An Investigation for Implicatures in Chinese : Implicatures in Chinese and in English are similar !

Lingjia Deng
Intelligent Systems Program
University of Pittsburgh
lid29@pitt.edu

Janyce Wiebe
Department of Computer Science
University of Pittsburgh
wiebe@cs.pitt.edu

Abstract

Implicit opinions are commonly seen in opinion-oriented documents, such as political editorials. Previous work have utilized opinion inference rules to detect implicit opinions evoked by events that positively/negatively affect entities (*goodFor/badFor*) to improve sentiment analysis for English text. Since people in different languages may express implicit opinions in different ways, in this work we investigate implicit opinions expressed via goodFor/badFor events in Chinese. The positive results have provided evidences that such implicit opinions and inference rules are similar in Chinese and in English. Moreover, we have observed cases where the inferences are blocked.

1 Introduction

In the opinion-oriented documents, many opinions are expressed implicitly rather than explicitly. Consider the following example from (Deng and Wiebe, 2014):

> EX(1.1) The reform would **lower** health care costs, which would be *a tremendous positive change* across the entire health-care system.

There is an explicit positive sentiment (*positive*) toward the event of *reform lower costs*. In expressing this sentiment, the writer implies he is negative toward the *costs*, because he's happy to see the costs being decreased. The writer may be positive toward *reform* since it conducts the *lower* event. Such inferences may be seen as opinion-oriented *implicatures* (i.e., defeasible inferences) [1].

We create an annotated corpus (denoted *DCW corpus*) (Deng et al., 2013)[2] and generalizes such events, defining a ***badFor (bf)*** event to be an event that negatively affects the object and a ***goodFor (gf)*** event to be an event that positively affects the object of the event. Here, *lower* is a bf event. According to the annotation scheme, ***goodFor/badFor*** (hereafter ***gfbf***) events have NP agents and objects (though the agent may be implicit), and the polarity of a gf event may be changed to bf by a ***reverser*** (and vice versa). We have developed a set of rules for inferring implicit sentiments, from explicit sentiments and gfbf events (Deng and Wiebe, 2014). We incorporate the rules into a graph-based model, which significantly improves classifying the sentiments toward agents and objects in the gfbf events.

The contribution of this work is investigating implicatures in a second language, specifically in Chinese. People in different languages may express implicit opinions in different ways, so it is better to first assess similarity of implicatures in the two languages, rather than to directly utilize the English resources. In this work we conduct an agreement study for gfbf information in Chinese. The good agreement scores provide evidence for the existence of similar implicature in Chinese. During the analysis of disagreement, we have observed interesting gfbf events triggered by Chinese syntax, which are rare in English but common in Chinese. We should provide additional guidance for such events when developing a Chinese gfbf manual in the future.

We run the graph-based model on the annotated Chinese corpus. The good evaluation results support our hypothesis that the inference rules in English apply for Chinese. Moreover, we have observed gfbf cases where the sentiment inferences are blocked, which are similar to what we have found in English (Wiebe and Deng, 2014).

[1] *Implicatures* "normally accompany the utterances of a given sentence unless special factors exclude that possibility (p. 39)." (Huddleston and Pullum, 2002)

[2] Available at: http://mpqa.cs.pitt.edu/

Proceedings of the 5th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, pages 8–17,
Baltimore, Maryland, USA. June 27, 2014.

Further, we analyze gfbf words and syntax of agents/objects in Chinese. Our analysis shows that it is feasible to extract components of Chinese gfbf events utilizing the existing resources. In the last section we briefly talk bout the Chinese explicit sentiment analysis.

2 Related Work

In addition to researches focusing on explicit sentiments (Wiebe et al., 2005; Johansson and Moschitti, 2013; Yang and Cardie, 2013), recently there are work investigating features that directly indicate implicit sentiments (Zhang and Liu, 2011; Feng et al., 2013), or working on inferring implicit opinions (Choi and Cardie, 2008; Zhang and Liu, 2011; Anand and Reschke, 2010; Reschke and Anand, 2011; Goyal et al., 2013). Different from their work, which do not cover all the inferences of implicit opinions over explicit opinions and gfbf events, we define a generalized set of inference rules and incorporate the rules into a graph-based model to achieve sentiment propagation between the agents and objects of gfbf events (Deng and Wiebe, 2014). The result shows that the graph-based model itself is able to assign the unknown nodes with correct labels 89% of the time.

