experiments on three different configurations show that the MCP-based approach achieves remarkable improvement on contradiction detection and can significantly improve the performance of textual entailment recognition.

Categories and Subject Descriptors: I.2.7 [**Artificial Intelligence**]: Natural Language Processing—*Text analysis*

General Terms: Languages

Additional Key Words and Phrases: Textual entailment, contradiction detection, Web mining, Chinese

ACM Reference Format:
Shih, C., Lee, C., Tsai, R. T., and Hsu, W. 2012. Validating contradiction in texts using online co-mention pattern checking. ACM Trans. Asian Lang. Inform. Process. 11, 4, Article 17 (December 2012), 21 pages.
DOI = 10.1145/2382593.2382599 http://doi.acm.org/10.1145/2382593.2382599

1. INTRODUCTION

In natural language processing (NLP), the term textual entailment (TE) refers to a directional relation between two text segments, usually sentences. TE can be said to exist when the truth of one text segment follows from that of another. The entailing segment is termed t while the entailed text (hypothesis) is termed h. Textual entailment has a more relaxed definition than pure logical entailment: "t entails h" ($t \Rightarrow h$) if, normally, a person reading t would infer that h is probably true [Monz and Rijke 2001]. TE relations are considered directional because even if "t entails h", the reverse "h entails t" may not be true [Marneffe et al. 2008].

Many NLP applications, such as question answering, information extraction, and machine translation evaluation, could make use of a model that would allow them to recognize whether the same meaning can be inferred from different sentence variants. At the 2004 PASCAL challenge, Recognizing Textual Entailment (RTE) was first

Author's address: C. Shih, Institute of Information Science, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei 11S, Taiwan; email: dapi0428@mail2000.com.tw.

DOI 10.1145/2382593.2382599 http://doi.acm.org/10.1145/2382593.2382599

proposed as a generic task that can fulfill semantic inference needs in many NLP applications [Dagan et al. 2005]. Since then, RTE has become a popular research topic in NLP research.

Textual entailment can generally be categorized into three different relation types [Giampiccolo and Magnini 2007]: Positive TE (text entails hypothesis), negative TE (text contradicts hypothesis), and non-TE (text neither entails nor contradicts hypothesis). Early RTE challenges asked participants to determine whether or not a given sentence positively entails another sentence, which is a binary classification task. Since RTE-3 (2007), identification of the latter two relation types has been included. The Recognizing Inference in TExt (RITE) task, another RTE task held at the 9th Annual Meeting of NII Test Collection for IR Systems (NTCIR), further extended the relation types to five and included Japanese and traditional and simplified Chinese language datasets [Shima et al. 2011]. Most teams today, however, still focus on identifying positive TEs and usually obtain better performances for this type.

As a participant of RITE at NTCIR-9, we proposed a knowledge-based hybrid approach and developed a Chinese RTE system consisting of several independent modules [Shih et al. 2011]. We simultaneously classify the entailment direction (forward, reverse, bidirectional), contradiction and independent (otherwise) by using modules which use shallow text features such as sentence alignment, token similarity, presence of named entities, syntactically dependent word pairs, negative expressions, and length difference between text and hypothesis sentences. The scores for all entailment relations are integrated by a voting strategy, and the relation with the highest vote is considered the result. Our system achieved accuracy scores of 0.661 and 0.501 for traditional Chinese binary and multiple-class subtasks, respectively. The system was ranked the highest among fully automatic systems in the RITE traditional Chinese category.

After doing error analysis on the RITE results, we found a possible direction to improve the ability of identifying contradictory statements. This relation type got a maximum F-score of 0.368, which effectively lowered the overall accuracy. Poor performance in this type indicates that relying only on the occurrence of negations and some antonyms for contradiction detection is insufficient to discriminate conflicts from unrelated (unknown) text pairs. This observation motivated us to investigate the characteristics of contradictory statements, strategies for detecting them, and the effects of more accurate contradiction detection.

In propositional logic, the definition of contradiction is that two statements cannot both be true at the same time; nevertheless, in text comprehension, contradictory cases do not necessarily match such a strict definition but rely on human intuition. According to Marneffe et al. [2008], the primary types of contradiction can be defined as: (1) those occurring via overt negations such as antonyms, negations, and time/date/number mismatches; and (2) contradictions resulting from factive/modal words, text structure, certain lexical contrasts, and word knowledge. Unlike the former type, which may only need a little external information for detection, conflict cases of the second type cannot be solved without large amounts of presupposition, background knowledge, and even specific know-how. Take the following sentences, for example.

茱莉安德魯斯在「麻雀變公主」裡飾演女王
(Julie Andrews played a queen in the film "The Princess Diaries")

茱莉安德魯斯在「麻雀變公主」裡飾演公主
(Julie Andrews played a princess in the film "The Princess Diaries") (1)

There are no negations or antonym pairs in these two sentences. The mismatches "女王" (queen) and "公主" (princess) in this sentence pair might be treated as two ordinary words. However, these two sentences are contradictory for the notion that generally one person may not play two characters in a single movie. Although a dual role is possible in filmmaking, its possibility is quite low that most people would regard these two as contradictory. One possible way to deal with this issue is to create or collect rules to detect the mismatch in a pair, but it would be extremely difficult to have a set of rules that is complete enough to handle most contradictory cases because many of them are defined by the context, not by the mismatched parts. In the example pair, "女王" (queen) and "公主" (princess) can co-exist in some situations, such as eating together or traveling together, however, relations such, *playARole (PersonA, Queen)* and *playARole (PersonA, Princess)*, rarely co-exist in an event. In other words, the contradiction is true because of the same relation and same subject, which defines the contradictory context. The cost of collecting enough rules to cope with all possible mismatches is very high. For this reason, it is clear that we do not have a good chance of identifying contradictions between two sentences without external information. Such difficult cases, in which contradiction can only be detected with specific background knowledge, are widespread in written language.

In this article, we focus on detecting contradiction in traditional Chinese (CT) text. More specifically, we first investigate a RITE dataset and categorize types of contradictory statements according to the linguistic phenomena and which manifest the problem of knowledge shortage. Furthermore, we propose an approach of using a search engine to aid in the detection of contradictions. Critical mismatches are validated along with their contexts based on search engine result counts returned by suitable query strings. These query strings are carefully designed so that the contradiction level can be measured in terms of search result counts. The purpose of this research is to emphasize the seriousness of knowledge insufficiency in contradiction detection, and provide a novel approach of using co-mention patterns to extract essential but scanty knowledge from the Internet for contradiction judgment.

The rest of this article is organized as follows: Section 2 gives a brief review of contradiction detection within the framework of textual entailment recognition. Section 3 describes the categories of linguistic phenomena and the problem of knowledge shortage. Section 4 introduces our method of using the mismatch conjunction of phrase and web query to detect contradictions in a text. Section 5 introduces our baseline RITE system, IASLD, and then the experiments and the results from two different aspects for estimating the effect of our approach will be conducted in Section 6. We analyze and discuss the approach on contradiction detection in Section 7, and finally present the conclusion in Section 8.

2. RELATED WORKS

The amount of research dedicated to finding contradictions in texts is limited. Crouch et al. [2003] was the first study to suggest that contradiction detection should be considered as important as positive entailment detection. Later, Harabagiu et al. [2006] proposed an approach to discover contradictions such as negations and antonyms using two strategies: (1) identifying and removing negations of propositions, and (2) extracting linguistic information of negations, contrasts and antonyms from the text for training a classifier. Ritter et al. [2008], on the other hand, conducted a deep analysis of the RTE-3 extended task and addressed some issues with and potential directions for contradictory statement generation and justification. Marneffe et al. [2008] not only described corpus annotation guidelines and several types of contradiction but also proposed some salient features (such as number/date/time, negations, antonyms,

factivity, modality, relations, and structures) useful in identifying incompatible information in texts. Building on this work, Voorhees [2008] focused on contradictions that can only be resolved with background knowledge and described a case study of contradiction detection based on functional relations. Magnini and Cabrio [2010] introduced a novel methodology for the qualitative analysis of a TE system focusing on contradiction judgments based on decomposing text-hypothesis pairs into monothematic pairs.

