



ELSEVIER

Contents lists available at [ScienceDirect](http://www.sciencedirect.com)

Information Processing and Management

journal homepage: www.elsevier.com/locate/infoproman

Using compositional semantics and discourse consistency to improve Chinese trigger identification [☆]

Pei-Feng Li, Qiao-Ming Zhu, Guo-Dong Zhou ^{*}

School of Computer Science & Technology, Soochow University, 215006 Suzhou, China

ARTICLE INFO

Article history:

Received 23 October 2012

Received in revised form 11 October 2013

Accepted 18 October 2013

Available online 20 November 2013

Keywords:

Chinese event extraction

Trigger identification

Compositional semantics

Morphological structure

Head morpheme

Discourse consistency

ABSTRACT

Due to the special characteristics and challenges in Chinese language, event extraction in Chinese is much more difficult than that in English. In particular, the state-of-the-art Chinese event extraction systems suffer much from the low recall in trigger identification due to the failure in identifying unknown triggers and the inconsistency in identifying trigger mentions. To resolve these two issues, this paper proposes an inference mechanism to infer unknown triggers via the compositional semantics inside Chinese words and another inference mechanism to recover trigger mentions via the discourse consistency between Chinese trigger mentions. Here, various morphological structures are explored to better represent the compositional semantics inside Chinese triggers and automatically identify the head morpheme as the governing sememe of a trigger in inferring unknown triggers. Evaluation on the ACE 2005 Chinese corpus justifies the effectiveness of our approach over a strong baseline in Chinese event extraction, in particular trigger identification.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

As a compromise to natural language understanding, Information Extraction (IE) aims to extract structured information (e.g., entities, relations and events) from a text. As a classic IE task, event extraction is to identify instances of a predefined event type and can be typically divided into four components: trigger identification, event type classification, argument identification and argument role classification. In the literature, most studies focus on English event extraction (e.g., [Ahn, 2006](#); [Finkel, Grenager, & Manning, 2005](#); [Grishman, Westbrook, & Meyers, 2005](#); [Hardy, Kanchakouskaya, & Strzalkowski, 2006](#); [Hong et al., 2011](#); [Ji & Grishman, 2008](#); [Liao & Grishman, 2010](#); [Llorens, Saquete, & Navarro-Colorado, 2013](#); [Lu & Roth, 2012](#); [Maslennikov & Chua, 2007](#); [Patwardhan & Riloff, 2007](#); [Patwardhan & Riloff, 2009](#)).

In comparison, there are few successful stories regarding Chinese event extraction due to the special characteristics in Chinese trigger identification. In particular, there are two major sources for the low performance (particularly the recall): the occurrence of unknown triggers¹ and the lack of sentence-level information.

For the occurrence of unknown triggers, [Table 1](#) gives the statistics in both the ACE (Automatic Content Extraction) 2005 Chinese and English corpora² using 10-fold cross-validation. In each validation, we leave 10% trigger mentions as the test set and the remaining ones as the training set. It shows that unknown triggers cover about 21% of Chinese trigger mentions while

[☆] This paper is a much extended version of [Li et al. \(2012\)](#) at EMNLP-CoNLL 2012.

^{*} Corresponding author. Tel.: +86 512 65214851; fax: +86 512 65243205.

E-mail addresses: pfli@suda.edu.cn (P.-F. Li), qmzhu@suda.edu.cn (Q.-M. Zhu), gdzhou@suda.edu.cn (G.-D. Zhou).

¹ If a trigger in the test set doesn't occur in the training set, we regard it as an unknown trigger. Otherwise, a known trigger.

² The whole Chinese ACE corpus has 3332 event mentions. For the sake of fair comparison, we randomly choose the same number of event mentions from the English corpus as the cross-validation data.

Table 1
Statistics of unknown triggers: a comparison between Chinese and English.

Language	Chinese	English
% unknown trigger types (mentions)	33.7% (20.9%)	18.5% (8.9%)
# trigger types (mentions)	763 (3332)	586 (3332)

this figure reduces to only about 9% in English. It also shows that given the same number of event mentions, there are about 30% more different trigger types in Chinese than that in English. This justifies the low performance (particularly the recall) of a Chinese event extraction system, which normally extracts those known trigger mentions occurring in the training data as candidate instances and uses a classifier to distinguish true trigger mentions from pseudo ones.

For the lack of sentence-level information, some event mentions fail to provide enough intra-sentential information to the classifier and can be only inferred confidently from the discourse-level information, in particular for Chinese event extraction, especially when the word triggering an event is polysemous or an unknown trigger. Kim (2000) compares the use of overt subjects in English and Chinese, and finds out that overt subjects occupy over 96% in English, while this figure drops to only 64% in Chinese. Our statistics in the ACE 2005 Chinese corpora shows that almost 55% event arguments are missing and this imposes extra difficulty in Chinese event extraction.

To resolve above two issues, this paper proposes two novel inference mechanisms to Chinese trigger identification by employing the compositional semantics inside Chinese triggers to infer unknown triggers and imposing the discourse consistency between Chinese trigger mentions to recover trigger mentions.

The first mechanism is motivated by the compositional nature of Chinese words, whose semantics can be often determined by the component morphemes. Hence, it is natural to infer unknown triggers by employing the compositional semantics inside Chinese triggers. Take following two sentences as examples:

-
- (E1) 4名学生被玻璃划伤/4-Ming-Xue-Sheng-Bei-Bo-Li-Hua-Shang/
(Four students were **scratched** by the glass.) –Known trigger
- (E2) 1名乘客被刺伤/1-Ming-Cheng-Ke-Bei-Chi-Shang/
(A passenger was **stabbed**.) –Unknown trigger
-

where “划伤/Hua-Shang/” is a known trigger and “刺伤/Chi-Shang/” is an unknown one.

In the above examples, the semantics of “划伤/Hua-Shang/” (injure by scratching) in (E1) can be largely determined from those of “划/Hua/” (scratch) and “伤/Shang/” (injure) while the semantics of “刺伤/Chi-Shang/” (injure by stabbing) in (E2) from those of “刺/Chi/” (stab) and “伤/Shang/” (injure). Since these two triggers have similar morphological structures, we can easily infer that “刺伤/Chi-Shang/”(injure by stabbing) in (E2) is a trigger if “划伤/Hua-Shang/”(injure by scratching) in (E1) is known as a trigger. Similarly, we can infer more triggers in the event type *injure*, such as “灼伤/Zhuo-Shang/” (injure by burning), “撞伤/Zhuang-Shang/” (injure by hitting), “压伤/Ya-Shang/” (injure by pressing), etc.

The second mechanism is enlightened by the wide existence of the discourse consistency in natural languages, particularly for Chinese, due to its discourse-driven nature (Zhu, 1980). Very often, distinguishing true trigger mentions from pseudo ones is only possible with discourse-level information. Take following two contingent sentences as examples:

-
- (E3) 美国与北韩3号在吉隆坡结束飞弹会谈/Mei-Guo-Yu-Bei-Han-3-Hao-Zai-Ji-Long-Po-Jie-Shu-Fei-Dan-Hui-Tan/
(The United States and the Democratic People’s Republic of Korea finished missile **talks** in Kuala Lumpur.)
- (E4) 会谈的气氛严肃/Hui-Tan-De-Qi-Fen-Yan-Su/
(The **talks** are serious.)
-

While it is easy to determine that mention “会谈/Hui-Tan/”(talk) in the sentence (E3) indicates a *meet* event from the intra-sentential information (entities, such as agents, time and place in the sentence) and difficult for mention “会谈/Hui-Tan/”(talk) in the sentence (E4), we can easily infer from the sentence (E3) that “会谈/Hui-Tan/”(talk) in the sentence (E4) also indicates a *meet* event, using the one sense per discourse principle: if one instance of a word in a discourse is a trigger mention of an event type, it is very likely that its other instances in the same discourse will be a trigger mention of the same event type.

The rest of this paper is organized as follows. Section 2 overviews the related work. Section 3 introduces a state-of-the-art baseline system for Chinese event extraction. Section 4 describes the inference mechanism to infer unknown triggers for Chinese trigger identification by employing the compositional semantics inside Chinese triggers. Section 5 describes the inference mechanism to recover trigger mentions by imposing the discourse consistency between Chinese trigger mentions. Section 6 presents the experimental results. Section 7 discusses the impact of our two mechanisms. Section 8 concludes the paper and points out future work.

2. Related work

Almost all the existing studies on trigger identification and overall event extraction concern English. While earlier studies focus on sentence-level extraction (Ahn, 2006; Grishman et al., 2005; Hardy et al., 2006), later ones turn to employ higher-level information, such as document (Finkel et al., 2005; Maslennikov & Chua, 2007; Patwardhan & Riloff, 2009), cross-document (Ji & Grishman, 2008), cross-event (Gupta & Ji, 2009; Liao & Grishman, 2010) and cross-entity (Hong et al., 2011) information.

Compared with the tremendous work on English trigger identification, there are only a few studies on Chinese trigger identification with focus on either feature engineering or trigger expansion, under the same framework as English trigger identification.

On feature engineering, Tan, Zhao, and Zheng (2008) first employ a local feature selection method to ensure the performance of trigger classification and then apply multiple levels of patterns to improve the coverage in argument classification. Fu et al. (2010) apply a feature weighting scheme to re-weight various features for trigger identification and event type classification. Chen and Ji (2009b) apply various kinds of lexical, syntactic and semantic features to address the special issues in Chinese. Wang (2012) introduces various kinds of surface and semantic features, such as term frequency, sentence location, sentence length, title words overlap rate, semantic role features, to select the most informative sentences as event candidates.