Many works in Chinese sentiment analysis develop heuristics for adapting methods in English to methods appropriate for Chinese (Tsou et al., 2005; Wang et al., 2007; Li and Sun, 2007). Instead of projecting English methods and resources into Chinese versions, there are also works leveraging Chinese-English parallel corpus to assist Chinese sentiment analysis. Wan (2008) translates Chinese sentiment sentences into English and ensemble the sentiment classification results from both English and Chinese sentiment classifiers. Wan (2009) adopt co-training methods, utilizing labeled English sentences and unlabelled Chinese sentences. Lu et al. (2011) assumes parallel sentences in different languages bear the same sentiment. They utilize unlabelled Chinese-English parallel corpus to jointly improve sentiment classification in both languages. Boyd-Graber and Resnik (2010) present a generative model, jointly modeling topics that are consistent across languages, to improve sentiment rating predictions.

3 Implicature in Chinese

The definition of a gfbf event is from (Deng et al., 2013). A *goodFor (gf)* event is an event that positively affects an entity (similarly, for *badFor (bf)* events). A gfbf triple has the structure of ⟨ agent, gfbf, object⟩, though the agent can be implicit. For example, in the sentence from (Deng et al., 2013), "Repealing the Affordable Care Act (ACA) would hurt our economy.", there are two gfbf triples. One is ⟨Repealing the ACA, hurt, families, our economy⟩, which is a bf. The other is ⟨*implicit*, Repealing, the ACA⟩, which is bf and the agent is implicit. The DCW corpus contains manually annotated gfbf events, the gfbf polarities, the corresponding agents and objects and the writer's attitudes toward the agents and objects.

Because people in different languages may express their opinions in different ways. In this section, we conduct an agreement study for Chinese gfbf information in Section 3.1 and achieve good agreement scores, reported in Section 3.2, which provide supporting evidences for detecting Chinese gfbf events. In the disagreement analysis, we have observed interesting cases which are gfbf events in semantics but are triggered by Chinese own syntax. We explain the cases in Section 3.3.

3.1 Agreement Study Design

Data: We collect 100 political editorials from the Opinion Column in the Chinese version of New York Times[3], where each political editorial has an English version and a Chinese version. The Chinese editorial is a translated and paraphrased version of the corresponding English editorial, written by professional translators. The English version and the Chinese version are paragraph paralleled. In the previous agreement study of (Deng et al., 2013), the annotators are asked to annotate the whole document. Because not all the sentences contain gfbf events and the documents are long, a large proportion of disagreement we find that is due to negligence. In order to reduce negligence and provide a more dense data for annotation, first, we collect a lexicon of English gfbf words in the DCW corpus. Then we find the English sentences containing English gfbf words and select the paragraphs containing those sentences. The parallel Chinese paragraphs are collected. Though a paragraph may contain more than one sentence and some sentences do not have gfbf events, it is much more dense to annotate than the document as a whole. When presenting data to the annotators, we do not provide an isolated paragraph since it may

[3]http://cn.nytimes.com/opinion/

lose the context information. Instead, we present the original Chinese editorials and highlight the selected paragraphs. The annotators are told to read through the whole document but only need to annotate the highlighted paragraphs.

Procedure: We adopt our English manual in (Deng et al., 2013) to train the annotators. The annotators read through the manual and several Chinese gfbf examples. Then, the annotators label several paragraphs and discuss their disagreements to reconcile their differences. For the formal agreement study, we randomly selected 60 paragraphs, which have a total of 253 Chinese sentences. These paragraphs are different from the paragraphs discussed during training. The annotators then independently annotated the 60 selected paragraphs.

3.2 Agreement Study Evaluation and Result

We use the same measurement for agreement for all types of spans. (The type is either gfbf, agent, or object). Suppose A is a set of annotations of a particular type and B is the set of annotations of the same type from the other annotator. For any text span $a \in A$ and $b \in B$, the span coverage c counts the percentage of overlapping Chinese characters between a and b,

$$c(a,b) = \frac{|a \cap b|}{|b|} \quad (1)$$

where $|a|$ is the number of characters in span a, and $\cap$ gives the set of characters that two spans have in common (Johansson and Moschitti, 2013).

Following (Wilson and Wiebe, 2003), we treat each set A and B in turn as the gold-standard and calculate the average F-measure ($agr(A,B)$).

$$agr(A||B) = \frac{\sum_{a \in A, b \in B,\ |a \cap b|>0} c(a,b)}{|B|} \quad (2)$$

$$agr(A,B) = \frac{agr(A||B) + agr(B||A)}{2} \quad (3)$$

Now that we have the sets of annotations on which the annotators agree, we use κ (Artstein and Poesio, 2008) to measure agreement for the attributes. We report three κ values: one for the polarities of the gfbf events, and the other two for the writer's attitudes toward the agents and objects.