In RTE-4 and RTE-5, more than 20 participants in the three-type RTE subtask presented strategies for dealing with contradictions. Most of these approaches looked for the presence of negations and antonyms. Other cues, such as time, date, number, location name, quantifiers were widely used to improve the accuracy of contradiction detection. In addition, utilization of dictionaries and databases such as WordNet [Fellbaum 1998], VerbNet [Kipper-Schuler 2005], Wikipedia, or inference rule sets such as DIRT [Lin and Pantel 2001] were also common. Systems such as Clark and Harrison [2008] introduced the use of logical forms for RTE. Pado et al. [2008] relied on filtering non-coreferent events to distinguish irrelevant sentence pairs from mismatched cases. Machine-learning-based RTE systems such as those described in Castillo [2009], on the other hand, tend to gather all available features to train a textual relation classifier which can automatically distinguish entailed texts from non-entailed cases. Moreover, some participants, such as Iftene [2008] and Iftene and Moruz [2009], proposed a sophisticated hybrid system, applying a mapping from every node in the hypothesis tree to one node from the text tree to compute a fitness score for differentiating entailments from non-entailed pairs.

3. BACKGROUND KNOWLEDGE FOR DETECTING CONTRADICTIONS

We discuss linguistic phenomena of contradictions and the knowledge insufficiency problem in this section.

3.1. Linguistic Phenomena of Contradictions in RITE

As mentioned earlier, Marneffe et al. [2008] suggested a primary contradiction classification strategy and clearly pointed out that the difficulty of tackling the second type of contradiction, which is associated with factive/modal words, text structure, certain lexical contrasts, and word knowledge. Magnini and Cabrio [2010] also found that the second type of contradiction is much more frequent in textual entailment datasets than in ordinary newswire documents and Wikipedia; and overt negations, which often appear in these datasets, are not as effective for entailment judgment. Moreover, the linguistic phenomena that cause two statements to become contradictive are quite varied and complicated. These factors make contradiction detection in RTE datasets a more challenging task.

In order to successfully recognize entailments and contradictions, developers and researchers have created a variety of new features, algorithms, and knowledge resources to address specific type of linguistic phenomena in entailed and contradictory cases. However, as Bentivogli et al. [2010] pointed out, it is very difficult to measure the impact of these new methods due to (1) the sparseness of the linguistic phenomena and (2) the difficulty of isolating each linguistic phenomenon for independent evaluation.

To deal with the above problems, some previous research attempted to break the textual entailment problem down into certain isolated "aspects". Magnini and Cabrio [2009] first proposed subdividing the TE problem into a set of *monothematic pairs*.

By means of human judgment, monothematic pairs can be created based on linguistic phenomena relevant to the entailment relation. For example, the TH pair,

娜拉提諾娃一共獲得 18 座大滿貫金杯
(Martina Navrátilová won 18 Grand Slam championships in total)

娜拉提洛娃一共取得了 58 個大滿貫的金杯
(Martina Navrátilová won 58 Grand Slam championships in total) (2)

can be resolved (i.e., judged as contradictive) by means of quantity inconsistency between the two sentences (18 vs. 58). There is no other knowledge involved in the entailment. Magnini and Cabrio defined such an instance as a monothematic pair with respect to the quantity phenomenon. Later, Bentivogli et al. [2010] extended and refined Magnini and Cabrio's methodology, providing five macro categories of linguistic phenomena related to contradiction that are present in the RTE-5 dataset. The five macro categories are as follows.

— Lexical, which includes identity, format, acronymy, demonymy, synonymy, semantic opposition, hyperonymy, and geographical knowledge.
— Lexical-syntactic, which includes transparent heads, nominalization, verbalization, causatives, and paraphrases.
— Syntactic, which includes negation, modifiers, argument realization, apposition, lists, coordination, and active/passive alternation.
— Discourse, which includes coreference, apposition, zero anaphora, ellipsis, and statements.
— Reasoning, which includes apposition, modifiers, genitives, relative clauses, elliptic expressions, meronymy, metonymy, membership, representativeness, reasoning on quantities, temporal and spatial reasoning, and all the general inferences using background knowledge.

Bentivogli's work took a new perspective on using linguistic phenomena of TH pairs to simplify the complexity of TE problems. It also pointed out the possibility of classifying TE problems via linguistic phenomena relevant to entailment relation determination.

Returning to contradiction detection, Magnini and Cabrio [2010] followed the work of Bentivogli et al. [2010] to investigate linguistic phenomena related to contradictory pairs in the RTE-5 dataset. According to their observations, contradictions are mainly triggered by phenomena such as quantity mismatching, antonymy, apposition mismatching, and general inference. Negations, on the other hand, were surprisingly relevant to only one TH pair's contradiction. Their work also revealed the contribution probability of linguistic phenomena toward positive or negative assignment. According to their statistics, some linguistic phenomena, like antonyms, are more reliable in detecting contradictions, whereas other linguistic phenomena that appeared in contradictions may also be found in entailed text pairs. Furthermore, the distribution of the five macro categories of contradiction-related linguistic phenomena is still unknown in the Chinese RITE dataset. In order to gain a deeper understanding of contradiction behavior in RITE, we examined contradictory cases in the Traditional Chinese (TC) RITE development dataset, and analyzed the distribution of the linguistic phenomena related to negative relation assignment. Table I shows the linguistic phenomena and the number of corresponding contradictory TH pairs in the RITE TC multiple classification (MC) development set.

3.2. Incompleteness of Background Knowledge

Decomposing TH pairs into monothematic pairs reveals the relevance and contribution of linguistic phenomena to contradiction detection. But for some linguistic phenomena, it is difficult to create corresponding monothematic pairs from TH pairs without human supervision, especially for those TH pairs with deep reasoning or complicated grammatical structures. Most of the contradiction detection systems focus their attention on investigating more explicit linguistic phenomena such as negations, antonyms, modifiers, and lexicon mismatches. These linguistic phenomena, which can mostly be found in literal mismatches between T and H, are the main clues for contradiction detection for these systems. In our analysis of the RITE TC corpus, problems relating to the incompleteness of background knowledge were discovered in the following linguistic phenomena.

(1) Incompleteness of background knowledge in antonym identification.

Antonymy relations, which are one of the most trustworthy phenomenon types in RITE, are generally obtained from thesauri such as WordNet. In Chinese, knowledge resources such as E-hownet [Chen et al. 2005] and Sinica Bow [Huang et al. 2004] provide opposing terms and antonymy information as well. Unfortunately, dictionary-based antonym identification suffers from the fact that only a limited number of antonym instances are available, as can be seen in the following TH pair:

巴基斯坦的宿敵印度 (*Pakistan, an old enemy of India*)

巴基斯坦的好朋友印度 (*Pakistan, a good friend of India*) (3)

Obviously the mismatches between two sentences "*宿敵*" (an old enemy) and "*好朋友*" (a good friend) are antonyms. However, we find hardly any antonymy relations between these two terms in all available Chinese knowledge resources.

(2) Incompleteness of background knowledge in lexicon mismatches.

Similar background knowledge shortages not only happen with antonyms, but can also be detected in ordinary lexicon mismatches in TH pairs, such as in the following example.

亞伯雷斯克獎得主經常兩年內再獲得諾貝爾獎
(The winner of the Lasker Award usually receives the Nobel Prize within two years)

亞伯雷斯克獎得主都是兩年內再獲得諾貝爾獎
(The winner of the Lasker Award always receives the Nobel Prize within two years) (4)

In this example, the two sentences conflict due to the difference between "*經常*" (usually) and "*都是* " (always). Even if the meanings of these mismatches can be found in dictionaries, it is hard to determine if they cause these TH pairs contradictive.

(3) Incompleteness of background knowledge in reasoning.