On trigger expansion, almost all the related studies use predefined or automatically-clustered synsets, a common mechanism widely used in various NLP applications, to infer new triggers. Chen and Ji (2009a) propose a bootstrapping framework to exploit extra information captured by an English event extraction system. Particularly, they first translate the English trigger and its arguments in a special event mention to Chinese via manual cross-lingual projection and then add those additional event mentions to their Chinese event extraction system via cross-lingual co-training. Ji (2009) first extracts some cross-lingual predicate clusters using bilingual parallel corpora and a cross-lingual information extraction system, and then employs the derived clusters to expand the triggers. Both Zhao, Qin, Che, and Liu (2008) and Qin, Zhao, Ding, Liu, and Zhai (2010) employ a semantic dictionary “TongYiCi Ciling (expanded version)” to expand triggers for Chinese event type classification. The problem with current trigger expansion is that it fails to consider the sense shifting of a word in different contexts and thus may introduce too many pseudo triggers and harm the precision.

Regarding the computational semantics, almost all the related studies focus on how to combine words together to convey complex meanings (Liang, Joedan, & Klein, 2011; Wong & Mooney, 2007; Zettlemoyer & Collins, 2007). However, the compositional semantics mentioned in this paper is more fine-grained with focus on how to combine Chinese characters into a word and mine its semantics from its morphological structure, especially of verb and noun as an event trigger. To our knowledge, there are only two papers systematically addressing the compositional semantics inside Chinese words. Li (2011) explores the internal structures inside Chinese nouns and employs it in word segmentation, while Li, Zhou, Zhu, and Hou (2012) further explore various kinds of Chinese morphological structures in the joint parsing of Chinese morphological and syntactic structures.

As an important hypothesis in natural languages, the discourse consistency has been widely applied to many natural language processing applications, such as named entity recognition (Finkel & Manning, 2009) and coreference resolution (Klenner, 2007). Specially, several studies have successfully incorporated trigger or entity consistency constraints into English event extraction. Yarowsky (1995) and Yangarber & Jokipii (2005), Yangarber et al. (2007) apply a cross-document inference mechanism to refine local extraction results for the disease name, location and start/end time. Mann (2007) proposes some constraints on relationship rescoring to impose the discourse consistency on the CEO's personal information. Ji and Grishman (2008) employ a rule-based approach to propagate consistent triggers and arguments across topic-related documents. Gupta and Ji (2009) use a similar approach to recover implicit time information for events. Liao and Grishman (2011) integrates a similar approach with a self-training strategy to extract events. Besides, Liao and Grishman (2010) employ the cross-event consistency information to improve sentence-level event extraction. Hong et al. (2011) regard the entity type consistency as a key feature to predict event mentions and propose a cross-entity inference mechanism to improve the traditional event extraction system.

In summary, there are no explicit approaches to capture the special characteristics and challenges in Chinese language to improve the performance of Chinese event extraction, in particular Chinese trigger identification.

3. Task definition and baseline system

In this section, we define the task firstly and then introduce a strong baseline system.

3.1. Task definition

To better understand the Chinese event extraction task as defined in ACE evaluations, where an event is defined as a specific occurrence involving participants, we list some ACE terminologies:

Event mention: a phrase or sentence within which an event is described;

Trigger: the main word which most clearly expresses the occurrence of an event, so recognizing an event can be recast as identifying a corresponding trigger;

Trigger mention: a reference to a trigger;
 Event type: the type of an event;
 Argument: an entity involved in an event;
 Argument role: the relation of an argument to an event where it participates.

In particular, the event extraction task can be divided into four components:

Trigger identification: to distinguish true trigger mentions from pseudo trigger mentions;
 Event type classification: to classify trigger mentions by event types;
 Argument identification: to distinguish true arguments from pseudo arguments;
 Argument role classification: to classify arguments by argument roles.

Take the following sentence as an example:

(E5) 沙龙今天到达华盛顿/Sharon arrived Washington today.

The trigger identifier should first recognize “到达/Dao-Da/” (arrive) as a trigger mention and then the event type classifier should determine it as a *Movement* event. Finally, the argument identifier should select “沙龙/Sharon/” (Sharon), “今天/Jin-Tian/” (today) and “华盛顿/Hua-Sheng-Dun/” (Washington) as the arguments of this event mention and the argument role classifier should assign the roles *Artifact*, *Time* and *Destination* to them respectively.

3.2. Baseline system

For comparison, we re-implement the state-of-the-art Chinese event extraction system, as described in [Chen and Ji \(2009b\)](#), which consists of four typical components (i.e. trigger identification, event type classification, argument identification and argument role classification) in a pipeline way, with the same set of features and the same Maximum-Entropy (ME) model to train individual component classifiers. During testing, each word in the test set is first scanned for instances of known triggers from the training set. When an instance is found, the trigger identifier is applied to distinguish those true trigger mentions from pseudo ones. If true, the event type classifier is then applied to recognize its event type. For any entity mention in a sentence which is identified as an event, the argument identifier is employed to assign its possible arguments afterwards. Finally, the argument role classifier is introduced to assign a role to each argument.

One problem with Chen and Ji's system is its ignoring effective long-distance features. In order to resolve this problem and provide a stronger baseline, we introduce more features into four components:

Trigger identification and event type classification: (1) syntactic features: the path to the root of the governing clause, the subject and the object of the trigger mention when they are entities; (2) nearest entity information: the entity type of left syntactically/physically nearest entity to the trigger mention + entity, the entity type of right syntactically/physically nearest entity to the trigger mention + entity;

Argument identification and argument role classification: (1) syntactic features: POS (Part-Of-Speech) of the trigger mention, the dependency path from the entity to the trigger mention; (2) neighbor words: the left neighbor word of the entity + its POS, the right neighbor word of the entity + its POS, the left neighbor word of the trigger mention + its POS, the right neighbor word of the trigger mention + its POS; (3) semantic role features: Arg0 and Arg1 of the trigger mention, returned by a semantic role labeler ([Zhou, Li, Fan, & Zhu, 2011](#)).

Normally, Chinese trigger identification may suffer much from the errors propagated from upstream processing such as part-of-speech tagging and parsing, especially word segmentation. To alleviate word segmentation errors to known triggers, [Chen and Ji \(2009b\)](#) constructs a global errata table to deal with such errors in the training set. In this paper, a split-merge strategy is applied to recover those wrongly segmented known triggers and explore possible unknown triggers. Particularly, for each single-morpheme predicate in a document after word segmentation, it is merged with either previous or next single-morpheme word to form a trigger candidate if the trigger candidate has the same morphological structure of a known trigger.

3.3. Experimental setting

The ACE 2005 Chinese corpus (only the training data is available) is used in all our experiments. The corpus contains 633 Chinese documents annotated with 8 predefined event types and 33 predefined subtypes. In this paper, we adopt the same experimental setting as [Chen and Ji \(2009b\)](#) and randomly select 567 documents as the training set and the remaining 66 documents as the test set. Besides, we reserve 33 documents in the training set as the development set, following the setting of ACE diagnostic tasks, and use the ground truth entities, times and values for our training and testing.

For evaluation, we follow the same standards as defined in [Ji \(2009\)](#):

Table 2

Performance of trigger identification, event type classification, argument identification and argument role classification.

System	Trigger identification			Event type Classification			Argument identification			Argument role classification		
	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
Chen and Ji (2009b)	71.5	51.2	59.7	66.5	47.7	55.6	56.1	38.2	45.4	53.1	36.2	43.1
Our baseline:+f	75.2	52.0	61.5	70.3	49.0	57.8	58.4	42.7	49.3	55.2	38.6	45.4
Our baseline:+f + SM	75.0	53.5	62.4	70.2	50.1	58.5	58.2	43.4	49.7	55.1	39.1	45.7

A trigger is correctly identified if its position in the document matches a reference trigger;

An event type is correctly determined if the event type of a trigger mention and its position in the document match a reference trigger;

An argument is correctly identified if its involved event type and position in the document match any of the reference argument mentions;

An argument role is correctly determined if its involved event type, position in the document, and role match any of the reference argument mentions.

Finally, all sentences in the corpus are divided into words using a word segmenter ICTCLAS³ with all entities annotated in the corpus kept. Besides, we use Stanford Parser (Chang, Tseng, Jurafsky, & Manning, 2009; Levy & Manning, 2003) to create both the constituent and dependency parse trees and employ the ME model⁴ to train individual component classifiers.

3.4. Experimental results

Table 2 shows the Precision (P), Recall (R) and F1-Measure (F) on the held-out test set. It shows that our newly-incorporated features improve the performance by 1.8 units, 2.2 units, 3.9 units and 2.3 units in F1-measure on trigger identification, event type classification, argument identification and argument role classification, respectively, with both gains in precision and recall. It also shows that our split-merge strategy (SM) further improves the performance by 0.9 units, 0.7 units, 0.4 units and 0.3 units in F1-measure in the four components respectively. However, this improvement is very limited compared to possible 8.7% word segmentation errors to known and unknown triggers.