Three annotator participate in the agreement study. All of them are Chinese graduate students studying in US. One of them is the co-author of this work (*Anno 1*), while the other two do not know details of gfbf and implicature before (*Anno2, Anno3*). Since *Anno1* is familiar with this work, we compare the other two's annotations to *Anno1*'s. In Table 1, the upper half is the agreement for span overlapping ($agr(A,B)$), and the lower half is the agreement for attribute (κ).

$agr(A,B)$	gfbf	agent	object
Anno 1& 2	0.7929	0.9091	0.9091
Anno 1 & 3	0.7044	0.9524	1.0
κ	gfbf polarity	agent attitude	object attitude
Anno 1 & 2	0.9385	0.7830	0.7238
Anno 1 & 3	0.8966	0.5913	0.8478

Table 1: Results for Agreement Study Analysis.

The result have shown that the annotators have good agreement scores, though our training period is not long and our training data cover multiple topics. In particular, the annotators agree quite well on recognizing the agents and objects and judging the polarity of gfbf events.

For recognizing gfbf events, we have found two interesting gfbf cases caused by the Chinese syntax that is different from English, elaborated in the next section. Among the spans only one annotator marks, one third is due to the two cases above; one third are borderlines that could be marked; one third are incorrect. For the spans two annotator mark but the third doesn't, we regard it as negligence.

For judging the writer's attitudes toward agents and objects, we can see from Table 1 that *Anno 2* and *Anno 3* behave differently. This is understandable because we are marking the implicit opinions of the writer. Though trained, different annotators have different thresholds for judging whether an opinion is expressed here. Some annotators may be more sensitive than the others. If we don't count the spans that one annotator marks it as *none* (i.e. neutral) but the other doesn't, the κ scores increase a lot, as Row *Polar* shows in Table 2. This indicates that the annotators mainly disagree on whether the sentiment is neutral or not, rather than the polarity of opinions.

To further investigate whether the disagreement is caused by Chinese, or is due to the annotators' inherent different sensitivities of opinions, we randomly select 5 documents from the DCW corpus, delete the writer's attitude toward agents and objects but keep the remaining annotations. The an-

	Anno 1 & 2		Anno 1& 3	
	agent	object	agent	object
Table 1	0.783	0.723	0.591	0.848
Polar	0.875	0.915	1	0.88
Eng	0.738	0.652	0.4633	0.8734

Table 2: κ for Agreement Study Analysis.

notators are then told to mark the attitudes. As Row *Eng* in Table 2 shows, we have got consistent agreement results within the same annotators when they annotate in English and in Chinese. This supports the idea that the differences between the annotators are differences on the underlying task, regardless of the language.

3.3 GoodFor/Badfor Triggered by Chinese Syntax

During the analysis of disagreement, we have found gfbf cases which are triggered by the Chinese syntax that is different from English. Since the annotators are trained by the English manual, some annotators stay consistent with the English syntax, but the others go beyond syntax and identify gfbf according to semantics and pragmatics, which lead to disagreement. In this section we list two major cases due to the Chinese own syntax. This suggests that additional guidance to annotate such cases should be added to the English manual to develop a Chinese gfbf manual.

The first case is due to unclear expression of passive voice in Chinese. In English, the noun phrase that would be the object of an active sentence (Our troops **defeated** the enemy) appears as the subject of a sentence with passive voice (The enemy **was defeated** by our troops)[4]. It is clear that *enemy* is the object and *our troops* is the agent in both sentences. However, this is not intuitive for some Chinese sentences.

A Chinese example is " 经济潜力似乎**得以释放** ", whose English translation is: "The economic potential ... appeared to **be unleashed**". A word-to-word translation would be "...appeared to **have got unleashed**". In the two English versions, *potential* is obviously the object of *unleashed* event. However, some annotators analyze this sentence according to syntax[5]. The dependency syntax between the object *potential* (潜力) and the gfbf *unleash* (释放) is **nsubj**(释放-5, 潜力-2) so it is not marked. Some annotators view from pragmatics and read as a passive voice. Since there is no word transformation of Chinese verbs for passive voice (e.g. *unleash* changes to *unleashed* in English), this raises disagreement.

The other case is related to one constraint defined in (Deng et al., 2013). According to the manual, the polarity of a gfbf triple must be determined within the triple. As explained in the manual, in the sentence "Tom has left his cousin a big trouble", the triple ⟨Tom, left, his cousin⟩ is not a gfbf event, since we cannot judge whether this event is good for or bad for his cousin without knowing what Tom leaves to his cousin. While in the sentence "They decrease the manufacturing costs", the event *decrease* is a bf no matter how many or by what means the costs are decreased.