Another knowledge shortage problem worth noting regarding contradiction is reasoning. As Magnini and Cabrio [2010] reported, the reasoning category is the most frequent of all linguistic phenomena in the RTE-5 dataset. The requirements of background knowledge, however, make resolving statements containing implicit incompatibilities difficult. This argument can be supported by the low contribution probability

Table I. Distribution of Linguistic Phenomena and Their Examples in Contradictory RITE CT MC Pairs

Phenomena		# TH Pairs	Example
Lexicon	Identity	10	茱莉安德魯絲 **1930 年**出生 **1935 年**出生於英國的茱莉・安德魯絲
	Format	8	和平號太空站於 **3 月 23 日下午 1 時 59 分**墜落 和平號太空站，**三月二十三日下午二時半**左右…
	Demonym	6	**車臣人**一九九六年擊敗俄軍，取得實質獨立 …**車臣籍**的前蘇聯空軍將領杜達耶夫…
	Synonym	3	張藝謀 1987 年以「紅高粱」**拿下**柏林影展金熊獎 …他執導的《紅高粱》**贏得**了最佳影片 金熊獎…
	Antonym	10	…廢氣排放量**降到**比一九九零年**低**百分之五點二... …廢氣排放量**升到**比 1990 年**高**百分之五點二…
Syntactic	Negation	11	小泉純一郎 2001 年贏得自民黨總裁選戰 小泉純一郎 2001 年**未**贏得自民黨總裁選戰
	Modifier	10	亞伯雷斯克獎得主**經常**兩年內再獲得諾貝爾獎 亞伯雷斯克獎得主**都是**兩年內再獲得諾貝爾獎
	List	2	…是**不宣布獨立，不更改國號，不推動兩國論入憲，不推動改變現狀的統獨公投，也沒有廢除**… ..是**不獨，不改國號，無兩國論，不統以及沒有廢除國統綱領與國統會**…
	Apposition	6	張學良 1928 年(廿八歲)當上**總司令**成為「東北王」 張學良的**總司令**叫做「東北王」
Discourse	Coreference	1	**張藝謀** 1987 年以「紅高粱」拿下柏林影展金熊獎 …**他**執導的《紅高粱》贏得了最佳影片 金熊獎…
	Anaphora-Zero	5	**鞏俐** 1966 年生 …，祖籍山東濟南並在此長大
Reasoning	Temporal	9	何厚鏵**現任**澳門國際機場專營公司董事局副主席 何厚鏵**曾任**澳門國際機場專營公司董事局副主席
	Spatial	1	…突然在**印尼蘇門答臘叢林**中墜毀 …墜毀於**印尼的一條河**中
	Quantity	4	小泉純一郎 1998 年再度角逐自民黨總裁，**仍失敗** 小泉純一郎角逐自民黨總裁**失敗一次**
	Modifier	3	…有**一半 (六十一位)** 的亞伯雷斯克獎得主再獲得諾貝爾獎 亞伯雷斯克獎得主**都會**再獲得諾貝爾獎
	Apposition	9	茱莉安德魯絲**丈夫**是布萊克艾德華 茱莉 安德絲在 1959 年與**東尼 華爾頓**(Tony Walton)結婚
	General-inference	15	鞏俐**老公**是黃和祥 鞏俐與新加坡老公黃和祥**離婚** 1 年多
	Background knowledge	19	若望保祿二世是天主教的**領袖** 若望保祿二世是天主教的**總理**

of reasoning phenomena in contradiction judgment (0.41). For example in the sentence pair,

小李穿了一件藍色的毛衣 (Lee wore a blue sweater)
小李穿了一件藍色的長褲 (Lee wore blue pants) (5)

there is no contradiction . However, if we rewrite Example (5) as

小李穿了一件藍色的裙子 (Lee wore a blue skirt)
小李穿了一件藍色的長褲 (Lee wore blue pants), (6)

there seems to be a conflict because generally people do not wear a skirt and pants at the same time. If we switch the main verb of Example (5) to the following,

小李買了一件藍色的裙子 (Lee bought a blue skirt)
小李買了一件藍色的長褲 (*Lee bought blue pants*) (7)

the sentences in Example (6) turn out to be compatible again. The reason why Example (6) is incompatible is obvious, but such general knowledge facts, which are almost always unstated, are not usually found in dictionaries. These types of contradictions arising from predicate mismatches are rarely considered.

Generally speaking, the issues described above are due to a shortage of background knowledge. Unlike other linguistic phenomena which have more easily identifiable patterns or features, contradictory statements caused by specific background knowledge may occur in any kind of TH pair mismatches. Furthermore, the knowledge insufficiency problem is encountered in most linguistic phenomena in the lexicon, syntactic, and reasoning categories. To the best of our knowledge, few previous studies have endeavored to explore contradictions caused by background knowledge and common sense knowledge, and there has been little research into solving the deficiency of required information like oppositions (antonyms). Thus, researchers and developers of entailment recognition or contradiction detection systems usually face the predicament that even if there is only one mismatch between two literally matched statements, it is still difficult to determine whether or not the in-between relation is contradictive because of the unavailability of necessary background knowledge about the mismatches. For this reason, we believe that developing a complementary mechanism to obtain essential background knowledge is vital.

4. METHODS OF USING MISMATCH CONJUNCTION PHRASE FOR CONTRADICTION DETECTION

When considering the derivation of knowledge for contradiction detection, Web content is always a good choice as the resource to retrieve useful information. We soon realized the serious problem that people take background knowledge such as lexicon meaning, common sense, term usage, term relations, and presuppositions for granted and do not provide extra explanations when making an utterance on the Web. Fortunately, on the other hand, those utterances will not contradict the common sense, presuppositions, and implicit language regulations, such as term usage constraints. We may further argue that if a statement never happened on the web, this might be an implication that this statement is incorrect or unreasonable. In view of this, merging mismatch segments between texts into a single statement and checking its availability on the Web might be a workable and practical solution for detecting implicit incompatibilities, since those incorrect or unreasonable merging which may go against the common usage of language or background knowledge will be recognized for getting no hits or

an extremely low number of hits in a Web search. Some previous research utilized text distributions for exploring synonyms from the Web. Lin et al. [2003] proposed a methodology of using patterns such as "from *X* to *Y*", "either *X* or *Y*" to distinguish synonyms and antonyms. However, their work did not address the contradiction occurring via background knowledge but only focused on acquiring a synonym list from a Web query. Further, some patterns they used in their work do not have suitable correspondents in Chinese. Thus, we need to design a new procedure with the goal of detecting combinations of incompatible mismatch segments in Chinese TE via the infrequency of unreasonable mismatch conjunction phrases on the Web. Details of this will be described in the next paragraph.

4.1. Definition of Mismatch Conjunction Phrase

After analyzing contradiction pairs in the RITE corpus, we noticed a finding from the observations: if both T and H contain only one mismatch between TH pairs, then the relation between the mismatch segments mostly indicates the relation between TH pairs. We define such a mismatch as a *critical mismatch*. After deriving the critical mismatch segments of a TH pair, we can try to connect them by certain conjunctions to form a new statement.

聖嬰現象最強的 1 次是從 1982 年-1983 年
(The most powerful El Niño was from 1982 to 1983)

聖嬰現象最顯著的 1 次是從 1982 年-1983 年
(The most noticeable El Niño was from 1982 to 1983) (8)

In Example (8), the underlined sequences are critical mismatch segments of the two sentences. We can combine these two critical mismatch segments as "*強和顯著*" (the most powerful and most noticeable) by inserting a conjunction "*和*" (and). As a matter of course, the connection process may produce strange statements, such as the following instance.

昆明世界園藝博覽會地點在中國
(Flora Exposition in Kunming was located in China)

昆明世界園藝博覽會為期 184 天
(Flora Exposition in Kunming lasted 184 days) (9)

The connected mismatch segments "*地點在中國和為期 184 天*" (was located in China and lasted 184 days) looks uncommon in Chinese. Another noticeable case of merging occurs in the connection of two opposite statements, like the mismatch segments "*宿敵*" (an old enemy) and "*好朋友*" (a good friend) in Example (2) that generate a statement such as "*宿敵和好朋友*" (an old enemy and a good friend).