For our baseline system, given the small performance gaps between trigger identification and event type classification (3.9 units in F1-measure: 62.4 units vs. 58.5 units) and between argument identification and argument role classification (4.0 units in F1-measure: 49.7 vs. 45.7), the performance bottlenecks of our baseline system mainly exist in trigger identification and argument identification, particularly the former one. While argument identification has the performance gap of 8.8 units in F1-measure compared to trigger type classification (49.7 units vs. 58.5 units), the former one, trigger identification, can only achieve the performance of 62.4 units in F1-measure (in particular the recall with only 53.5 units). This indicates the importance of trigger identification in overall Chinese event extraction.

In this paper, we will focus on trigger identification to improve its performance, particularly for the recall, via employing the compositional semantics inside Chinese triggers to infer unknown triggers and imposing the discourse consistency between Chinese trigger mentions to recover trigger mentions.

4. Employing compositional semantics inside Chinese triggers to infer unknown triggers

Language is perhaps the only communicative system in nature, which compositionally builds structured meanings from smaller pieces, and this compositionality is the cognitive mechanism that allows for what Humboldt called language's "infinite use of finite means." While the lexical semantics is the smallest in most Chinese language processing applications, this paper introduces a more fine-grained semantics, the compositional semantics inside Chinese triggers, and unveils its effect and usage in Chinese language processing by employing it into Chinese trigger identification.

4.1. Compositional semantics and morphological structures in Chinese words

Both in English and Chinese languages, a word is composed of one or more characters. However, a component character in English is just the basic unit to form a word instead of a semantic unit. In comparison, most Chinese characters have their own meanings as morphemes, the minimal meaningful unit in Chinese language. If a Chinese word contains more than one morpheme, its meaning can be often interpreted in terms of its composite morphemes. This more fine-grained semantics are the compositional semantics inside Chinese words namely. Actually, it is also a normal way to understand a new Chinese word in everyday life for a Chinese native speaker. Table 3 shows four examples of such compositional semantics in Chinese

³ <http://ictclas.org/>

⁴ <http://mallet.cs.umass.edu/>

Table 3
Examples of compositional semantics in Chinese words.

Word	Morphemes
会见/Hui-Jian/ (interview ^a)	会/Hui/ (meet) 见/Jian/ (meet)
击毙/Ji-Bi/ (shoot and kill)	击/Ji/ (shoot) 毙/Bi/ (kill)
来到/Lai-Dao/(come)	来/Lai/ (come) 到/Dao/ (to)
私信/Si-Xin/ (private letter)	私/Si/ (private) 信/Xin/ (letter)

^a “会见” has more than one sense. Here, we just give the one when it acts as an event trigger in the ACE 2005 Chinese corpus.

words. For example, “会见/Hui-Jian/” is composed of two morphemes, “会/Hui/” and “见/Jian/”, which have their own semantics, and the semantics of “会见/Hui-Jian/” comes from those of its component morphemes “会/Hui/” and “见/Jian/”.

Without doubt, a general method to represent the compositional semantics inside Chinese words is to systematically explore the morphological structures in Chinese words since it is the nature of compound words. As the basic word-building mechanism to form morphemes into words, the morphological structures of Chinese words are formulated by three major processes: compounding, affixation, and conversion. Among them, compounding concatenates two or more morphemes together to form a compound word, affixation is a morphological process to add grammatical or lexical information to a base form, and conversion changes a word from one POS into another without the addition or deletion of any morphemes. Besides, compounding is the most productive way to compose a Chinese word while affixation is the most popular way to construct an English word. As for conversion, it's really not a way to construct a word and just represents the fact that some words have more than one grammatical role.

4.2. Morphological structures in Chinese triggers

Since almost all triggers in Chinese events are verbs and nouns, we focus on the morphological structures of Chinese verbs and nouns. Actually, the statistics on the ACE 2005 Chinese corpus shows that more than 95% of the triggers are either verbs or verbal nouns and nearly 5% of the triggers are nouns (e.g., “公开信/Gong-Kai-Xin/” (open letter), “大会/Da-Hui/” (plenary session)). In the ACE 2005 English corpus, there are some adjectives triggering an event of a special type. However, no adjective acts as a trigger in the ACE 2005 Chinese corpus. Besides, almost 95% of the triggers in the ACE 2005 Chinese corpus contain one or two morphemes. Therefore, this paper only considers the one-morpheme and two-morpheme triggers of verbs and nouns. Besides, this paper focuses on following morphological structures in compounding a two-morpheme trigger (besides the one-morpheme structure) due to its dominance:

Coordinative Structure: The two morphemes of a trigger play coordinative roles. For example, “合/He/” (combine) and “并/Bing/” (merge) are coordinative in trigger “合并/He-Bing/” (merge).

Modifier-Head Structure: The modified morpheme follows the modifying one in a trigger. For example, “婚/Hun/” (marry) is modified by “新/Xin/” (new) in trigger “新婚/Xin-Hun/” (newly-married).

Subject-Predicate Structure: One morpheme is the subject and the other one tells something about the subject. This structure is like a subject-predicate sentence condensed in a trigger. For example, “身/Shen/” (body) is a subject of predicate “亡/Wang/” (die) in trigger “身亡/Shen-Wang/” (die).

Predicate-Object Structure: The first morpheme (predicate) governs the second one (object) in a trigger. For example, “业/Ye/” (business) serves as the object of predicate “开/Kai/” (start) in trigger “开业/Kai-Ye/” (start business).

Predicate-Complement Structure: The first morpheme is a predicate and the second one interprets the first one from different aspects (e.g., direction, result and tense) in a trigger. For example, morpheme “入/Ru/” (into) expresses the direction of action “进/Jin/” (go) in trigger “进入/Jin-Ru/” (go into).

4.3. Determining morphological structures in Chinese triggers

A general method to determine the morphological structures in Chinese triggers is to first annotate some instances manually and then train a classifier. Alternatively, a simple way is employed in this paper to determine the morphological structures in Chinese triggers via their POS structures, due to our finding that the morphological structures in Chinese triggers can be largely inferred from their POS structures. Following are the inference rules employed in this paper for different morphological structures:

Single-Morpheme Structure: For a single-morpheme trigger whose POS is noun or verb, its morphological structure is *Single-Morpheme*. The statistics on the training set shows that this inference rule covers 100% of cases given correct POSs.

Predicate-Complement Structure: If the POS structure of a trigger is (*verb + preposition*) or (*verb + auxiliary*), its morphological structure is *Predicate-Complement*. The statistics on the training set shows that this inference rule covers 99.6% of cases given correct POSs.

Table 4
Distribution of different morphological structures in Chinese triggers.

Structure	% Trigger types(mentions)
Single-Morpheme	12.3(19.1)
Coordinative	38.1(46.3)
Modifier-Head	15.8(13.3)
Predicate-Object	14.1(11.4)
Predicate-Complement	16.1(8.7)
Other Words (length ≥ 3)	4.6(1.2)

Predicate-Object Structure: If the POS structure of a trigger is (*verb + noun*), its morphological structure is *Predicate-Object*. The statistics on the training set shows that this inference rule covers 99.2% of cases given correct POSs.

Coordinative Structure: If the POS structure of a trigger is (*verb + verb*) (e.g., “捐/Juan/VV 赠/Zeng/VV” (donate), “购/Gou/VV 买/Mai/VV” (buy), etc.), its morphological structure is *Coordinative*. The statistics on the training set shows that this inference rule covers 97.6% of cases given correct POSs. The only exception to this inference rule is that it ignores those triggers whose POS structure is (*noun + noun*). This happens in Chinese triggers, though seldom. In such cases, i.e. if the POS structure of a trigger is (*noun + noun*), its morphological structure can be either *Modifier-Head* or *Coordinative* (e.g., “婚/Hun/NN 姻/Yin/NN” (marriage)).

Modifier-Head Structure: The morphological structure of a trigger is *Modifier-Head*, if its POS structure is one of following four structures: (1) (*adjective + verb*); (2) (*adjective + noun*); (3) (*noun + noun*); (4) (*noun + verb*). The statistics on the training set shows that this inference rule covers 95.5% of cases given correct POSs. The only exceptions to this inference rule are that if the POS structure of a trigger is (*noun + noun*) or (*noun + verb*), its morphological structure can also be *Coordinative* or *Subject-Predicate*, respectively.

Subject-Predicate Structure: Our exploration on the ACE 2005 Chinese corpus shows that only one trigger (i.e. “身亡/Shen-Wang/” (die)) has the *Subject-Predicate* structure. Therefore, we ignore this structure in this paper.

To obtain the POS structures of Chinese triggers, we split all triggers into characters and employ a Chinese POS tool – ICT-CLAS to tag their POSs. Table 4 shows the distribution of the morphological structures in Chinese triggers in the training set, extracted using above inference rules. Random manual evaluation of 1000 instances shows that our inference rules achieve the accuracy of 91.2% given automatically-tagged POSs.

4.4. Identifying head morphemes in Chinese triggers

Normally, almost all Chinese verbs or nouns contain one morpheme as the governing semantic element, called Head Morpheme (HM) in this paper. Since the semantics of a Chinese trigger can be often inferred from its HM, it's natural to infer unknown triggers via HMs. For example, given verb “死/Si/” (die) as HM in trigger “烧死/Shao-Si/” (burn to death, trigger of the *Die* event) whose morphological structure is *Coordinative*, it is reasonable to infer “砸死/Za-Si/” (crush to death), “炸死/Zha-Si/” (burst to death), “闷死/Men-Si/” (stifle to death) to be triggers of the *Die* event, due to their same morphological structures and HMs as “烧死/Shao-Si/”. This also applies to noun as HM. For example, given “私信/Si-Xin/” (private letter) a trigger of the *Phone-Write* event and “信/Xin/” (letter) the HM of this trigger, it is reasonable to infer those words, with HM “信/Xin/” (letter) and morphological structure *Modifier-Head* (e.g., “贺信/He-Xin/” (congratulatory letter), “密信/Mi-Xin/” (secret letter), etc.), as new triggers.