However, a Chinese instance is, "**把**改革**置于死地** ", whose translation is "**put** the reform **to die**". Whether the event *put* (把) is good for or bad for the object *reform* (改革), depends on whether the agent puts the reform to die or puts the reform to revive, for instance. However, in Chinese, 把 is not main verb (Li and Thompson, 1989), the object (改革, reform) of the main verb (置于死地, die) is placed after the function word (把), and the verb is placed after the object, forming a subject–object–verb (SOV) sentence (Chao, 1968)[6], which is defined as *ba structure* (Chao, 1968; Li and Thompson, 1989; Sybesma, 1992). Thus, in Chinese the sentence is read as: "kill the reform", which could be seen as a gfbf event. This structure is very common in Chinese.

In conclusion, there are very similar implicatures in Chinese. However, in order to fully study the gfbf events in Chinese, the manual should be revised to provide guidance for annotating the cases mentioned above.

4 Implicature Inference in Chinese

We propose a set of sentiment inference rules and incorporate them into a graph-based model to conduct sentiment propagation among entities (agents and objects) of gfbf events (Deng and Wiebe, 2014). In Section 4.1, we run this graph-based model on the Chinese annotations. The positive results of sentiment propagation support our hypothesis that the inference rules apply for Chinese as well. Further, we categorize interesting gfbf cases where the inferences are blocked in Section

[4]http://en.wikipedia.org/wiki/English_passive_voice.

[5]We use Stanford's dependency parser in this work.

[6]http://en.wikipedia.org/wiki/B%C7%8E_construction.

4.2. From our observation, the blocking inferences are similar to what we have found in English (Wiebe and Deng, 2014).

4.1 Graph-based Model

In the graph-based model, a node represents an entity (agent, or object), and an edge exists between two nodes if the two entities participate in one or more gfbf events with each other. Scores on the nodes represent the explicit sentiments, if any, expressed by the writer toward the entities. Scores on the edges are based on constraints derived from the rules. Loopy Belief Propagation (Pearl, 1982; Yedidia et al., 2005) is applied to accomplish sentiment propagation in the graph. Given a graph built from manually annotations, an evaluation is carried out to assess the ability to propagate sentiment of the model. In the study, for each subgraph (connected component), we assign one of the nodes in the subgraph with its gold-standard polarity. Then we run LBP on each node in the subgraph. The experiment is run on the subgraph $|S|$ times, where $|S|$ is the number of nodes in the subgraph. Therefore, each node is assigned its gold-standard polarity exactly once, and each node is given a propagated value $|S|-1$ times, as propagated by each of the other nodes in its subgraph. We use Equations (4) and (5) to evaluate the chance of a node given a correct propagated label.

$$correct(a|b) = \begin{cases} 1 & \textit{a is correct} \\ 0 & \textit{otherwise} \end{cases} \quad (4)$$

$$correctness(a) = \frac{\sum_{b \in S_a, b \neq a} correct(a|b)}{|S_a| - 1} \quad (5)$$

Here we run the graph-based model on the Chinese annotations. The data we use include the training and testing paragraphs in the agreement study, in total 85 paragraphs, 341 sentences and 160 gfbf triples. Later we use this corpus of 160 gfbf triples for analysis (denoted *Chinese gfbf corpus*). Since the edge scores of the model are defined according to the inference rules, if the sentiments are propagated correctly, this is a good evidence that the inference rules apply to Chinese.

The performances of the sentiment propagation are really good, reported in Table 3. The model has an 70%-83% chance of propagating sentiments correctly in Chinese. This gives us confidence that the inference rules apply in Chinese and further we can utilize these rules to assist Chinese sentiment analysis. Compared to the scores of *correctness* reported in (Deng and Wiebe, 2014), which are 0.8874 for all subgraphs and 0.9030 for multi-node subgraphs, our scores are lower. We analyze the reasons for the gap between our scores in Chinese and in English in the next section.

Dataset	# subgraph	correctness
all subgraphs	136	0.7058
multi-node subgraphs	61	0.8251

Table 3: Performance of Graph-Based Model in Chinese.

4.2 Blocking the Inference

A wrong propagation indicates the inferences related to that propagation are blocked. During the error analysis, we have found three interesting categories of cases where the inferences are blocked. Interestingly, we have observed these cases in English as well (Wiebe and Deng, 2014). In other words, we didn't find any blocking case specific to Chinese. The lower scores of *correctness* in Chinese might be due to the smaller amount of experiment data and more blocking cases in this corpus.