Hence, a definition of combining mismatch segments, which we call *mismatch conjunction phrases* (MCP), is given.

Definition 4.1 (*Mismatch Conjunction Phrases,* MCP). assume Mis(X) and Mis(Y) represent the mismatch segments in sentence X and sentence Y. A mismatch conjunctions phrase MCP(X,Y) is defined as

$$MCP(X, Y) = FuncP + Mis(X) + FuncS + Conj + FuncP + Mis(Y) + FuncS$$

Where "+" concatenates two strings. Conj stands for the conjunction. *FuncP* and *FuncS*, which represent functional prefix and suffix, consist of manually collected

Table II. Conjunctions, Functions, and Constituent's POS Used in the Mismatched Conjunction Phrase Generation Process

Conjunction	Function	POS
"與", "及", "和"	and	Noun
"並", "且", "並且", "而且", "也"	and	Verb
"又", "且"	and	Adjective
"且", "又"	and	Adverb
"或", "還", "還是"	or	All
"到"	to	All

Chinese function characters such as "*的*"(de), "*是*" (is). These functional prefixes and suffixes are used for three reasons. (1) to make the generated MCP more fluent, (2) to increase the matched webpage numbers, and (3) to avoid finding Web pages with unmatched or misaligned Chinese segmentation.

We introduce the process of generating mismatch conjunction phrases later in this section.

4.2. Mismatch Conjunction Phrase Generation

Two issues need to be addressed in the process of generating MCPs of TH pairs. First, the conjunction used to combine two mismatch segments together should be refined beforehand. Second, the TH pair types, which include or exclude from the process of generating MCPs, should be determined.

Let us assume first that a mismatch is critical. Coordinating conjunctions, which are one type of conjunction and can join two or more items of equal syntactic importance [Curzan and Adams 2005], are applied to connect two mismatch segments to generate an MCP. With different coordinating conjunctions, different MCPs are produced. Coordinating conjunctions such as "*與*" (and), "*而且*" (and), and "*和*" (and) are used to bind two co-existing constituents; the other coordinating conjunctions such as "*但*" (denoting an exception), "*到*" (denoting a range), "*或*" (denoting an option), and "*等*" (denoting "and so forth") join two or more opposite, unrelated, or similar items together. Moreover, according to Academia Sinica [1993], the association or collocation of coordinating conjunctions and their constituents still vary from the constituents' part-of-speech (POS). To be more specific, the coordinating conjunctions used to connect nouns are different from those which are used to connect verbs, adjectives, and adverbs in practical use. Therefore, while generating MCPs, we use coordinating conjunctions corresponding to the POS tag of the constituents. Table II shows the list of conjunctions, their functions, and the constituents' POS in the generation processing. We use mismatched segments "*強*" (powerful) and "*顯著*" (noticeable) in Example (8) to demonstrate the process of generating an MCP: the conjunctions for adjectives are "*又*" (and), and "*且*" (and). And the functional suffix for adjectives is "*的*" (de). According to Definition 4.1, two MCPs of this mismatch segments pair, "*強的又顯著的*" and "*強的且顯著的*" (powerful and noticeable), are generated.

Not all the TH pairs are suitable for generating MCPs. We skip TH pairs with more than two mismatched segments on each side. This constraint, based on the assumption that if there is only one mismatch in a TH pair, then the mismatch is critical and is used to remove mismatches of little significance to the relation judgment. Furthermore, TH pairs with highly complicated linguistic phenomena are also excluded by

this constraint because more mismatches may indicate more complicated contradiction phenomena within; in any case, the TH pairs with highly complicated linguistic phenomena are not our targets to resolve.

4.3. Checking Mismatched Conjunction Phrase for Detecting Contradictions

As mentioned before, our purpose is to find unreasonable MCPs, which may indicate the existence of a contradictory binding of mismatched segments. Therefore, a strategy to distinguish the reasonableness/unreasonableness of MCPs is needed. We found that the AND conjunction phrases, which contain the conjunction "and", tend to be utilized for connecting two co-existing and non-conflicting constituents. On the other hand, the disjunctive conjunction, such as "or", mostly link two opposite concepts. The distribution of MCPs with these two types of conjunctions may display the compatible degree between mismatched segments. In other words, the reasonableness or unreasonableness of mismatching segment combinations leads the distribution difference of the MCPs. In addition, some TE contradictions are constructed from the opposition of the meaning between critical mismatch segments, but some incompatibilities result in the unreasonable co-existence of combinations of verb and mismatch segments. Example (6) demonstrates how two verb-mismatch combinations bring about a contradiction. Accordingly, while searching for incompatibilities in texts, taking the verbs of mismatched segments into account is necessary.

We apply a step-by-step Web-based checking procedure to show how we use MCPs querying to identify contradictions. Initially, assume $Mis(T)$ and $Mis(H)$ are critical mismatch segments of a TH pair. A set of MCPs, $MCP_{and}(T, H)$, which consist of all possible combinations of AND conjunctions and functional prefixes/suffixes, is generated. Considering the connecting order, the $MCP_{and}(H, T)$, in which the order of constituents are exchanged, is produced as well. Thereafter, the query strings of MCPs, $QSMCP$, are generated. The definition of $QSMCP$ is defined as follows.

Definition 4.2 (*Query Strings of MCP, QSMCP*). Assume MCP(X, Y) represents a mismatch conjunction phrase of mismatch segments Mis(X) and Mis(Y). The verb of the Mis(X) and Mis(Y) in sentence X and sentence Y is represented as V. Then the query strings of MCP(X, Y), QSMCP(X, Y) are defined as:

$$QSMCP(X, Y) = \begin{Bmatrix} MCP(X, Y) \cup V, if\, Mis(X) \neq V \text{ and } Mis(Y) \neq V \\ MCP(X, Y), otherwise \end{Bmatrix}.$$

According to Definition 4.2, query strings of mismatched segments, *Mis(T)* and *Mis(H)*, and the MCPs $MCP_{and}(T, H)$ and $MCP_{and}(T, H)$, are produced as *QSMis(T)*, *QSMis(H)*, $QSMCP_{and}(T, H)$ and $QSMCP_{and}(H, T)$ respectively; all the query strings are thrown into a search engine to get the number of found Web pages. The availability value AV_{and} is proposed to estimate the prevalence of mismatch conjunctions. The definition of availability value is as follows.

Definition 4.3 *(Availability Value, AV)*. Assume Mis(X) and Mis(Y) represent the mismatched segments in sentence X and sentence Y. $QSMCP_{and}(T, H)$ and $QSMCP_{and}(H, T)$ are the query strings of $MCP_{and}(T, H)$, $MCP_{and}(T, H)$. Then the availability value of Mis(X), Mis(Y) and AND conjunction is defined as:

$$AV_{and}(Mis(X), Mis(Y)) = \frac{Max(HitNum(QSMCP_{and}(X, Y)), HitNum(QSMCP_{and}(X, Y)))}{HitNum(QSMis(X)) * HitNum(QSMis(Y))},$$

where HitNum(X) is the number of matched Web pages with Web query string X.

As just mentioned, MCPs with OR conjunctions should also be produced. We compute the availability value of "or" MCPs as

$$AV_{or}(Mis(X), Mis(Y)) = \frac{Max(HitNum(QSMCP_{or}(\mathrm{X}, \mathrm{Y})), HitNum(QSMCP_{or}(\mathrm{X}, \mathrm{Y})))}{HitNum(QSMis(\mathrm{X})) * HitNum(QSMis(\mathrm{Y}))},$$

where $QSMCP_{or}$(X, Y) and $QSMCP_{or}$(X, Y) are appended from optional MCPs MCR_{or}(X, Y) and MCR_{or}(X, Y).