Therefore, how to identify the HM in a Chinese trigger becomes the key to infer unknown triggers. Table 5 shows our automatic mechanism to identify HM, where $LM(w)$ and $RM(w)$ are used to obtain the left and right morphemes from one-morpheme or two-morphemes word w respectively.

For a trigger whose morphological structure is *Single-morpheme*, *Predicate-Complement* or *Modifier-Head*, it's easy to identify its HM from the relationship between its two morphemes. If the structure of a trigger is *Predicate-Object*, we select the noun (object) as HM because it better represents the semantics of the trigger than the predicate, i.e. the governing semantic element always comes from the object. However, without additional information, it's hard to select HM from a trigger whose morphological structure is *Coordinative*. For example, trigger “访问/Fang-Wen/” (visit) has the morphological structure *Coordinative*, whose two component morphemes, “访/Fang/” (visit) and “问/Wen/” (ask), have their own semantics. Fortunately, we can find out that morpheme “访/Fang/” (visit) has the same meaning as trigger “访问/Fang-Wen/” (visit). So an effective way to identify HM in a trigger of the *Coordinative* structure is via the semantic similarity.

In this paper, we employ HowNet⁵ (Dong and Dong, 2006) to obtain the semantics of Chinese words. Similar to Wordnet in English, HowNet is a structured Chinese lexical semantic resource. In HowNet, sememe is a basic semantic unit and represents the meaning of a word. In total, about 2200 sememes are used to define 95000 Chinese words. In this paper, the governing sememe is introduced to recognize HMs from those triggers with the *Coordinative* structure. That is, if a morpheme represents the

⁵ <http://www.keenage.com>

Table 5
Inference rules on different morphological structures to extract HMs.

Morphological structure	Inference rule to extract HM
Single-Morpheme	<i>trigger</i>
Coordinative	$LM(trigger)$: if $SemSim(trigger, LM(trigger)) > \alpha$ $RM(trigger)$: if $SemSim(trigger, RM(trigger)) > \alpha$
Modifier-Head	$RM(trigger)$
Predicate-Object	$RM(trigger)$
Predicate-Complement	$LM(trigger)$

Table 6
Distribution of HMs over different triggers.

Number	% Distribution over Trigger types (mentions)
1	14.7(10.9)
2–4	29.1(23.9)
5–9	28.1(32.9)
≥ 10	28.1(32.3)

governing sememe, it is recognized as HM of that trigger. Following Liu and Li (2002), function $SemSim(x, y)$ is used to calculate the semantic similarity between the sememes of the trigger x and its morpheme y as follow:

$$SemSim(x, y) = \frac{\phi}{Dis(x, y) + \phi} \quad (1)$$

where $Dis(x, y)$ is the distance between the sememe of x and y in HowNet's sememe hierarchical architecture and ϕ is an adjustable parameter. In preliminary study, we have tried to experiment on the development set to find an optimal parameter for ϕ . However, our experimental results show this parameter is not sensitive enough: when parameter ϕ changes from 0.5 to 2.0, the performance of identifying head morphemes is relatively stable. Therefore, we simply set it to 0.75, following Liu and Li (2002).

If the sememe of a trigger is similar to one of its morphemes, we extract this morpheme as HM. For those words that have more than one sememe, we choose the most similar sememe. Table 6 shows the distribution of HMs over different triggers. It shows that 85.3% of HMs appear in more than one trigger type and 56.2% of them appear in more than 4 trigger types. As for trigger mentions, the percentages become 89.1% and 65.2% respectively. An extreme example is that 85.2% (75/88) of the trigger types in the *Trial-Hearing* event mentions contain HM “审/Shen/” (trial) and 85.4% (117/138) of the trigger types in the *injure* event mentions contains HM “伤/Shang/” (injure).

4.5. Inferring unknown triggers

To better represent the nature of the compositional semantics inside Chinese words, we infer unknown triggers both on the head morphemes and governing sememes, according to their morphological structures.

Firstly, following the compositional semantics principle, all the one-morpheme or two-morpheme words in the test set are extracted as candidates if candidate word w satisfies following conditions:

- (1) w contains at least one HM;
- (2) The POS of w is noun or verb;
- (3) w occurs as a pseudo trigger less than 5 times in the training set;
- (4) The conditional probability that the similar known triggers of word w refer to true trigger mentions is greater than 0. Here, similar triggers are those with the same morphological structure and the same HM.

It is very likely that a candidate word w will not be a trigger if it appears in the training set several times but never triggers an event. The statistics on the training set shows that this filter can eliminate 28% of pseudo triggers and 3% of true triggers.

Secondly, for each candidate word w in the candidate set, we apply following inferences to distinguish the true unknown triggers from the pseudo ones on different morphological structures, according to its morphological structure.

Single-Morpheme: We expand those HMs in two-morpheme triggers as the single-morpheme triggers. So we apply a simple inference to infer unknown triggers: if the maximum score of the semantic similarity $SemSim$ between word w and each known trigger which contain word w is equal to 1, we accept w as an unknown trigger.

Predicate-Complement: The first morpheme is usually a verb, so the sememe of word w always is similar to the sememe of its first morpheme. We accept word w as an unknown trigger when its left morpheme is in the set of left morphemes of those *Predicate-Complement* triggers in the training set.

Predicate-Object: we regard word w as an unknown trigger following two conditions:

- (1) Its right morpheme is a HM;
- (2) The maximum score of the semantic similarity *SemSim* between the right morpheme of word w and its similar triggers in the training set is larger than or equal to a threshold β .

Similar triggers are those triggers with the same morphological structure and the same HM. For example, if there are two triggers “离职/Li-Zhi/” (resign) and “辞职/Ci-zhi/” (resign), and their HMs are “职/Zhi/” (job) too. For a candidate “免职/Main-zhi/” (resign), its morphological structure is as same as the above two and its HM also is “职” (job). We call them similar triggers and calculate the similarities between “免/Mian/”(dismiss) and the predicates (“离/Li/” (leave), “辞/Chi/” (dismiss)) in its similar triggers in the training set.

Modifier-Head: The first morpheme of word w modifies the second one, so that the semantics of word w comes from its second morpheme. We apply following rules based on POS consistency and semantic similarity:

- (1) Its right morpheme is a HM;
- (2) The maximum score of the semantic similarity *SemSim* between the right morpheme of word w and each known trigger which contain the right morpheme of word w is equal to 1;
- (3) The POS of the left morpheme of word w belongs to the POS set of left morphemes of all known triggers whose structures are *Modifier-Head*.

Coordinative: Since the two composite morphemes of word w are homogeneous and its semantics is flexible and maybe comes from the combination of its two morphemes or one of its morpheme. We calculate the average score of the similarities to infer trigger of this type as follows:

$$\frac{\sum_{co \in \{w, LM(w), RM(w)\}} \text{MAX}_{m_i \in SC}(\text{SemSim}(co, m_i))}{3} > \lambda \quad (2)$$

where SC is the set of known triggers with following two constraints: (1) their morphological structures are *Coordinative*; (2) their left/right morphemes and the left/right morpheme of word w are the same HM. Besides, λ is a threshold to select unknown triggers.

5. Imposing discourse consistency between Chinese trigger mentions to recover trigger mentions

Some trigger mentions fail to provide enough intra-sentential information to the trigger identifier and their triggering event mentions can be only inferred confidently from the discourse-level information, especially when the word triggering an event is polysemous or an unknown trigger. In this paper, we impose the discourse consistency between Chinese trigger mentions to recover trigger mentions for Chinese trigger identification.

5.1. Discourse consistency between Chinese trigger mentions

Take the following discourse as an example:

-
- (E6) 8日晚犹太人和阿拉伯人在拿萨勒市发生冲 (TM1)/8-Ri-Wan-You-Tai-Ren-He-A-La-Bo-Ren-Zai-Na-Sa-Lei-Shi-Fa-Shen-Chong-Tu/, 在冲 (TM2)中/Zai-Chong-Tu-Zhong/,当地一名以色列籍阿拉伯人丧生/Dang-Di-Yi-Ming-Yi-Se-Lie-Ji- A-La-Bo-Ren-Sang-Sheng/ 另外/Ling-Wai/,该冲 (TM3)还导致数十人受伤/Gai-Chong-Tu-Hai-Dao-Zhi-Shuo-Shi-Ren-Shou-Shang/
- (**Clashes (TM1)** erupted between Jews and Arabs in Nazareth at the night of the 8th, , A local Israeli Arabian died during the **clashes (TM2)**. , Otherwise, the **clashes (TM3)** caused the injuries of dozens of people.)
-

Obviously, it is easy for a trigger identifier to determine that mention TM1 of trigger “冲 /Chong-Tu/” (clash) indicates an *Attack* event from the contained intra-sentential information (e.g., attacker, target, time and place) and difficult to determine that mention TM2 and TM3 are an *Attack* event from the contained intra-sentential information. However, the latter two trigger mentions can be easily inferred from the discourse using the one sense per discourse principle, i.e., if one mention of a word in a discourse triggers an event of one type, it is very likely that its other mentions will trigger events of the same type.

