Annotation Projection of Textual Negation

Examination Number: B083090

Word Count: 9620

MSc Speech and Language Processing

The University of Edinburgh

2016

Acknowledgments

Table of Contents

Abstract

The current dissertation has two research goals: (1) to create linguistically-informed knowledge representation of negation in the narrative domain in Chinese (HB-neg-ch) (2) to project negation annotation from English to Chinese, and provide quantitative and qualitative error analysis. The annotation guidelines and the projection results reveal the different ways English and Chinese express negation. The necessity of linguistic analysis in creating additional language resources is thus highlighted.

Keywords: annotation guidelines, annotation projection, negation

1. Introduction

Negation has been widely studied in linguistics and philosophy as an operator that reverses the polarity of a statement. The incorporation of negation into the knowledge representation of semantic relations is crucial for a range of natural language processing (NLP) tasks (Blanco & Moldovan, 2011) especially when we would like to know whether an action or a state described in a statement is factual or counter-factual.

There are three key elements in negation: cue, scope and event (Morante et al., 2011). A cue is the marker that expresses negation. Examples of cues are "no", "not", "nothing" in English, and 不 *bu4* ("not")*,* 没有 *mei2you3* ("did not") in Chinese. A scope is the part of the sentence that is affected by the polarity change associated with the cue. An event is the specific action or state that is factually negated in the scope. For example, the cue in the sentence "I do not eat fish" is "not", and the scope that undergoes the change of polarity is "I do eat fish" in which the negated event is "eat".

However, there has not been a corpus annotated with all of the three elements of negation in a language other than English. The current study fills the research gap to present the first linguistically-informed knowledge representation of negation annotated for cue, event and scope in the narrative genre of Chinese. Both the

guidelines and the annotation are provided while the difficulties and challenges in applying an English annotation scheme to another language are systematically examined. To explore more convenient ways of creating annotation and to quantify the linguistic differences between English and Chinese, the current study has a second research goal to project the negation annotation from English to Chinese using a word alignment model trained from parallel corpora. Chinese is chosen to be the first foreign language for this annotation projection task because of its similar syntactic order to English. Error analysis of the projection results will also be provided to better understand the hindrances to negation projection and translation. The resources and methods developed in this study will be useful for a variety of NLP tasks such as machine translation (MT), information extraction, question answering and so on.

In MT, for example, negative sentences are usually translated worse with lower BLEU scores than positive sentences (Fancellu and Webber, 2014; Wetzel and Bond, 2012). Negation is particularly hard for MT because languages differ considerably in expressing negation. For example, Chinese doe not have an equivalent form for the negative pronoun "nothing" in English. As a result, Google translate misses the cue in translating the English sentence "we bought nothing" into 我 们 买 了 什 么 *wo3men mai3 le shen2me* ("We bought anything/what.").

There are currently no evaluation metrics that could tell us how well a MT system translates negative sentences. The traditional MT metric (BLEU) measures the n-gram overlap between MT output and translation references without providing any semantic analysis. The annotation of negation presented in the current study can be used to create evaluation metrics targeted for negation. For example, we could compare the negation annotation of MT output and references, and categorize errors according to cue, event and scope.

The main body of the thesis is organized into 8 sections. The next section surveys related work in projects of negation annotation and annotation projection. Section 3

introduces the annotation corpus and most importantly presents a summary of the annotation guidelines (the full guidelines can be found in appendix A). Section 4 lays out the methodology in the projection task. Section 5 introduces the evaluation metrics, soft comparison and strict comparison F1, for measuring projection quality. Section 6 presents the results of the projection with both qualitative and quantitative error analysis. The implications and future research recommendations are highlighted in section 7. The final section concludes the thesis with a brief summary of the findings and implications.