Low availability value may indicate unreasonableness. The difference between AV_{and} and AV_{or} show the appearance tendency of two mismatch segments combined in one single conjunction phrase, and point out the possible semantic relation between two mismatch segments. MCPs with higher AV_{and} and lower AV_{or} are rarely regarded as containing two compatible mismatched segments; MCPs with lower AV_{and} and higher AV_{or}, on the contrary, may merge two contradictory statements into one phrase. For those MCPs combining two synonyms, with completely unrelated items, and with long, complicated mismatches, it is highly possible to get zero or an extremely low number of both AV_{and} and AV_{and} since people do not tend to put these statements together into a conjunction phrase. Thus, an empirical threshold τ is predefined. If AV $< \tau$, the TH pair will be considered as either having synonymy mismatches or being unrelated. We use the following describe the determination rule of contradictions.

$$Mis(T) \text{ and } Mis(H) \text{ are Contradictory } \Leftrightarrow AV_{or} > AV_{and}, \text{and } AV_{or} > \tau$$

Hence, the process of MCP-based contradiction detection contains the following steps.

(1) Extract critical mismatch segments Mis(T) and Mis(H) from the TH pair.
(2) Generate $\mathrm{MCP_{and}}$(T, H), $\mathrm{MCP_{and}}$(H, T), $\mathrm{MCP_{or}}$(T, H), and $\mathrm{MCP_{or}}$(H, T) from Mis(T) and Mis(H)
(3) Use all MCPs along with the verb V to form $QSMCP_{and}(T, H)$, $QSMCP_{and}(H, T)$ and $QSMCP_{or}(T, H)$ and $QSMCP_{or}(H, T)$ for Web query.
(4) Obtain $\mathrm{AV_{and}}$(Mis(T), Mis(H)) and $\mathrm{AV_{or}}$(Mis(T), Mis(H)) from the Web.
(5) If $\mathrm{AV_{and}} = 0$ and $\mathrm{AV_{or}} = 0$, eliminate V in all QSMCPs and go to step 4.
(6) Examine $\mathrm{AV_{and}}$ and $\mathrm{AV_{or}}$ by the determination rule for contradiction to assign the result.

4.4. Threshold Determination

In order to determine the threshold τ used in the availability value filtering, we manually created 113 Chinese word pairs to simulate critical mismatched segments of TH pairs. These simulated mismatch segments, including 48 contradictive pairs and other non-contradictive pairs, were used to generate sample MCPs. Thereafter, we applied MCP-based Web queries on these sample MCPs, and the contradiction detection accuracy on these simulated sample MCPs was optimized by tuning the value of threshold τ. The optimized threshold τ is applied in the following MCP experiments.

We found that while the threshold τ was set at 5×10^{-15}, the recall, precision, and F-score of contradiction detection on these simulated mismatch pairs achieved a maximum of 70.83% (34/48), 72.34% (34/47), and 71.57% respectively. Table III shows examples of the simulated mismatch segments pairs, their part of speech tags, and the contradictiveness.

5. IASLD – THE BASELINE SYSTEM

The architecture of our baseline RTE system [Shih et al. 2011], as shown in Figure 1, is introduced in this section. The key points of the system, includes

Table III. Examples of Statement Pairs for Determining the Threshold

Pairs	POS	Is this a contradiction?
偉大(great), 渺小(lowly)	Adjective	Yes
增加(increase), 減少(decrease)	Verb	Yes
長處(advantage), 短處(shortcoming)	Noun	Yes
利用(utilize), 使用(use)	Verb	No
能夠(be able to), 可以(can)	Adverb	No
公司(company), 他們(they)	Noun	No

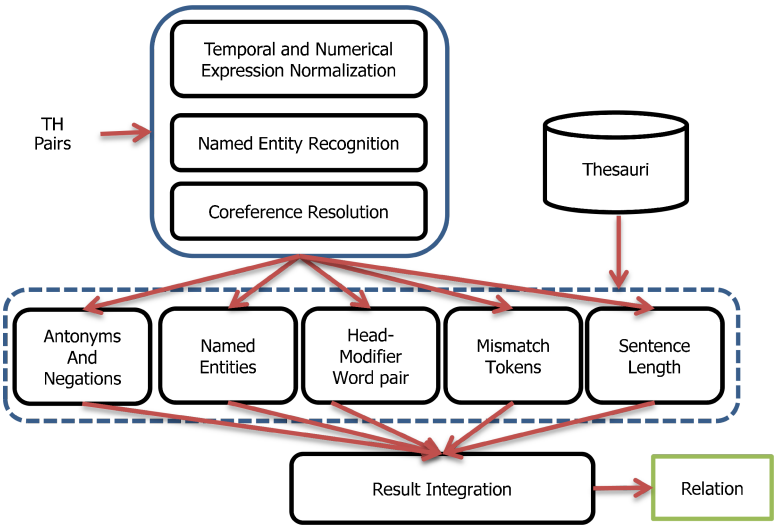

Fig. 1. Architecture of the baseline TE system.

preprocessing, entailment determination, and result integrating strategy are introduced below.

5.1. Preprocessing

Three preprocessing steps are required before the TH pair judgment. These preprocesses, which include normalization of temporal and numerical expressions, named entity recognition, and co-reference resolution, are designed for reducing the complexity.

(1) Normalization of temporal and numerical expressions. The variations of temporal and numerical expressions are very common in Chinese texts. As a result, we collected some frequent temporal and numerical expressions patterns, and designed a normalizing algorithm by using these patterns. All the temporal and numerical sequences are identified by these predefined patterns, and converted into digits according to manually-constructed transformation rules. For example, a sentence such as "*一九九九年十二月十日*" (December 10, 1999) is converted as "*1999 年 12 月 10 日*". According to the TE recognition results, approximately 2.1% (9/421) of the TH pairs became equal after the normalization. Table IV shows the selected temporal and numerical expression types and their examples.

Table IV. Examples of Frequent Patterns and Their Types Used in Temporal and Numerical Expression Normalization. These Patterns Are Stated in Regular Expressions

Regular Expressions of the Patterns	Type
(西元|公元)*[(possible digit characters)]{2,4}年	Years
[(possible digit characters)]{1,2}月	Month
(百分之|千分之)[(possible digit characters)]{1,}	Percentage
[(possible digit characters)\\.]{1,}%	Percentage
[(possible digit characters)]{1,}分之[(possible digit characters)]{1,}	Fraction
(西元|公元)*[(possible digit characters)]{2,4}[年]{0,1}(到|至|\\-|\\～)[(possible digit characters)]{2}年	Duration

(2) Named entity recognition. In our RITE participating system, we developed two different modules for named entity recognition (NER). The first module, which is statistical-based, used a training set from the SIGHAN NER share task [Levow 2006] and CRF++, a machine-learning toolkit. Another module, which depends on manually generated rules and context templates, located NEs by pattern matching. We integrated these two modules into a hybrid Chinese named entity recognition system to identify entities such as person, location, organization, temporal and numerical expressions, and certain artifacts such as tools and animals. The Chinese named entity corpus (CNEC) [Shih et al. 2004] and the test set from SIGHAN NER share task were used for evaluation. The result shows that our NER module achieved 73% and 88.6% in *F*-score, respectively [Wu et al. 2006].
(3) Coreference and Zero-Pernoun resolution. Coreference resolution is defined as finding the anaphoric noun phrases in the text that refer to the same real-world entity [Ng 2008]. However, unlike English, a prominent phenomenon in Chinese coreference resolution is the higher frequency of pronoun dropping. According to Zhao and Ng [2007], the dropping of pronouns, which is called *zero pronouns*, is much more prevalent in Chinese and poses a unique challenge in coreference resolution for Chinese. In IASLD, we use a parsing based approach to eliminate coreference and zero pronouns as much as possible.

5.2. Entailment Determination

Features such as NE, mismatch tokens, head-modifier word pairs, and sentence length are used independently to determine the relation of TH pairs.