Table 7 compares the discourse consistency between trigger mentions in the ACE 2005 Chinese and English corpora. Here, a trigger is considered full-consistent when each of its appearances is annotated to trigger an event of the same type or not.

Table 7
Comparison of discourse consistency between trigger mentions in ACE 2005 Chinese and English corpora.

Language	Full-consistent (%)	Adjacent pair-consistent (%)
English	70.2	87.5
Chinese	90.5	95.4

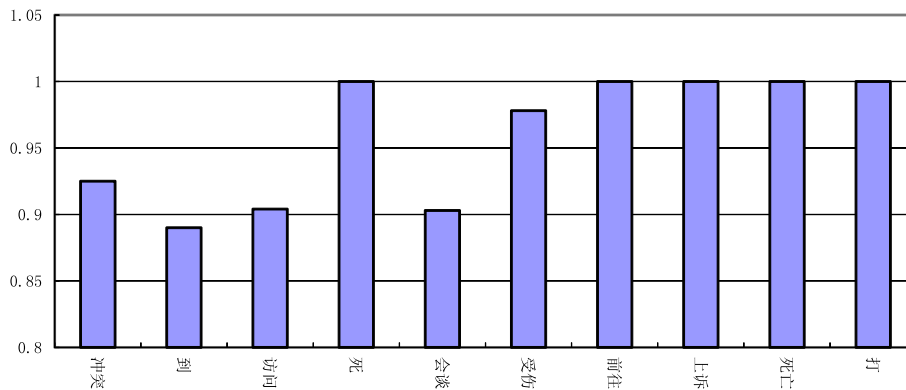


Fig. 1. Discourse consistency for top 10 frequent triggers in the ACE 2005 Chinese corpus.

For comparison, Table 7 also includes the adjacent pair consistency for adjacent trigger mention pairs. For example, mentions TM1, TM2 and TM3 of trigger “冲 /Chong-Tu/” in (E6) are annotated full-consistent while both TM1/TM2 and TM2/TM3 are adjacent pair-consistent. It shows that the discourse consistency between Chinese trigger mentions holds much more likely than the English counterpart. Fig. 1 gives the discourse consistency for top 10 frequent triggers, which occupy 18% of the event mentions in the ACE 2005 Chinese corpus.

Inspired by the observation on the adjacent pair consistency between the mentions of the same trigger, we find out that the adjacent pair consistency also exists between those trigger mentions with the same HM. That is, if one mention of a word triggers an event of one type, it is likely that the adjacent mention of a word in the same discourse with the same HM will also trigger an event of the same type. Take the following discourse as an example:

- (E7) 武绍祖26岁的时候就担任(TM4)全国学联主席/Wu-Sa0-zu-26-Sui-De-Shi-Hou-Jiu-Dan-Ren-Quan-Guo-Xue-Lian-Zhu-Xi/,后来又出任(TM5)国务院副总理王震的秘书/Hou-Lai-You-Chu-Ren-Guo-Wu-Yuan-Fu-Zong-Li-Wang-Zhen-De-Mi-Shu/,1985年更是升任(TM6)国防科工委政委/1985-Nian-Gen-Shi-Sheng-Ren-Guo-Fang-Ke-Gong-Wei-Zheng-Wei/
(Shaozu Wu **served (TM4)** as the president of All-China Students' Federation when he was 26 and then he **took up the post of (TM5)** the secretary of Zhen Wang, the Chinese vice premier. In 1985, he was **promoted (TM6)** as the commissar of National Defense Division.)

In (E7), trigger mentions TM4, TM5 and TM6 have the same HM “任/Ren/” (serve). This is largely due to people's tendency to avoid literal repetition. The statistics on the ACE 2005 Chinese corpus shows that the adjacent pair consistency between trigger mentions with the same HM holds in more than 90% cases.

5.2. Recovering trigger mentions via discourse consistency between Chinese trigger mentions

In a discourse, all mentions of a trigger can be linked into an undirected graph. This can also be applied to all mentions of those triggers with the same HM. Due to the higher probability of the adjacent pair consistency than the full consistency and the more difficulty of modeling the full consistency than the adjacent pair consistency, this paper adopts the adjacent pair consistency to impose the discourse consistency between Chinese trigger mentions. Particularly, we apply the Markov Network (Markov Random Field) (Kindermann & Snell, 1980), an undirected graphical model, to represent the relationship and impose the discourse consistency between Chinese trigger mentions and introduce the First Order Logic to infer trigger mentions.

Following the discourse consistency principle, our method infers a trigger mention based on other mentions of the same trigger in a discourse. So it's necessary to recognize some trigger mentions in the test set beforehand and then impose the discourse consistency to recover missed trigger mentions. For this purpose, we collect relevant information from the trigger

Table 8
Main evidence and query predicates used in our MLN.

Predicate	Description
$Token(i,d,w)$	Token i in discourse d has word w ;
$Event(i,d)$	Token i in discourse d is a trigger mention and refers to an event;
$EventType(i,d,p)$	Token i in discourse d refers to an event with type p ;
$HeadMorph(i,d,m)$	HM of token i in discourse d is morpheme m ;
$Next(i,j)$	Token j is next to i while token i and j are two mentions of a same word w ;
$Prev(i,j)$	Token j is previous to i while token i and j are two mentions of a same word w ;
$NextMorph(i,j)$	Token j is next to i while token i and j are two mentions with the same HM m ;
$PrevMorph(i,j)$	Token j is previous to i while token i and j are two mentions with the same HM m ;

identifier and event type classifier, two components in our event extraction system. As probability-based ME models, they can provide a confidence that indicates how likely a trigger mention is a true one and how likely a trigger mention belongs to a special event type. Given a discourse and different mentions of a trigger returned by the trigger identifier and the event type classifier, we can simply accept those mentions with high probability as true mentions of the trigger or as a special event type, and regard those with low probability as pseudo trigger mentions. Specifically, for those mentions in-between, we apply Markov Logic Network (MLN) (Richardson & Domingos, 2006) to recover true trigger mentions.

As a statistical relational learning language, Markov Logic Network is based on the First Order Logic and Markov Network. Basically, it can be seen as an expressive template language that uses First Order Logic formulas to instantiate Markov Networks of repetitive structure (Riedel, Chun, Takagi, & Tsujii, 2009). Recently, MLN has been used successfully in many NLP applications (e.g., Poon and Domingos, 2007; Poon and Vanderwende, 2010; Song, Jiang, Zhao, Li, & Wang, 2012).

One issue in MLN is the introduction of logical predicates according to the nature of a task. Table 8 shows the main evidence and query predicates used in our MLN, along with brief descriptions.

Another issue is the specification of weighted first order formulas, which define a distribution over above ground logical predicates following the discourse consistency as described in Section 5.1.

The first formula is adopted to recover trigger mentions with the consistency between trigger identification and event type classification. That is, a trigger mention classified to an event with a special type by the event type classifier must be a trigger mention in the trigger identifier.

$$EventType(i, d, +p) \Rightarrow Event(i, d) \quad (3)$$

Here, the notation “+” signifies that the MLN contains an instance of the formula, with a separate weight, which is learnt from the training set.

The second and third formulas are adopted, following the discourse consistency between Chinese trigger mentions, to recover a trigger mention from its previous or next mention of the same trigger based on the adjacent pair consistency in a discourse.

$$Prev(i, j) \wedge Event(j, d) \Rightarrow Event(i, d) \quad (4)$$

$$Next(i, j) \wedge Event(j, d) \Rightarrow Event(i, d) \quad (5)$$

The fourth and fifth formulas are adopted to recover trigger mentions with the same HM. That is, if one mention of a word triggers an event, the previous or the next mention of a word with the same HM will also trigger an event of the same type.

$$NextMorph(i, j) \wedge Event(j, d) \wedge (\exists m HeadMorph(i, d, m) \wedge HeadMorph(j, d, m)) \Rightarrow Event(i, d) \quad (6)$$

$$PrevMorph(i, j) \wedge Event(j, d) \wedge (\exists m HeadMorph(i, d, m) \wedge HeadMorph(j, d, m)) \Rightarrow Event(i, d) \quad (7)$$

In particular, the open-source Alchemy package⁶ is employed for learning and inference. Like Poon and Vanderwende (2010), we use the Stochastic Gradient Descent (SGD) to learn weights and introduce MC-SAT, a slice sampling Markov chain Monte Carlo algorithm, to make the inference. To obtain a final assignment, we set the query atoms with probability no less than 0.35 (fine-tuned to maximize F1 on the development set) to true and the rest to false, in order to keep true trigger mentions.

6. Experimentation

In this section, we first evaluate our computational semantics and discourse consistency mechanisms in inferring unknown triggers and recovering trigger mentions, respectively, and then evaluate their contribution on Chinese trigger identification and overall Chinese event extraction, using the same experimental settings as described in Section 3.3.

⁶ <http://alchemy.cs.washington.edu/>

Table 9Performance of HM identification (with α fine-tuned to 0.85 using the development set).

System	P(%)	R(%)	F1
HM identification	80.1	82.1	81.4

Table 10Performance of compositional semantics in inferring unknown triggers (with β and λ fine-tuned to 0.7 using the development set).