Irrealis: This category contains gfbf events that haven't or will not happen. One of the case is when the agent tried to conduct the gfbf event, but failed. In Ex(4.1), the agent and objective are underlined and the gfbf event is boldfaced. By the rules, the writer has the same sentiment toward the agents and objects in gf events and opposite sentiments toward the agents and objects in bf events (Deng and Wiebe, 2014). In Ex(4.1), the writer is negative toward both the agent and the object, though this is a bf event. This is because the event *counter* does not exist due to the failure, which is implied by *intended to*. The inferences for gfbf events in this category are blocked because the writer expresses the sentiments toward entities based on what they have done so far.

> Ex(4.1) ...monetary policy activism intended to **counter** the cyclical bumps and grinds of the free market.

Forced GFBF: This category contains gfbf events whose agents don't intend to do that or being forced to conduct the event. For example, in Ex(4.2), though the triple ⟨Obama, delay, mandate⟩ is an event which does not happen, it

is different from Ex(4.1). Here, the agent *Obama* is forced to conduct the *delaying*, though he does not want to and the writer does not blame him if he does so. For the entities involved in forced events, (at least the writer believes the entities are involuntary,) the forced event will not affect the writer's sentiments toward the entities so that the inferences are blocked.

> Ex(4.2) Some of them even seem to think that they can bully Mr. Obama into **delaying** the individual mandate too.

Quoted GFBF: This category contains gfbf events in the quotations. Consider the Ex(4.3), where one of the gfbf triple is ⟨law, reduce, amount of labor ⟩. In the original editorial, the writer supports *the law* and the writer has a positive sentiment toward *the number of jobs* (because he/she expects to see more job opportunities). But merely from the annotated gfbf triple, it is inferred that the law has negative effect since it reduces the number of jobs. This is not contradictory with the writer's stance because the writer regards the event as *a deliberate misreading* he/she doesn't believe. The actual agent of the event should be (misreading, Obama). This example shows that inferences of a triple in the quotation are blocked, or event flipped, based on the writer's sentiment toward the agent saying the quotation. The agent in a quoted gfbf is similar to the notion of *nested source* in sentiment analysis (Wilson and Wiebe, 2003).

> Ex(4.3) Some of the job-killer scare stories are based on *a deliberate misreading* that estimated the law would "**reduce** the amount of labor used in the economy" by about 800,000 jobs.

In conclusion, the good performance in our pilot study gives supporting evidence for our hypothesis. That is, the inference rules apply for Chinese. Moreover, there is no evidence showing that the cases where the inferences are blocked only happen in Chinese.

5 Chinese GoodFor/BadFor Lexicon

Above all we have assessed the similarity of implicatures and inference rules in Chinese and English. In the following sections, we will analyze whether Chinese gfbf components could be captured by similar techniques in English.

Description	Count (Percentage %)	
Parallel Span	122	(76.25%)
Chinese Adding GFBF	10	(6.875%)
Chinese Adding Object	6	(3.75%)
English Out Of Triple	5	(3.125%)
English Neutral	6	(3.125%)
Paraphrase	11	(6.875%)

Table 4: Counts of Chinese-English Corresponds

In this section, we compare the gfbf spans in the Chinese gfbf corpus and the English version, to investigate the possibility of deriving a bilingual gfbf lexicon. Though the Chinese and English editorials are paragraph paralleled, they are not sentence paralleled, because an English sentence may be translated into multiple Chinese sentences and several English sentences may be merged into one Chinese sentence. Therefore, instead of automatic word-alignment, we manually pick up the English parallel spans of the Chinese annotated gfbfs. The correspondences of Chinese and English spans are categorized in Table 4. We present pairs of examples from the Chinese gfbf corpus, beginning with the original English sentence (*Eng*), followed by another English sentence which is the word-by-word translation of the Chinese sentence (*Chi*).

Parallel Span: This category contains instances where the Chinese annotated gfbf spans have the corresponding translations in the English sentences, and the English spans are also gfbf words.

Chinese Adding GFBF: In the original English sentence below, *its own making* is a noun phrase rather than a gfbf verb used as a noun. However, in the Chinese version, there is a clear triple, ⟨itself, makes, a monetary prison⟩. In such case the Chinese version adds a gfbf event into the sentence.

Eng: ...the Fed is domiciled in a monetary prison of **its own making**.
Chi: ...the Fed is domiciled in a monetary prison **which itself makes**.

Chinese Adding Object: As stated in the manual, all gfbf triples should have objects. Thus, in the original sentence below, we will not mark *exclusion* because the object is implicit. However, the Chinese version clearly states the object, *patients*.

Eng: ...no more exclusion based on pre-existing conditions...
Chi: ...no more exclusion **of the patients** based on pre-existing conditions...