(1) Named entities. We set a rule that in TH pair, a sentence with more NE means that it contains more information, and is likely to entail another sentence.
(2) Mismatch tokens. We use the CKIP Chinese segmentation tool [Ma and Chen 2003] to tokenize TH pairs and see which token is located in mismatches. Thesauri such as E-hownet [Chen et al. 2005], the Chinese concept dictionary [Yu and Yu 2001], Tongyichichilin [梅家驹 1983], and Sinica Bow [Huang et al. 2004] are used to compute the similarity of mismatch tokens between TH pairs. Thereafter, a simple rule is used to derive the entailment relations:
Assume T and H have mismatch token lists $MT_T = \{t_T1, t_T2, \ldots t_Tm\}$ and $MT_H = \{t_H1, t_H2, \ldots t_Hn\}$ respectively. The semantic relation labels SR between two words SR = {synonym, hypernym, hyponym, holonym, entailment, ... etc}, then the

Algorithm 1. Head-modifier word pair based textual entailment recognition

Input: word pair $\mathbf{WP_T} = \{\mathbf{h_T}, \mathbf{m_T}\}$ and $\mathbf{WP_H} = \{\mathbf{h_H}, \mathbf{m_H}\}$ where $\mathbf{h_T}, \mathbf{m_T} \in \mathbf{T}$ and $\mathbf{h_H}, \mathbf{m_H} \in \mathbf{H}$
Output: The word pair entailment relation ($\mathbf{R_{WP_TWP_H}}$)· gets assigned.
for each word pair $\mathbf{WP_T} = \{\mathbf{h_T}, \mathbf{m_T}\}$·
  for each word pair $\mathbf{WP_H} = \{\mathbf{h_H}, \mathbf{m_H}\}$
    if $\mathbf{ED_{h_Th_H}}$ **= Bidirectional**
      then The relation $\mathbf{R_{WP_TWP_H}}$ between $\mathbf{WP_T}$ and $\mathbf{WP_H}$ are assigned as $\mathbf{ED_{m_Tm_H}}$;
    else if $\mathbf{ED_{m_Tm_H}}$**= Bidirectional**
      then The relation $\mathbf{R_{WP_TWP_H}}$ between $\mathbf{WP_T}$ and $\mathbf{WP_H}$ are assigned as $\mathbf{ED_{h_Th_H}}$
    else
      then The relation $\mathbf{R_{WP_TWP_H}}$ between $\mathbf{WP_T}$ and $\mathbf{WP_H}$ are assigned as Independent
  end
end

entailment direction $ED_{t_Tt_H}$ between each two mismatch tokens t_T and t_H is defined as the following.

$$\mathrm{ED}_{t_Tt_H} = \left\{ \begin{array}{c} \text{bidirectional, if } \mathrm{SR}_{t_Tt_H} = \text{synonym} \\ \text{forward, if } \mathrm{SR}_{t_Tt_H} = \text{hypernym|Entailment} \\ \text{reverse, if } \mathrm{SR}_{t_Tt_H} = \text{hyponym} \\ \text{independent, if } \mathrm{SR}_{t_Tt_H} = \text{otherwise} \end{array} \right\}$$

The entailment relation of T and H is the relation supported by the most mismatch tokens.

(3) Head-modifier word pair. CKIP Chinese parser [You and Chen 2004] is used to extract all the head-modifier relations between the tokens in IASLD. Meanwhile, a part-of-speech filter is manually constructed to remove word pairs with stopwords and function words. Only five POS combinations are allowed to form a head-modifier word pair: noun-verb, verb-noun, adjective-noun, adverb-verb, and noun-noun. A predefined algorithm is used to capture the semantic relation by checking the head-modifier word pairs. This algorithm is described in Algorithm 1.

(4) Sentence length. Sentence length is a powerful feature in the RITE dataset. Taking advantage of sentence length greatly increases the chance to obtain correct results, especially for TH pairs with a large difference in length. We decided against considering the literal length and instead used the number of tokens for the notion that the number difference of tokens may indicate the distribution of information more precisely. The rule of sentence length based recognition is presented as follows:

Assume N_T and N_H represent the token number of T and H respectively, then the relation of sentence length module $R_{T,H}$ is as follows.

$$\mathrm{R}_{\mathrm{T,H}} = \left\{ \begin{array}{c} \text{Forward, if } \mathrm{N_T} > \mathrm{N_H} \\ \text{Reverse, if } \mathrm{N_T} < \mathrm{N_H} \\ \text{Bidirectional, if } \mathrm{N_T} = \mathrm{N_H} \end{array} \right\}$$

5.3. Result Integration

We integrate the results from different features by a voting strategy. The relation with the most feature supports wins the integration. Furthermore, various heuristics are used as tie-breakers to deal with TH pairs with relations of more than one winner.

Table V. The Results of TE Recognition and Contradiction Detection on a RITE TC MC Development Set with Manually-Annotated and Automatic-Extracted Mismatched Information

	Accuracy	**Contradiction Precision**	**Contradiction Recall**	**Contradiction *F*-Score**
Manually annotated	78.05% (64/82)	66.67% (16/24)	59.26% (16/27)	62.7%
Automatically extracted	68.67% (57/82)	45% (9/20)	33.33% (9/27)	38.3%
Difference	9.38%	21.67%	25.93%	24.4%

6. EXPERIMENTS AND RESULTS

We designed two different experiments to estimate the effectiveness of MCP querying for contradiction detection. First, a TE corpus with manually annotated mismatch information was used to evaluate how well the system detects contradiction in real TE data. Second, we tested our MCP-based contradiction detection system on two different RITE datasets to estimate its impact on the performance of our previous RTE system.

6.1. Experiment 1: TE Corpus with Manually-Annotated and Automatic-Extracted Mismatched Information

This experiment was designed to observe the impact of failed critical mismatched recognition on our MCP-based system's performance using a real TE dataset. First, we manually constructed a TE corpus in which all critical mismatches in TH pairs were annotated. Our annotators checked 421 TH pairs in RITE TC MC development set and labeled both critical mismatches and the verb in the selected 82 TH pairs. For instance, the mismatch segments and the related verb of Example (2) and (3) are manually annotated as "*獲得*" (obtain), "*18座*" (18), "*58座*" (58) and "*印度*" (India), "*宿敵*" (old enemy), "*好朋友*" (good friend). Secondly, we developed a critical mismatch extraction module that uses the CKIP Chinese parser [You and Chen 2004] to obtain the grammatical dependences among all the tokens in T and H. We checked all tokens T_T from T and t_H from H to see if both t_T and t_H referred to a common token $ct_{T,H}$. All token pairs (t_T, t_H) that referred to $ct_{T,H}$ and that passed the examination can be regarded as critical mismatch segments between T and H. MCP-based contradiction detection was carried out on these selected TH pairs, and the results of using manually annotated critical mismatch information were compared to the results with automatically extracted mismatches. The evaluation results are shown in Table V.

6.2. Experiment 2: Contradiction Detection on a Different RITE Dataset

In this experiment, we examined the impact on performance by adding MCP-based contradiction detection to our NTCIR-9 TE system. According to Shima et al. [2011], the RITE TC development set was mostly created from the NTCIR-7 CCLQA dataset, whereas the RITE TC test set was created from two different sources: the NTCIR-8 CCLQA answer set and the relevant documents retrieved in the past NTCIR CLIR tasks. We use these two different datasets to measure the effectiveness of the MCP-based approach on contradiction detection. A classifier which is based on a Chinese parser was implemented to filter out TH pairs with multiple, complicated mismatches. Tables VI and VII present the comparison between the baseline system and the improved system (baseline + contradiction detection) on both RITE TC development sets and test sets.

MCP-based approach yields large contradiction detection improvement in both development and test data, especially on recall. It may indicate that some implicit incompatibilities in TH pairs were successfully indentified by our approach.