#Known triggers	#Infered words	#True unknown triggers	#Pseudo unknown triggers
171	62	43(69.4%)	19(30.6%)

Table 11

Impact of compositional semantics in inferring unknown trigger mentions (incremental) (#Gold: 367).

System	#True trigger mentions	#Pseudo trigger mentions
Baseline	266	629
Candidate selection ^a	343	1296
+Single-Morpheme	282	1141
+Predicate-Complement	290	1058
+Modifier-Head	298	862
+Predicate-Object	314	764
+Coordinative (overall)	326	508

^a We extract all candidate trigger mentions following four conditions in Subsection 4.5.

6.1. Computational semantics: inferring unknown triggers

As the key to infer unknown triggers, Table 9 shows the performance of HM identification. For evaluation, the HMs of all the known triggers in the ACE 2005 Chinese corpus are manually labeled by three annotators and we accept those morphemes as HMs when at least two annotators agree on them.

Manual inspection shows that 57% of errors are related to two morphological structures, *Predicate-Object* and *Coordinative*, due to the limited coverage of simple inference rules, besides the upstream errors in the morphological structure identification (34%), largely due to wrong POS tagging.

For a trigger with the *Predicate-Object* structure, the simple inference rule, which extracts the object of such structure as HM, much harms the precision by 7.5 units, due to the fact that the semantics of a word with the *Predicate-Object* structure may be a combination of its component morphemes instead of just the head morpheme. For example, while “*职/Zhi/*” (post) is a useful HM to infer triggers “*离职/Li-Zhi/*” (resign), “*任职/Ren-Zhi/*” (hold a post), “*辞职/Ci-Zhi/*” (resign), etc., the predicates also play an important role in composing their semantics (e.g., “*离/Li/*” (leave)+“*职/Zhi/*” (post): leave + post = resign). For a trigger with the *Coordinative* structure, it is possible that none of its component morphemes is HM.

Table 10 shows the performance of inferring unknown triggers via the compositional semantics. Manual inspection shows that 62 words are inferred as unknown triggers, among which 69.4% are true triggers.

Table 11 shows the contribution of inference rules on different morphological structures to infer unknown trigger mentions in the test set. Here, the baseline just extracts those trigger mentions occurring in the training data, as adopted in Chen and Ji (2009b) and our baseline event extraction system

Compared with the baseline, our mechanism recovers 16.3% (60) of true trigger mentions. This improvement further proves that our mechanism is effective to infer unknown trigger mentions. Moreover, combining HMs and sememes of Chinese words not only can infer unknown trigger mentions, but also can filter out more pseudo trigger mentions. Table 11 shows 60.8% (788) of pseudo trigger mentions are filtered out and that will help the trigger identifier to improve the precision. Besides, some triggers in the training set are seldom used as trigger mentions and the Non-trigger Filtering also is applied to filter out the mentions of those known triggers in the step of candidate selection. Almost 28% of pseudo trigger mentions of known triggers are filtered out (The percent of the filtered true trigger mentions is less than 3%), so the overall number of pseudo trigger mentions is reduced to only 508, less than that of the Baseline.

6.2. Discourse consistency: recovering trigger mentions

We collect relevant information from the trigger identifier and event type classifier and simply accept those mentions with high probability as true mentions of the trigger or as a special event type, and regard those with low probability as pseudo trigger mentions. The high and low probability thresholds are fine-tuned to 95% and 5% respectively, using the

Table 12
Impact of discourse consistency between trigger mentions in recovering trigger mentions.

Inference	Trigger identification		
	P(%)	R(%)	F1
Consistency between trigger identification & event type classification	+2.3	−0.3	+0.9
Consistency between mentions of the same triggers	+2.1	+1.1	+1.6
Consistency between mentions of the same HMs	+1.2	+0.8	+1.0
Overall	+5.6	+1.6	+3.5

Table 13
Overall contribution to Chinese trigger identification and overall event extraction.

System	Trigger identification			Event type classification			Argument identification			Argument role classification		
	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1	P(%)	R(%)	F1
Our baseline	75.0	53.5	62.4	70.2	50.1	58.5	58.2	43.4	49.7	55.1	39.1	45.7
+Compositional semantics	73.3	69.8	71.5	69.9	66.5	68.2	57.6	55.0	56.3	54.1	50.1	52.0
++Discourse consistency	78.9	71.4	75.0	74.0	67.6	70.7	59.9	55.7	57.7	56.4	50.6	53.3
Our baseline + synsets	72.1	56.8	63.5	67.5	52.9	59.3	56.8	45.2	50.3	53.6	40.8	46.3
Human annotator1(blind)	63.3	62.9	63.1	61.7	59.5	60.6	64.6	54.1	58.9	60.9	48.2	53.8
Human annotator2(familiar)	72.6	74.3	73.4	69.1	70.2	69.6	71.5	65.9	68.6	66.4	54.6	59.9
Inter-Annotator Agreement	45.8	42.9	44.3	45.3	42.5	43.8	60.4	49.7	54.5	55.1	45.9	50.1

development set. Specifically, for those mentions in-between, we apply our MLN-based mechanism to recover true trigger mentions.

Table 12 shows that our mechanism of imposing discourse consistency further improves the performance by 3.5 units in F1-measure, due to the gain in both precision and recall by 5.6 and 1.6 units respectively. This justifies the effectiveness of our discourse consistency mechanism in recovering trigger mentions. The bigger gain of 5.6 units in precision also indicates that our mechanism is more effective to impose the discourse consistency between those pseudo trigger mentions.

6.3. Chinese trigger identification and overall event extraction

Table 13 illustrates the contribution of our compositional semantics and discourse consistency mechanisms on trigger identification and overall event extraction on the held-out test set. In addition, we also report the performance of two human annotators (The human annotator 1 is a first year postgraduate student with no background to Chinese event extraction while the human annotator 2 is a third year postgraduate student working on Chinese event extraction) on 33 texts (a subset of the held-out test set). It shows that our approach can improve the F1-measure for trigger identification by 12.6 units, trigger type classification by 12.2 units, argument identification by 8.0 units and argument role classification (i.e. overall event extraction) by 7.6 units, largely due to the dramatic increase in recall of 17.9 units, 17.5 units, 12.3 units and 11.5 units, respectively.

Table 13 also shows that that our compositional semantics mechanism significantly enhances the F1-measure of trigger identification by 9.1 units, largely due to a dramatic increase of 16.3 units in the recall. This justifies the effectiveness of our compositional semantics mechanism in inferring unknown triggers. Further analysis also shows that there are two sources for errors. One is that almost 4.7% of the trigger mentions in the test set doesn't have a morpheme appeared in HMs of the training set. For example, there are many ways to express an *injure* event and only a few triggers or its HMs occurring in the training set. The another is due to wrong POS tagging, resulting in wrongly determining the morphological structure of a trigger.

For fair comparison with the traditional method of expanding triggers using predefined or automatically-clustered synsets, we re-implement a Chinese trigger expanding system following Qin et al. (2010), which expanded new triggers using a predefined synset *TongYiCi CiLin*. Table 13 illustrates the impact of this traditional trigger expanding system on event extraction. It shows that the synset-based expanding method only slightly improves the performance over the baseline and this result is consistent with that described in Qin et al. (2010). This may be due to its failure of considering the sense shifting of a word in complex contexts in that too many pseudo triggers may harm the precision.

The low inter-annotator agreement of two annotators indicates that Chinese event extraction is really challenging even for a well-educated human being. However, it is not surprising: the inter-annotator agreement on the English ACE 2005 corpus is only about 40% for trigger identification (Ji & Grishman, 2008). A detailed analysis shows that a human annotator tends to make more mistakes in trigger identification for two reasons. The first one is that a human annotator always misses some event mentions when a sentence contains more than one event mention. The second reason is that it is hard to identify an event mention due to the failure of following specified annotation guidelines, as mentioned in Ji and Grishman (2008). Table 13 also shows the performance gaps of human annotators between trigger identification and event type classification are very small (2.5% and 3.8% in F1-measure). It verifies the importance of trigger identification in Chinese event extraction even

Table 14
Some of fine-grained morphological structures for *Predicate-Object*.

Fine-grained morphological structure	Sample
Event/Action + Source (place/site)	劫狱/Jie-Yu/(break into a jail and rescue a prisoner)
Event/Action + Origin (place/site)	出境/Chu-Jing/(leave the country)
Event/Action + Destination (place/position)	入狱/Ru-Yu/(send to jail)
Event/Status/Action + Things on earth	通车/Tong-Che/(open to traffic)
Event/Entertainment/Exercise + Tool (musical instrument/sporting equipment)	打球/Da-Qiu/ (play the ball)
Event/Status/Action + Agent (astronomical phenomena)	下雪/Xia-Xue/ (snow)

for a human being. For human annotators, it is much easier to determine the event type of a trigger, identify its arguments and determine the role of each argument, all with more than 90% in accuracy, once a trigger is identified correctly.

7. Discussion

In this section, we discuss our two novel mechanisms for inferring triggers and trigger mentions.

7.1. Compositional semantics in inferring unknown triggers

This paper shows that the morphological structures can well represent the compositional semantics in Chinese words and in particular, provide an ideal way to expand the coverage of the triggers in Chinese event extraction. Manual inspection shows that more than half of the errors are related to POS tagging. If we can correct all POS tagging errors, the accuracy of determining the morphological structures in Chinese triggers can reach 96% (+5%) and the F1-measure of HM identification can reach 88%. As a consequence, an additional 3.5% of the true unknown triggers will be inferred and 26% of pseudo unknown triggers will be filtered out.