English Out Of Triple: Recall from Section 3.3, the gfbf polarity must be sufficient to perceive the gfbf polarity within the triple. The ⟨the Fed, get, unemployment⟩ below cannot be considered as a gfbf, since whether it is good for or bad for the *unemployment* depending on whether it is **below** 6.5% or **up** 6.5%, for instance. On the contrary, the Chinese version uses the word *decrease*, which is a bf word, no matter how many percents are changed.

Eng: If and when the Fed — which now promises to **get** unemployment **below 6.5%**...
Chi: If and when the Fed — which now promises to **decrease** the unemployment to 6.5%...

English Neutral: Sometimes the English word doesn't have a gfbf meaning but the Chinese word has one, based on the translator's interpretation of the whole editorial, though the triple structures are the same in English and Chinese versions.

Eng: We've **had eight decades of** increasingly frenetic monetary policy activism...
Chi: We've been **insisting** increasingly frenetic monetary policy activism for eight decades...

In the original English sentence, *had eight decades of* is hardly regarded as a gfbf word. However, in the translated version, the word *insisting* is a gf word. The change of wording introduces a new gfbf event into the sentence.

Paraphrase: There are other cases where the sentences are paraphrased so largely that we cannot find a corresponding parallel span of the annotated Chinese span in the original English sentence. A majority of cases in this category are gfbf events triggered by the Chinese syntax in Section 3.3.

In conclusion, the percentage of 76.25% in Row *Parallel Span* indicates that it is applicable to derive a bilingual gfbf lexicon from a parallel corpus. However, we need to take into consideration the 23.75% mismatches for higher precision.

5.1 Chinese Reversers

The polarity of a gfbf event could be changed by a reverser (Deng et al., 2013). A common class of reversers is negation. For example, in the sentence, "the bill will not increase the costs", the gf *increase* is changed to be bf via the negation *not*. In this section, we analyze the Chinese reversers.

All of the reversers in the Chinese gfbf corpus happen to be negations. In the English sentences, the negations are easily extracted by *neg* dependency relation. About 50% of the Chinese negations are linked to the gfbf events via *neg* as well. Among this half, there are two negations commonly seen. One is 不 (Not), often labeled as AD (adverb) in terms of Part-Of-Speech, the other is 没有 (do not have), labeled as VV (verb), shown below. The negation is underlined and the gfbf event it negates is boldfaced.

Ex(5.1)不/AD **接受/VV** 同性恋/NN

Ex(5.2)没有/VV **刺激/VV** 贷款/NN

For the other half, the error mostly arises from segmentations. For the sentence below, though 没有 (*doesn't have*), often labeled as VB, could be regarded as a complete token, if we segment the two characters into two independent tokens, the parse is more similar to the English one. Below we only list the most relevant part of the parses.

Eng: He does n't have ability control war budget
Eng dep: neg(have-4, n't-3), root(ROOT-0, have-4), dobj(have-4, ability-6)
Chi: 他 没 有 能力 控制 战争 预算
wrong dep: root(ROOT-0, 没有-2), nsubj(控制-4, 能力-3), dep(没有-2, 控制-4)
correct dep: neg(有-3, 没-2), root(ROOT-0, 有-3), nsubj(控制-5, 能力-4)

In conclusion, it is feasible to recognize reversers in Chinese but it calls for a suitable word segmentation as input.

6 Syntax of Agent/Object in Chinese

According to (Deng et al., 2013), the agent is the entity conducting the gfbf event and the object is the entity that the gfbf event affects. This definition is very similar to subject and (in)direct object in semantic role labeling. Xue and Palmer (2004) investigate the Chinese semantic role labeling. They utilize the PropBank and the constituency parser. However, from a preliminary analysis of constituency parse, we cannot distinguish the agent and object merely from the parse tree, because the sentences in the editorials are usually complicated and it is difficult to classify whether a noun phrase (NP) constituency is agent or object in terms of its position. Kozhevnikov and Titov (2013) adopt a model transfer between different languages using dependency parser. In our case, the dependency parser has labels such

as "nsubj" and "dobj", which are strong indications of agents and objects. Thus, we use the Stanford dependency parser, which has both English and Chinese parsers, to analyze the syntax of agents/objects in the gfbf events. We count the types of dependencies on the path in a dependency parse between the tokens of agents/objects and the tokens of gfbf events in the DCW corpus and the Chinese gfbf corpus.

Among all the dependency types, 19.57% of the labels between agents and gfbfs are the ones specially designed for Chinese and 25.82% between objects and gfbf are the ones specially designed for Chinese. This indicates there is a considerable number of differences in dependency types. Chang et al. (2009), who create the Chinese parser, discuss the differences between Chinese and English types, which are similar to our observations.