2. Related work

2.1. Negation corpora

Work in negation representation in information technology originated from the biomedical domain. The pioneer BioScope corpus (Vincze et al., 2008) provides annotation of negation/speculative cues and their scopes on clinical free texts, biological abstracts and biological full papers. The corpus facilitated research in cue and scope detection in biomedical information processing. The BioScope style of annotation has recently been adopted to create an equivalent Chinese corpus, CNeSp, that contains annotated scientific literature, product reviews and financial articles (Zhou et al., 2015).

The benefit from negation annotation in the biomedical domain has inspired research in developing annotation resources for the general domain. The most important work along this line is the ConanDoyle-neg corpus introduced by Morante & Daelemans (2012). The corpus contains annotated text from the Sherlock Holmes stories: The Hound of the Baskervilles (HB) and The Adventure of Wisteria Lodge (WL). ConanDoyle-neg has several crucial differences from BioScope. First of all, ConanDoyle-neg marks the event of a negation instance whereas BioScope doesn't. An event reveals the exact action or state that is factually negated. It is a crucial linguistic property of negation, and will be particularly useful in fact extraction.

Another difference is that the scope model in ConanDoyle-neg excludes the cue, but contains all arguments of the event including the subject. Moreover, a scope can be discontinuous, allowing for the inclusion of recoverable elliptical elements. Finally, ConanDoyle-neg annotates affixal negation which is absent in BioScope.

More crucially, ConanDoyle-neg is of the narrative genre which contains conversations and narrations with a mixture of colloquial and formal styles. There are also several unique linguistic constructions to this genre such as questions, quotations, ellipsis, which are not present in BioScope. The ConanDoyle-neg corpus contributed to the ultimate goal for a negation extractor: to deal with all possible negation occurrences in all possible domains in natural language.

However, the ConanDoyle-neg style of annotation is only available in English. It will be extremely useful if we could apply the annotation scheme to different languages and create additional language resources so that we will be able to bring the research boost in negation detection to other languages. The current study contributes to this goal by providing the first ConanDoyle-neg style of annotation in Chinese.

2.2. Annotation projection

Since there has not been any past studies on annotation projection of negation, the current section will give a brief overview of other methods of obtaining negation annotation that do not require labeled data, and experiments on projecting other semantic annotation cross-linguistically.

Since most research has been devoted to develop negation extraction algorithms in English, it will be good news if we could adapt the existing algorithms in English to other languages without training on labeled data in the target language. Such a model transfer has been performed in the biomedical domain. Skeppstedt (2011) adapted a regular expression algorithm called NegEx (Chapman et al., 2001) to a Swedish

biomedical corpus. A Swedish lexicon and a set of hand-crafted rules were constructed separately for Swedish. The results showed that the Swedish adaptation yielded a much lower precision score than the model's performance in English.

Similar results were reported from model transfer studies on negativeness detection tasks in sentiment analysis. It seems that adapting an existing model to other languages will inevitably result in much worse performance (Brook et al., 2009; Mihalcea et al., 2007). Researchers have turned to annotation projection for a more efficient way to leverage existing language resources. In fact, Mihalcea et al. (2007) compared the results from annotation projection and model transfer on the task of cross-lingual subjectivity extraction from English to Romanian. The performance of annotation projection (F-measure= 67.85) was much better than model transfer (F-measure = 47.93).

The most straightforward model of projection align words from the source language to words in the target language and transfers the annotation according to the word alignment. However, Hwa et al. (2002)'s work on the projection of parses from English to Chinese showed poor performance from using word alignment information alone (F-measure = 38.1). The projection performance was significantly improved after a set of linguistically-informed post-processing transformation rules were applied. For example, as English and Chinese are both head-initial in most constituents, there was a rule stating that whenever an English word is aligned to multiple Chinese words, the leftmost Chinese word should be the head of the phrase.

Another linguistically-informed improvement in the projection tasks comes from the incorporation of argument structures. For example, Padó & Lapata (2009) designed a constituent-based projection algorithm to project semantic role labels from English to German. Instead of matching words, the constituent-based method matches constituents using the word alignment information. Paired with perfect matching, the constituent-based projection performs the best (F-measure=75.1).