Compared with English, the morphological structures in Chinese are much more complex and diverse, causing a lot of troubles in Chinese language processing. We ensure the importance of the compositional semantics in Chinese words in a wide range of Chinese language processing applications, such as machine translation, semantic parsing, etc. However, how to employ the compositional semantics in Chinese words to Chinese language processing applications still is a challenge. Table 13 shows its contribution in Chinese event extraction. The reason that we can successfully employ the compositional semantics to infer unknown triggers in Chinese is our novel mechanism of adopting the morphological structures as a bridge.

Although simple, our mechanism is promising enough for further efforts in this direction. If more fine-grained morphological structures can be introduced, more unknown triggers will be inferred. For example, the *Predicate-Object* structure can be divided into a few fine-grained morphological structures, some of which are shown in Table 14. If “入狱/Ru-Yu” (send to jail) is a trigger of the *Arrest-Jail* event, we will infer “劫狱/Jie-Yu” (break into a jail and rescue a prisoner) as a trigger of the same event according to our current mechanism due to the same HM “狱/Yu” (jail) and morphological structure *Predicate-Object*. However, this inference is problematic since “劫狱/Jie-Yu” (break into a jail and rescue a prisoner) is a pseudo trigger. If we can introduce those more fine-grained morphological structures as shown in Table 14 into our mechanism, it will be filtered as pseudo trigger due to its different fine-grained morphological structure from that of “入狱/Ru-Yu” (send to jail).

Our experimentation shows that HM provides a simple but effective way to infer unknown trigger. In this paper, we only consider those words, which contain one or two morphemes, as unknown triggers, ignoring the words with more than two morphemes. Actually, our mechanism can be also applied to these longer words. This is because in Chinese, almost all these longer words are constructed (nestedly) by some one-morpheme or two-morpheme words. For example, as a trigger of the *phone-write* event, “公开信/Gong-Kai-Xin” (open letter) contains a one-morpheme word “信/Xin” (letter) and a two-morpheme word “公开/Gong-Kai” (open), with morphological structure *Modifier-Head* and HM “信/Xin” (letter). So we can infer it from known trigger “私信/Si-Xin” (private letter) whose morphological structure and HM are as same as “公开信/Gong-Kai-Xin” (open letter).

7.2. Discourse consistency

As a discourse-driven language, ellipsis is very common in Chinese, causing the discourse inference a fundamental mechanism to understand the meaning of a clause, sentence or discourse. A document always contains many events. Unfortunately, how to identify the scope of a discourse is challenging. In this paper, we adopt document as a discourse unit to simplify this task. Manual inspection on those inconsistent trigger mentions shows that 65.6% of them are in actually two obvious-different discourses. If the scope of a discourse can be identified exactly, more improvement can be expected.

Table 15 categorizes the samples of the inconsistency between trigger mentions. It shows,

Table 15
Inconsistency between trigger mentions in a discourse.

Category	% Trigger mentions
Specific/Generic	35.0
Annotation inconsistency	34.2
POS error	16.0
Sense shifting	13.6
Others	1.2

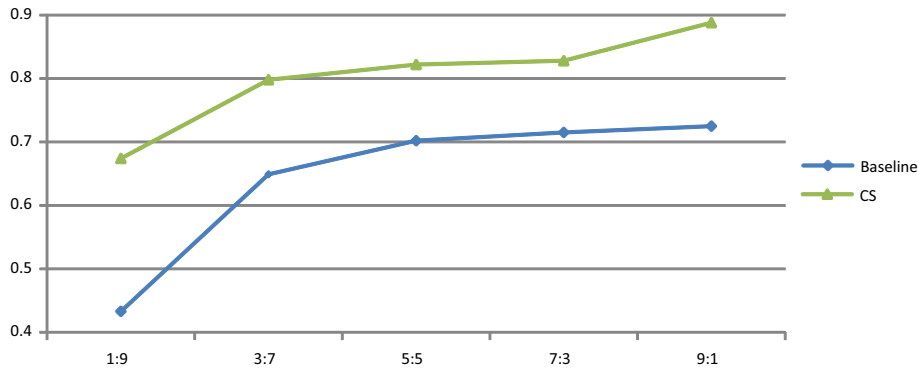


Fig. 2. The percentages of extracted true trigger mentions on different proportions of the training set to the test set.

- (1) 35.0% result from the generic event mentions (e.g., “人总是会死的/Ren-Zong-Shi-Hui-Si-De/ ” (All person always will die.)), which are ignored accordingly to the ACE guidelines;
- (2) 34.2% result from the annotation inconsistency due to the ACE guideline that ambiguous triggers should only be tagged when they are clearly co-referent with an unambiguous trigger within the same document.⁷ However, this guideline is too difficult for the annotators to conform;
- (3) 16.0% are caused by POS tagging errors. For example, “到/Dao/” (come, to) can act as a preposition, an auxiliary or a verb, among which only verb “到/Dao/” (come) can act as a trigger. If a preposition “到/Dao/” (to) is tagged as a verb by mistake, it will be selected as the candidate of the trigger mention and then leads to the inconsistency between this mention and the other mentions in the same discourse as verb and triggers an event of a specific event type;
- (4) Sense shifting covers 13.6% of the inconsistency. Most of them occur in two discourses in a document.

Manual inspection on those consistent trigger mentions shows that 62.5% belong to the discourse consistency between the pseudo trigger mentions and this figure reduces to 37.5% on the consistency between the true trigger mentions. This comparison further justifies our experimental results in Section 6.2 that more pseudo trigger mentions are filtered out and less true trigger mentions are recovered.

7.3. Training set size on the performance of inferring unknown triggers via compositional semantics

In order to evaluate the effect of the training set size on the performance, we modify the proportion of the training set to the test set from 9:1 to 1:9. Fig. 2 shows the percentages of true trigger mentions extracted by our baseline and our compositional semantics mechanism (CS). From Fig. 2, we can find out that our mechanism can extract much more true trigger mentions than that of the baseline, especially for a smaller training set. When the proportion of the training set to the test set is set to 1:9, our mechanism can extract 67.5% of true trigger mentions while the figure drops to 43.3% in our baseline. This justifies that our mechanism can be well applied to minimally-supervised event extraction.

8. Conclusion

In this paper we propose two novel inference mechanisms to Chinese trigger identification. In particular, the compositional semantics inside Chinese triggers are employed to infer unknown triggers and the discourse consistency between Chinese trigger mentions are imposed to recover trigger mentions. We give good reasons why this should be done, and present

⁷ http://www ldc.upenn.edu/Projects/ACE/docs/Chinese-Events-Guidelines_v5.5.1.pdf.

effective ways how this could be done. The evaluation shows that our proposed inference mechanisms for Chinese event extraction are linguistically justified and pragmatically beneficial to real applications.

Compared with our previous work (Li et al., 2012), the contributions of this paper are as follows:

In place of verb structures in Li et al. (2012), more general morphological structures are introduced to better represent the compositional semantics inside Chinese words (esp Chinese triggers). Besides, a novel mechanism is proposed to automatically identify the head morpheme (either verb or noun⁸) as the governing sememe of a trigger according to its morphological structure. Finally, a novel inference mechanism is proposed to infer unknown triggers both on the head morphemes and governing sememes in place of three simple inference rules in Li et al. (2012);

A Markov Logic Network (MLN)-based inference mechanism, is introduced to impose discourse consistency in place of the classifier-based approach in Li et al. (2012). This novel mechanism infers new trigger mentions by integrating various kinds of knowledge in a probabilistic way. As a result, it can achieve better performance on recovering true trigger mentions. Moreover, besides the consistency on the trigger mentions with the same trigger as Li et al. (2012), this paper also imposes the consistency on the trigger mentions with the same head morpheme.

In future work, we will focus on how to introduce the discourse information into the individual components of Chinese event extraction and their joint learning.

Acknowledgments

This research was supported by the National Natural Science Foundation of China under Grant Nos. 61070123, 61331011 and 61273320, the National 863 Project of China under Grant No. 2012AA011102.