First, there are more *nsubj* in Chinese for agents (21.53%) and more *dobj* in Chinese for objects (21.59%), compared to English (17.43% and 14.01%), which are easier for the parser to detect.

Second, the most common types specially designed for Chinese are *assm*, *assmod* and *cpm* (in total 12.23% for agents and 16.14% for objects). The relations *assm* is associative marker, *assmod* is associative modifier, and *cpm* is complementizer. These are defined because of the frequent usage of 的 (whose, of) in Chinese. Though there is not a direct mapping between Chinese and English dependency types, they are similar to two common types in English: *prep* and *pobj* (together 23.36% for agents and 31.62% for objects).

Third, there are more *rcmod* in Chinese than those in English. There are 7.05% and 6.5% *rcmod* in Chinese agents and objects, respectively. But there are only 1.7% and 2.16% in English agents and objects. The type *rcomd* is a relative clause modifier. If a verb is used as the modifier of a noun, it will be labelled *rcmod*. Instead, English writers tend to use more adjectives to modify nouns, which will be labeled *amod* (4.04% and 4.48%).

Fourth, there are 7.63% and 6.22% *punct* in Chinese agents and object, compared to both 0% in English. In addition, there are 3.36% and 3.31% *conj* in English agents and objects. Chang et al. (2009) explain that English use conjunctions (*conj*) to link clauses while Chinese tend to use punctuation. Another finding in our corpus is that, translators tend to break down a long English sentence into several Chinese clauses, linked by punctuations.

For the other Chinese types, most of them are modifiers, which may be grouped with similar English modifiers.

7 Chinese Explicit Sentiment Analysis

There are various available resources for Chinese sentiment analysis, such as sentiment lexicon from HowNet[7], NTU Sentiment Dictionary (NTUSD) (Ku and Chen, 2007)[8] and the sentiment lexicon from Tsinghua University (Li and Sun, 2007). The sentiments recognized from lexicon hits are explicit, meaning that the writers use sentiment words to express his/her opinions. These explicit sentiment results are provided to the graph-based model as input. Note that the model plays a role of sentiment inference, instead of directly detecting sentiments from the text. The inferred sentiments are implicit, meaning that the writers express his/her opinions even without using a sentiment lexical clue.

8 Conclusion

In this work we investigate implicit opinions expressed via goodFor/badFor events in Chinese. The positive results have provided evidences that such implicit opinions and inference rules are similar in Chinese and English. There are some gfbf events caused by the Chinese syntax, guidance for which could be added to the current English manual to develop a Chinese manual. Moreover, there is no evidence showing that the blocked inferences only happen in Chinese. We also assess the feasibility of acquiring components of gfbf events from Chinese text using current available resources. In the future, it is promising to utilize gfbf information to assist sentiment analysis in Chinese.

Acknowledgement This work was supported in part by DARPA-BAA-12-47 DEFT grant #12475008 and National Science Foundation grant #IIS-0916046. We would like to thank Changsheng Liu and Fan Zhang for their annotations in the agreement study, and thank anonymous reviewers for their feedback.

[7] Available at: http://www.keenage.com/html/e_index.html
[8] Available at: http://nlg18.csie.ntu.edu.tw:8080/lwku/pub1.html

References

Pranav Anand and Kevin Reschke. 2010. Verb classes as evaluativity functor classes. In *Interdisciplinary Workshop on Verbs. The Identification and Representation of Verb Features*.

Ron Artstein and Massimo Poesio. 2008. Inter-coder agreement for computational linguistics. *Comput. Linguist.*, 34(4):555–596, December.

Jordan Boyd-Graber and Philip Resnik. 2010. Holistic sentiment analysis across languages: Multilingual supervised latent dirichlet allocation. In *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, pages 45–55, Cambridge, MA, October. Association for Computational Linguistics.

Pi-Chuan Chang, Huihsin Tseng, Dan Jurafsky, and Christopher D Manning. 2009. Discriminative reordering with chinese grammatical relations features. In *Proceedings of the Third Workshop on Syntax and Structure in Statistical Translation*, pages 51–59. Association for Computational Linguistics.

Yuen Ren Chao. 1968. *A grammar of spoken Chinese*. Univ of California Press.

Yejin Choi and Claire Cardie. 2008. Learning with compositional semantics as structural inference for subsentential sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 793–801, Honolulu, Hawaii, October. Association for Computational Linguistics.

Lingjia Deng and Janyce Wiebe. 2014. Sentiment propagation via implicature constraints. In *Meeting of the European Chapter of the Association for Computational Linguistics (EACL-2014)*.