3. The annotation task

3.1. Creating the corpus

The translation (translated by Mengyuan Lin) used for the current annotation corpus HB-neg-ch, is one of the most popular versions of HB and can be downloaded from http://www.kanunu8.com/book3/8105/. The translator's footnotes are deleted before annotation.

HB-neg-ch follows the same format of the English annotation ConanDoyle-neg where each line corresponds to a token and each column contains one type of annotation, but the lemma, POS and parsing information (column 5-7) are left as "_". As Chinese has a regular pattern of sentence boundaries, the sentences in the corpus were segmented by regular expression.

As Chinese does not insert space between words, an additional step was taken to segment the words using Stanford segmenter 3.4.1 (Tseng et al., 2005). Here we chose the ctb version of the segmenter as it has shown better performance in machine translation (Aziz, 2016).

During the annotation process, we noticed that the segmented words are not always consistent nor intuitive. This is partly because there has not been consensus on how we could combine characters into words in Chinese. In fact, it has been suggested that the equivalent "sociological word" in Chinese should be the characters, as Chinese is a monosyllabic isolating language with one character usually presenting one morpheme (Chao, 1965:136). It is thus more natural to present linguistic annotation in Chinese on a character basis, and we decided to transfer the word-based annotation to character-based annotation with one annotation line per character. Below is an example annotation for the translation of "How is it that no one saw it?":

```
baskervilles03  7   0   "
baskervilles03  7   1   为    _   _   _   _   为   _
baskervilles03  7   2   什    _   _   _   _   什   _
baskervilles03  7   3   么    _   _   _   _   么   _
baskervilles03  7   4   别    _   _   _   _   别   _
baskervilles03  7   5   人    _   _   _   _   人   _
baskervilles03  7   6   就    _   _   _   _
baskervilles03  7   7   没    _   _   _   没   _   _
baskervilles03  7   8   有    _   _   _   有
baskervilles03  7   9   看    _   _   _   _   看   看
baskervilles03  7   10  到    _   _   _   _   到   到
baskervilles03  7   11  呢    _   _   _   _   _   _
baskervilles03  7   12  ?    _   _   _   _   _   _
baskervilles03  7   13  "    _   _   _   _   _   _
```

The annotation was carried out in the CoNLL format in excel by the author who is a native Mandarin speaker with linguistic background. The guidelines and annotation were also created by the author. Annotation problems were raised and discussed regularly with a PhD student and a professor from the Informatics School.

Information of the HB-neg-ch in comparison with the English HB corpus is presented in Table 3.1. We can already see from the table that there are more negation instances in Chinese than in English, as the percentage of negative sentences, the number of cues, scopes and events in Chinese are consistently higher than English.

Table 3.1. Corpus statistics of negation instances

HB corpora	#sent	#neg sent	#cues	#scopes	#events
English	3640	850	985	888	617
Chinese	3277	975	1211	1161	755

3.2. Some general principles

To serve the purpose of measuring projection quality, the current version of HB-neg-ch aims to achieve the maximal linguistic and translational equivalence with ConanDoyle-neg in presenting the sentential annotation of cue, event and scope in Chinese. The linguistic equivalence principle requires a strategy to apply rules from the English guidelines to the equivalent linguistic constructions in Chinese. For example, there are a series of rules (e.g. not annotating modal verbs as events except when they express the meaning of 'be able to') that deal with modal verbs in the English annotation. The same rules should be transferred to Chinese modal verbs. But this is not as easy as it seems, as not all Chinese linguists agree on the definition of Chinese modal verbs. Here, we chose a more semantically-defined set of modal verbs (Li, 2003) which also show good translational equivalence with the English modal verbs identified in ConanDoyle-neg.