Table VI. The Results of Contradiction Detection on Selected TH Pairs from RITE TC MC Development Sets

	Accuracy	Contradiction Precision	Contradiction Recall	Contradiction *F*-Score
Baseline System [Shih et al. 2011]	39.13% (18/46)	50% (2/4)	18.18% (2/11)	26.66%
Baseline+ Contradiction Detection	52.17% (24/46)	53.33% (8/15)	72.72% (8/11)	61.53%
Improvement	+13.04%	+3.33%	+54.54%	+34.87%

Table VII. The Results of Contradiction Detection on Selected TH Pairs from RITE TC MC Test Sets

	Accuracy	Contradiction Precision	Contradiction Recall	Contradiction *F*-Score
Baseline System [Shih et al. 2011]	39.75% (33/83)	44.44% (4/9)	15.38% (4/26)	22.85%
Baseline+ Contradiction Detection	45.78% (38/83)	57.89% (11/19)	42.3% (11/26)	48.88%
Improvement	+6.03%	+13.45%	+32.95%	+26.03%

Table VIII. Contradiction Detection Comparison on Different Linguistic Phenomena in RITE CT MC Pairs

Phenomena		Total # of Contradictive TH pairs	Identified by Baseline System (Shih et al., 2011)	Identified by Baseline + Contradiction Detection
Lexicon	Identity	10	4	6 (+2)
	Format	8	4	4 (+0)
	Demonym	6	2	2 (+0)
	Synonym	3	3	3 (+0)
	Antonym	10	1	1 (+0)
Syntactic	Negation	11	8	8 (+0)
	Modifier	10	3	5 (+2)
	List	2	2	2 (+0)
Discourse	Apposition	6	0	0 (+0)
	Coreference	1	1	1 (+0)
Reasoning	Anaphora-Zero	5	3	3 (+0)
	Temporal	9	3	4 (+1)
	Spatial	1	0	0 (+0)
	Quantity	4	2	2 (+0)
	Modifier	3	1	2 (+1)
	Apposition	9	1	1 (+0)
	General-inference	15	4	5 (+1)
	Background Knowledge	19	7	11 (+4)

6.3. Experiment 3: Contradiction Detection on Different Linguistic Phenomena

The coverage of our MCP-based contradiction detection approach on different linguistic phenomena is examined in this experiment. Two configurations of contradiction detection, including baseline system and the improved system (baseline + contradiction detection), are applied on all the categorized contradictive TH pairs in Table I. The comparison of these two contradiction detection system on different linguistic phenomena is shown in Table VIII.

7. DISCUSSIONS AND FUTURE WORKS

Based on the MCP-based contradiction detection, most of the background knowledge shortage problems are solved. For instance, the conflicts between "*關注*" (concerned)

and "*漠不關心*" (unconcerned), "*膨脹*" (swell) and "*緊縮*" (tighten), "*共同*" (together) and "*各自*" (individually) can be correctly identified by our approach. However, three issues are worth noticing. These three issues of potential exceptionable web contents, zero-hit MCPs, and critical extractions of mismatches, are discussed in this section, in addition to some interesting findings.

7.1. Potential Exceptionable Web Contents for MCP-Based Contradiction Detection

Four kinds of exceptionable Web contents decrease the performance of MCP-based contradiction detection.

(1) *Mismatch segments with similar but not identical meaning*. If the AV of an OR conjunction phrase is greater than that of its AND conjunction phrase, then the pair is recognized as contradictive. This affects mismatch segment pairs such as "*覺得*" (feel) and "*認為*" (think) with similar but not identical meanings. Such words that are not actually contradictive might fail in the MCP-based contradiction detection.
(2) *Intended uncommon word usage*. Similar detection mistakes happen to strange statements such as "*是爸爸還是父親*" (Dad or father). Different from the previous case, this kind of statement is mostly created as intended.
(3) *Oxymoron statement*. The MCPs of some antonym pairs, such as "*美麗又醜陋*" (beautiful and ugly) and "*沉重又輕鬆*" (heavy lightness), can be found online on many Web pages. The frequent use of AND conjunctions to join these antonym pairs, which are called oxymorons, can skew the MCP-based contradiction identification model causing false negatives.
(4) *Misaligned Chinese word segmentation*. The misaligned Chinese word segmentation may confuse the MCP based approach. For instance, the MCP "*強和顯著*" (the most powerful and most noticeable) in Example (7) may find some misaligned Web contents such as "*附著力強和顯著的熱反射性*" (strong adhesive force and good heat reflectivity),in which the critical mismatches "*強*" (strong) and "*顯著*" (good) are not the constituents of a single conjunction phrase but used to describe totally different entities.

7.2. Unexpected Zero-Hit Mismatch Conjunction Phrases

Zero-hit MCPs are phrases for which both MCP_{and} and MCP_{or} fail to get any matched results from search engine. We have observed that such MCPs often contain synonyms, unrelated words, or complicated statements. Unexpected zero-hit MCPs severely reduce the performance of MCP-based contradiction detection, especially on system recall. We propose two reasons for the occurrence of unexpected zero-hit MCPs.

(1) MCP with long mismatches. As the length of an MCP and/or the mismatches it contains increase, the chances of finding search snippets which match the sequences drastically decrease. According to our observations, if the mismatch segments length is larger than three characters, the probability of getting zero-hit query results dramatically increase.
(2) MCPs with atypical syntax. Automatically generated MCPs with atypical syntax may receive zero search hits. In Chinese, some words are unsuited to be directly connected by conjunctions. In order to solve this problem, functional prefixes and suffixes are used in the MCP generated process. Currently, although the selection of functional prefixes and suffixes depends on the POS of the mismatch segments,

it cannot completely avoid the occurrence of inappropriate phrase merging. Moreover, the number of candidate functional prefixes and suffixes are limited. For now, the way to constantly produce fluent MCPs is a crucial point of MCP-based process that needs further development.

7.3. Critical Mismatch Extraction

From the results of Experiment 1, we can see the influence of mismatch extraction error to MCP-based contradiction. Our token-dependency based extraction process achieves an acceptable result in identifying critical mismatches between TH pairs. But there is still more than 20% room in the F-score for improvement. Our present MCP-based contradiction detection system is limited to TH pairs with only one mismatch. In fact, a large number of TH pairs in RITE have more than one mismatch. However, since most mismatches in TH pairs are irrelevant (synonyms, function words, and appositions), before increasing the number of mismatches our system can handle, we have to design a way of filtering out irrelevant mismatches.

7.4. Threshold of Availability Value

We found that a large number of the false positive cases in both perfect or extracted mismatch configuration can be correctly identified by slightly shifting the threshold τ. This fact may indicate that the threshold value which is optimized to the simulated mismatch segments pairs in Section 4.4 is not suitable for real TE data. Also, despite the difficulty of acquiring resources with annotated contradictive statements, these finding points to a future research direction of using large corpora or Web content as bases to determine a more fitting threshold for MCP-based contradiction detection.

7.5. Other Findings and Research Direction

(1) According to the results of all the tests, we did not observe significant accuracy difference among mismatches of different POS types. This indicates that using MCPs to detect contradictive statements can work on different POS types as long as the corresponding function prefixes and suffixes are properly assigned in the MCP generating process.
(2) Both traditional Chinese and simplified Chinese Web pages can be used to compute the availability value of MCPs. In fact, the behavior of conjunction phrases is almost the same in both kinds of Chinese corpus.
(3) Although the proposed approach and the experiments are only applied in Chinese, we believe that the idea of using conjunctions to capture contradictions in text is feasible. Our work and some previous researches, such as Lin et al. [2003], may indicate that the problem of how to use specific language behavior to acquire useful knowledge from unprocessed data or resources is research-worthy.

8. CONCLUSION

Using Web mining for deriving information and knowledge is common in natural language processing. But few previous works apply Web mining to obtain the required background knowledge for contradiction detection in the textual entailment framework. The mismatched conjunction phrases, which consist of critical mismatches in TH pairs, provide a novel methodology for acquiring implicit knowledge from the Web, and have been proven to be a reliable feature reflecting the existence of incompatibilities between two sentences. In addition, by using Web-based checking, the shortage of background knowledge between words or phrases in entailment and contradiction judgment may receive a strong supplement.