Lingjia Deng, Yoonjung Choi, and Janyce Wiebe. 2013. Benefactive/malefactive event and writer attitude annotation. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 120–125, Sofia, Bulgaria, August. Association for Computational Linguistics.

Song Feng, Jun Sak Kang, Polina Kuznetsova, and Yejin Choi. 2013. Connotation lexicon: A dash of sentiment beneath the surface meaning. In *Proceedings of the 51th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Sofia, Bulgaria, Angust. Association for Computational Linguistics.

Amit Goyal, Ellen Riloff, and Hal Daumé III. 2013. A computational model for plot units. *Computational Intelligence*, 29(3):466–488.

Rodney D. Huddleston and Geoffrey K. Pullum. 2002. *The Cambridge Grammar of the English Language*. Cambridge University Press, April.

Richard Johansson and Alessandro Moschitti. 2013. Relational features in fine-grained opinion analysis. *Computational Linguistics*, 39(3).

Mikhail Kozhevnikov and Ivan Titov. 2013. Cross-lingual transfer of semantic role labeling models. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1190–1200, Sofia, Bulgaria, August. Association for Computational Linguistics.

Lun-Wei Ku and Hsin-Hsi Chen. 2007. Mining opinions from the web: Beyond relevance retrieval. *Journal of the American Society for Information Science and Technology*, 58(12):1838–1850.

Jun Li and Maosong Sun. 2007. Experimental study on sentiment classification of chinese review using machine learning techniques. In *Natural Language Processing and Knowledge Engineering, 2007. NLP-KE 2007. International Conference on*, pages 393–400. IEEE.

Charles N Li and Sandra A Thompson. 1989. *Mandarin Chinese: A functional reference grammar*. Univ of California Press.

Bin Lu, Chenhao Tan, Claire Cardie, and Benjamin K Tsou. 2011. Joint bilingual sentiment classification with unlabeled parallel corpora. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 320–330. Association for Computational Linguistics.

J. Pearl. 1982. Reverend bayes on inference engines: A distributed hierarchical approach. In *Proceedings of the American Association of Artificial Intelligence National Conference on AI*, pages 133–136, Pittsburgh, PA.

Kevin Reschke and Pranav Anand. 2011. Extracting contextual evaluativity. In *Proceedings of the Ninth International Conference on Computational Semantics*, IWCS '11, pages 370–374, Stroudsburg, PA, USA. Association for Computational Linguistics.

Rintje Pieter Eelke Sybesma. 1992. *Causatives and accomplishments: The case of Chinese ba*, volume 1. Holland Institute of Generative Linguistics.

Benjamin KY Tsou, Raymond WM Yuen, Oi Yee Kwong, TBY La, and Wei Lung Wong. 2005. Polarity classification of celebrity coverage in the chinese press. In *Proceedings of International Conference on Intelligence Analysis*.

Xiaojun Wan. 2008. Using bilingual knowledge and ensemble techniques for unsupervised Chinese sentiment analysis. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 553–561, Honolulu, Hawaii, October. Association for Computational Linguistics.

Xiaojun Wan. 2009. Co-training for cross-lingual sentiment classification. 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: Volume 1-Volume 1*, pages 235–243. Association for Computational Linguistics.

Suge Wang, Yingjie Wei, Deyu Li, Wu Zhang, and Wei Li. 2007. A hybrid method of feature selection for chinese text sentiment classification. In *Fuzzy Systems and Knowledge Discovery, 2007. FSKD 2007. Fourth International Conference on*, volume 3, pages 435–439. IEEE.

Janyce Wiebe and Lingjia Deng. 2014. An account of opinion implicatures. arXiv:1404.6491v1 [cs.CL].

Janyce Wiebe, Theresa Wilson, and Claire Cardie. 2005. Annotating expressions of opinions and emotions in language ann. *Language Resources and Evaluation*, 39(2/3):164–210.

Theresa Wilson and Janyce Wiebe. 2003. Annotating opinions in the world press. In *Proceedings of the 4th ACL SIGdial Workshop on Discourse and Dialogue (SIGdial-03)*, pages 13–22.

Nianwen Xue and Martha Palmer. 2004. Calibrating features for semantic role labeling. In *EMNLP*, pages 88–94.

Bishan Yang and Claire Cardie. 2013. Joint Inference for Fine-grained Opinion Extraction. In *Proceedings of ACL*, pages 1640–1649.

Jonathan S Yedidia, William T Freeman, and Yair Weiss. 2005. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory, IEEE Transactions on*, 51(7):2282–2312.

Lei Zhang and Bing Liu. 2011. Identifying noun product features that imply opinions. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 575–580, Portland, Oregon, USA, June. Association for Computational Linguistics.