References

- Ahn, D., 2006. The stages of event extraction. In *Proceedings of COLING/ACL 2006 workshop on annotating and reasoning about time and events* (pp. 1–8). Sydney, Australia.
- Chang, P.-C., Tseng, H., Jurafsky, D., & Manning, C., 2009. Discriminative reordering with Chinese grammatical relations features. In *Proceedings of third workshop on syntax and structure in statistical translation* (pp. 51–59). Boulder, Colorado.
- Chen, Z., & Ji, H., 2009a. Can one language bootstrap the other: a case study on event extraction. In *Proceedings of NAACL HLT workshop on semi-supervised learning for natural language processing* (pp. 66–74). Boulder, Colorado.
- Chen, Z., & Ji, H., 2009b. Language Specific Issue and Feature Exploration in Chinese Event Extraction. In *Proceedings of 2009 annual conference of the North American Chapter of the ACL (NAACL 2009)* (pp. 209–212). Boulder, Colorado.
- Dong, Zhengdong, & Dong, Qiang (2006). *HowNet and the computation of meaning*. Singapore: World Scientific Pub Co. Inc.
- Finkel, J., & Manning, C., 2009. Joint parsing and named entity recognition. In *Proceedings of 2009 annual conference of the North American chapter of the ACL (NAACL 2009)* (pp. 326–334). Boulder, Colorado.
- Finkel, J., Grenager, T., & Manning, C., 2005. Incorporating non-local information into information extraction systems by Gibbs sampling. In *Proceedings of the 43rd annual meeting of the association for computational linguistics (ACL 2005)* (pp. 363–370). Ann, Arbor, Michigan.
- Fu, Jianfeng, Liu, Zongtian, Zhong, Zhaoman, & Shan, Jianfang (2010). Chinese event extraction based on feature weighting. *Information Technology Journal*, 9, 184–187.
- Grishman, R., Westbrook, D., & Meyers, A., 2005. NYU's English ACE 2005 system description. In *Proceedings of ACE 2005 evaluation workshop*, Gaithersburg, Maryland.
- Gupta, P., & Ji, H., 2009. Predicting unknown time arguments based on cross-event propagation. In *Proceedings of the joint conference of the 47th annual meeting of the ACL and the 4th international joint conference on natural language processing of the AFNLP (ACL-IJCNLP 2009)* (pp. 369–272). Suntec, Singapore.
- Hardy, H., Kanchakouskaya, V., & Strzalkowski, T., 2006. Automatic event classification using surface text features. In *proceedings of aaai 2006 workshop on event extraction and synthesis* (pp. 36–41). Boston, Maryland.
- Hong, Y., Zhang, J., Ma, B., Yao, J., Zhou, G., & Zhu, Q., 2011. Using cross-entity inference to improve event extraction. In *Proceedings of the 49rd annual meeting of the association for computational linguistics (ACL 2011)* (pp. 1127–1136). Portland, Oregon.
- Ji, H., 2009. Cross-lingual predicate cluster acquisition to improve bilingual event extraction by inductive learning. In *Proceedings of NAACL HLT workshop on unsupervised and minimally supervised learning of lexical semantics* (pp. 27–35). Boulder, Colorado.
- Ji, H., & Grishman, R., 2008. Refining event extraction through cross-document inference. In *Proceedings of the 46rd annual meeting of the association for computational linguistics: human language technologies (ACL-HLT 2008)* (pp. 254–262). Columbus, Ohio.
- Kim, Young-Joo (2000). Subject/object drop in the acquisition of Korean: A cross-linguistic comparison. *Journal of East Asian Linguistics*, 9(4), 325–351.
- Kindermann, Ross, & Snell, Laurie James (1980). Markov random fields and their applications. *American Mathematical Society*.
- Klenner, M., 2007. Enforcing consistency on coreference sets. In *Proceedings of recent advances in natural language processing (RANLP 2007)* (pp. 323–328). Borovets, Bulgaria.
- Levy, R., & Manning, C., 2003. Is it harder to parse Chinese, or the Chinese Treebank?. In *Proceedings of the 41rd annual meeting of the association for computational linguistics (ACL 2003)* (pp. 439–446). Sapporo, Japan.
- Li, Z., 2011. Parsing the internal structure of words: a new paradigm for Chinese word segmentation. In *Proceedings of the 49rd annual meeting of the association for computational linguistics (ACL 2011)* (pp. 1405–1414). Portland, Oregon.
- Li, P., Zhou, G., Zhu, Q., & Hou, L., 2012. Employing compositional semantics and discourse consistency in Chinese event extraction. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL 2012)* (pp. 1006–1016). Jeju Island, Korea.
- Liang, P., Joedan, M., & Klein, D., 2011. Learning dependency-based compositional semantics. In *Proceedings of the 49rd annual meeting of the association for computational linguistics (ACL 2011)* (pp. 590–599). Portland, Oregon.
- Liao, S., & Grishman, R., 2010. Using document level cross-event inference to improve event extraction. In *Proceedings of the 48rd annual meeting of the association for computational linguistics (ACL 2010)* (pp. 789–797). Uppsala, Sweden.
- Liao, S., & Grishman, R., 2011. Using prediction from sentential scope to build a pseudo co-testing learner for event extraction. In *Proceedings of the 5th international joint conference on natural language processing (IJCNLP 2011)* (pp. 714–722). Chiang Mai, Thailand.
- Liu, Q., & Li, S., 2002. Word similarity computing based on HowNet. In *Proceedings of 3th Chinese lexical semantic Workshop*. Taipei, Taiwan.

⁸ Li et al. (2012) only focus on expanding those verbal trigger words.

- Llorens, Hector, Saquete, Estela, & Navarro-Colorado, Borja (2013). Applying semantic knowledge to the automatic processing of temporal expressions and events. *Information Processing & Management*, 49(1), 179–197.
- Lu, W., & Roth, D., 2012. Automatic event extraction with structured preference modeling. In *Proceedings of the 50th annual meeting of the association for computational linguistics (ACL 2012)* (pp. 835–844). Jeju, Republic of Korea.
- Mann, G., 2007. Multi-document relationship fusion via constraints on probabilistic databases. In *Proceedings of human language technologies 2007 and the conference of the north american chapter of the association for computational linguistics (HLT/NAACL 2007)* (pp. 332–229). Rochester, New York.
- Maslennikov, M., & Chua, T.-S., 2007. A multi resolution framework for information extraction from free text. In *Proceedings of the 45rd annual meeting of the association for computational linguistics (ACL 2007)* (pp. 592–599). Prague, Czech Republic.
- Patwardhan, S., & Riloff, E., 2007. Effective information extraction with semantic affinity patterns and relevant regions. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL 2007)* (pp. 717–727). Prague, Czech Republic.
- Patwardhan, S., & Riloff, E., 2009. A unified model of phrasal and sentential evidence for information extraction. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing (EMNLP 2009)* (pp. 151–160). Singapore.
- Poon, H., & Domingos, P., 2007. Joint inference in information extraction. In *Proceedings of the 22nd national conference on artificial intelligence (AAAI 2007)* (pp. 913–918). Vancouver, Canada.
- Poon, H., & Vanderwende, L., 2010. Joint inference for knowledge extraction from biomedical literature. In *Proceedings of human language technologies: The 2010 annual conference of the North American Chapter of the association for computational linguistics (NAACL-HLT 2010)* (pp. 813–821). Los Angeles, California.
- Qin, Bin, Zhao, Yanyan, Ding, Xiao, Liu, Ting, Zhai, Guofu, (2010). Event type recognition based on trigger expansion. *Tsinghua Science and Technology*. 15 (3): 251–258, Beijing, China.
- Richardson, Matthew, & Domingos, Pedro (2006). Markov Logic Network. *Machine Learning*, 62, 107–136.
- Riedel, S., Chun, H.-W., Takagi, T., Tsujii, J., 2009. A Markov logic approach to bio-molecular event extraction. In *Proceedings of the workshop on current trends in biomedical natural language processing: Shared task (BioNLP 2009)* (pp. 41–49). Boulder, Colorado.
- Song, Y., Jiang, J., Zhao, W.X., Li, S., Wang, H., 2012. Joint Learning for Coreference Resolution with Markov Logic. In *Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL 2012)* (pp. 1245–1254). Jeju Island, Korea.
- Tan, H., Zhao, T., & Zheng, J., 2008. Identification of Chinese event and their argument roles. In *Proceedings of the 2008 IEEE 8th international conference on computer and information technology workshops* (pp. 14–19). Sydney, Australia.
- Wang, W., 2012. Chinese news event 5W1H semantic elements extraction for event ontology population. In *Proceedings of the 21st international conference companion on World Wide Web* (pp. 197–202). Beijing, China.
- Wong, Y.W., & Mooney, R., 2007. Learning synchronous grammars for semantic parsing with Lambda calculus. In *Proceedings of the 45rd annual meeting of the association for computational linguistics (ACL 2007)* (pp. 960–967). Prague, Czech Republic.
- Yangarber, R., & Jokipii, L., 2005. Redundancy-based correction of automatically extracted facts. In *Proceedings of the 2005 Joint conference on empirical methods in natural language processing (EMNLP 2005)* (pp. 57–64). Vancouver, Canada.
- Yangarber, R., Best, C., Etter, v.P., Fuat, F., Horby, D., & Steinberger, R., 2007. Combining information about epidemic threats from multiple sources. In *Proceedings of RANLP 2007 workshop on Multi-source, multilingual information extraction and summarization*. (pp. 41–48). Borovets, Bulgaria.
- Yarowsky, D., 1995. Unsupervised word sense disambiguation rivaling supervised methods. In *Proceedings of the 1995 annual meeting of the association for computational linguistics (ACL 1995)* (pp. 189–196). Cambridge, Massachusetts.
- Zettlemoyer, S.L., & Collins, M., 2007. Online Learning of Relaxed CCG Grammars for Parsing to Logical Form. In *Proceedings of the 2007 joint conference on empirical methods in natural language processing and computational natural language learning (EMNLP-CoNLL 2007)* (pp. 678–687). Prague, Czech Republic.
- Zhao, YanYan, Qin, Bing, Che, Wanxiang, & Liu, Ting, (2008). Research on Chinese event extraction. *Journal of Chinese Information Processing*, 22(1), 3–8, Beijing, China.
- Zhou, GuoDong, Li, Junhui, Fan, Jianxi, & Zhu, Qiaoming (2011). Tree kernel-based semantic role labeling with enriched parse tree structure. *Information Processing & Management*, 47(3), 349–362.
- Zhu, Dexi. (1980). *Research on Chinese modern grammars*. Beijing, China: Chinese Commerce and Trade Press.