Our work can be considered pioneering in the use of Web content for dealing with the problems of textual entailment and contradiction detection. However, the need of extended work and enhancements are addressed in the discussion section. We hope the points we emphasize in this paper can offer some interesting and useful points to contradiction detection researchers and, more generally, text understanding.

REFERENCES

ACADEMIA SINICA. Research group of Chinese knowledge and information processing, 1993. Chinese part-of-speech analysis: Tech. rep. no. 93–05.

BENTIVOGLI, L., CABRIO, E., DAGAN, I., GIAMPICCOLO, D., LEGGIO, M. L., AND MAGNINI, B. 2010. Building textual entailment specialized data sets: A methodology for isolating linguistic phenomena relevant to inference. In *Proceedings of the 7th Annual Conference on Language Resources and Evaluation (LREC'10)*.

CASTILLO, J. 2009. Sagan in TAC2009: Using Support Vector Machines in Recognizing Textual Entailment and TE Search Pilot task. In *Proceedings of the 2nd Text Analysis Conference (TAC'09)*.

CLARK, P. AND HARRISON, P. 2008. Recognizing textual entailment with logical inference. In *Proceedings of the 1st Text Analysis Conference (TAC'08)*.

CHEN, K. J., HUANG, S. L., SHIH, Y. Y., AND CHEN, Y. J. 2005. Extended HowNet - A representational framework for concepts. In *Proceedings of Ontologies and Lexical Resources (IJCNLP'05)*.

CROUCH, D., CONDORAVDI, C., PAIVA, V. D., STOLLE, R., AND BOBROW, D. G. 2003. Entailment, intensionality and text understanding. In *Proceedings of the Human Language Technology Conference/North American Chapter of the Association for Computational Linguistics Workshop on Text Meaning (HLT-NAACL'03)*.

CURZAN, A., AND ADAMS, M. P. 2005. *How English Works: A Linguistic Introduction.* Pearson Longman, New York, 152.

DAGAN, I., GLICKMAN, O., AND MAGNINI, B. 2005. The PASCAL recognizing textual entailment challenge. In *Machine Learning Challenges*, J. Quiñonero-Candela, I. Dagan, B. Magnini, and F. d'Alché-Buc Eds., Lecture Notes in Computer Science, vol. 3944. Springer, 177–190.

FELLBAUM, C. 1998. *Wordnet: An Electronic Lexical Database*. MIT Press, Cambridge, Mass.

GIAMPICCOLO, D. AND MAGNINI, B. 2007. The 3rd PASCAL recognizing textual entailment challenge. In *Proceedings of the ACL-Pascal Workshop on Textual Entailment and Paraphrasing (ACL'07)*. 1–9.

HARABAGIU, S., HICKL, A., AND LACATUSU, F. 2006. Negation, contrast and contradiction in text processing. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI'06)*.

HUANG, C. R., CHANG, R. Y., AND LEE, S. B. 2004. Sinica BOW (Bilingual Ontological Wordnet): Integration of bilingual WordNet and SUMO. In *Proceedings of the Annual Conference on Language Resources and Evaluation (LREC'04)*. 1553–1556.

IFTENE, A. 2008. UAIC participation at RTE4. In *Proceedings of the 1st Text Analysis Conference (TAC'08)*.

IFTENE, A. AND MORUZ, M. 2009. UAIC participation at RTE5. In *Proceedings of the 1st Text Analysis Conference (TAC'09)*.

KIPPER-SCHULER, K. 2005. VerbNet: A broad-coverage, comprehensive verb lexicon. Ph.D. thesis, Computer and Information Science Department, University of Pennsylvania, Philadelphia, PA.

LEVOW, G. 2006. The 3rd International Chinese Language Processing Bakeoff: Word Segmentation and Named Entity Recognition. In *Proceedings of SIGHAN5 the 3rd International Chinese Language Processing Bakeoff at COLING/ACL (ACL'06)*. 108–117.

LIN, D. AND PANTEL, P. 2001. Discovery of inference rules for question answering. *Nat. Lang. Eng.* 7, 4, 343–360.

LIN, D., ZHAO, S., QIN, L., AND ZHOU, M. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI'03)*. 1492–1493.

MA, W. Y. AND CHEN, K. J. 2003. Introduction to CKIP Chinese Word Segmentation System for the First International Chinese Word Segmentation Bakeoff. In *Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing (ACL'03)*. 168–171.

MAGNINI, B. AND CABRIO, E. 2009. Combining specialized entailment engines. In *Proceedings of LTC'09*.

MAGNINI, B. AND CABRIO, E. 2010. Contradiction-focused qualitative evaluation of textual entailment. In *Proceedings of the Workshop on Negation and Speculation in Natural Language Processing (NeSpNLP'10)*. 86–94.

MARNEFFE, M. C. D, RAFFERTY, A. N., AND MANNING, C. D. 2008. Finding contradictions in text. In *Proceedings of Human Language Technology Conference (HLT'08)*.

MONZ, C. AND RIJKE, M. D. 2001. Lightweight entailment checking for computational semantics. In *Proceedings of the 3rd Workshop on Inference in Computational Semantics (WICS'01)*.

NG, V. 2008. Unsupervised models of co-reference resolution. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*. 640–649.

PADO, S., DE MARNEFFE, M., MACCARTNEY, B., RAFFERTY, A., YEH, E., AND MANNING, C. 2008. Deciding entailment and contradiction with stochastic and edit distance-based alignment. In *Proceedings of the 1st Text Analysis Conference (TAC'08)*.

RITTER, A., DOWNEY, D., SODERLAND, S., AND ETZIONI, O. 2008. It's a contradiction—No, it's not: A case study using functional relations. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'08)*. 11–20.

SHIH, C. W., TSAI, T. H., WU, S. H., HSIEH, C. C., AND HSU, W. L. 2004. The construction of a Chinese named entity tagged corpus: CNEC1.0. In *Proceedings of the Research on Computational Linguistics Conference (ROCLING'04)*.

SHIH, C. W., LEE, C. W., YANG, T. H., AND HSU, W. L. 2011. IASL RITE system at NTCIR-9. In *Proceedings of the 9th NII Test Collection for Information Retrieval Workshop (NTCIR'11)*.

SHIMA, H., KANAYAMA, H., LEE, C. W., LIN, C. J., MITAMURA, T., MIYAO, Y., SHI, S., AND TAKEDA, K. 2011. Overview of the NTCIR-9 RITE: Recognizing inference in TExt. In *Proceedings of the 9th NII Test Collection for Information Retrieval Workshop (NTCIR'11)*.

VOORHEES, E. M. 2008.Contradictions and justifications: Extensions to the textual entailment task. In *Proceedings of the Association for Computational Linguistics (ACL'08)*. 63–71.

WU, C. W., JAN, S. Y., TSAI, T. H., AND HSU, W. L. 2006. On using ensemble methods for Chinese named entity recognition. In *Proceedings of the Association of Computer Linguistics Special Interest Group on Chinese Language Processing Workshop (ACL-SIGHAN'06)*.

YOU, J. M. AND CHEN, K. J. 2004. Automatic semantic role assignment for a tee structure. In *Proceedings of the Association of Computer Linguistics Special Interest Group on Chinese Language Processing Workshop (ACL-SIGHAN'04)*.

YU, J. S. AND YU, S. W. 2001. Introduction to Chinese concept dictionary. In *Proceedings of the International Conference on Chinese Computing (ICCC'01)*. 361–367.

ZHAO, S. H. AND NG, H. T. 2007. Identification and resolution of Chinese zero pronouns: A machine learning approach. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP'07)*. 541–550.

梅家驹, 竺一鸣, and 高蕴琦. 1983. *Tongyichichilin*. Shanghai Lexicographical Publishing House, Shanghai.

Received May 2012; revised July 2012; accepted September 2012