The translational equivalence is another important principle in creating the current annotation especially when dealing with Chinese constructions that are not present in English. For example, Chinese has negated prepositions as in (1)[1], and can also move the cue to negate the verb phrase as in (2). Both (1) and (2) are translated into the same English sentence where the event is the main verb "see". Therefore, we decide to annotate the verb as the event when the negated preposition is attached to the verb. Accordingly, in both (1) and (2), 看到 *kan4dao4* ("see") will be annotated as the event. The translational equivalence principle will make the two languages more comparable for the projection task later on. Notice that for all the annotation examples displayed in this study, the cue is marked in bold; the scope is marked with an underline, and the event is marked in italics.

---

[1] For all the Chinese examples displayed in the study, pinyin (in italics) and word-by-word translation will be given. The English translation will be provided in parenthesis.

(1)

我	没有	在	报上	看到	他的	噩耗	。
wo3	*mei2you3*	*zai4*	*bao4shang4*	*kan4dao4*	*ta1de*	*jiang1hao4*	.
I	did NOT	in	newspaper	see	his	news of death	

("I did not see his death in the paper.")

(2)

我	在	报上	没有	看到	他的	噩耗	。
wo3	*zai4*	*bao4shang4*	*mei2you3*	*kan4dao4*	*ta1de*	*jiang1hao4*	.
I	in	newspaper	did NOT	see	his	news of death	

("I did not see his death in the paper.")

There are also some annotation decisions and guideline rules in English that either we do not fully agree with or are inconsistent themselves. A general rule of thumb is to discuss the disagreement in the guidelines, but follow the English annotation if the above two principles are satisfied and the annotation is consistent.

When inconsistencies occur in the guidelines or between the guidelines and the actual annotation, we should always rely on the annotation if it is consistent. If the annotation is itself inconsistent, we could either follow the majority of the annotation cases or rely on the language intuition from a native speaker.

3.3. Summary of the annotation guidelines

3.3.1. Cue

As defined in Morante et al. (2011), a cue is "the word that expresses negation". There are in total 45 types of cues and 10 types of core negation cues in the corpus. The core negation cues here refer to the single-morpheme negation markers in Chinese. They are typically one-character word (except for 没有 *mei2you3* ("did/have not")) that can be combined with bound morphemes to form multiword cues. The core negation cues

can be adverbs, auxiliary verbs and prefixes. Unlike English, Chinese does not have negation pronouns (e.g. "nothing") and negation prepositions ("without"). The most common core negation cue in Chinese is 不 *bu4* ("not"), an adverb that is able to negate all verbs and adjectives but 有 *you3* ("to have/to exist"). An example is in (3).

(3)

我	不	*知道*	*应该*	*相信*	*什么*	。
wo3	*bu4*	*zhi1dao4*	*ying1gai1*	*xiang1xin4*	*shen2me*	。
I	NOT	know	should	believe	what	.

("I don't know what I should believe.")

As an isolating language, Chinese has fewer affixes. The definition of an affix is also problematic in Chinese, as some common Chinese affixes that occur in fixed locations within a compound can also function as stand-alone free morphemes in other cases. Here we employ a rather broad definition of negative affixes that refer to local negation cues within a compound.

Except 别 *bie2* ("not"), all the other core negation cues can function as prefixes in a compound such as 没/无用 *mei2/wu2yong4* ("useless"). Interestingly, some of them are created through translation: e.g. 无线 *wu2xian4* ("no wire=wireless"), 非物质 *fei1wu4zhi4* ("not physical=non-physical"), 不道德 *bu4dao4de2* ("immoral").

Compared to English, the negation in negative affixes in Chinese compounds is much more active than the negative affixes in English. Chinese allows speakers to read meaning from every character and a character is more able to preserve its original meaning through history (Chao, 1965). There are cases such as "disgrace" in English where the negation meaning of "dis" only exists etymologically. By contrast, besides a small number of exceptions, Chinese negative affixes are still expressing negation. Three principles (compositionality, obsolescence and translational equivalence) were designed to detect these exceptions. For example, the negation marker 无 *wu2* ("not")

in 无聊 *wu2liao2* ("no-chatting boring") is treated as a false negative affix because the meaning of the compound cannot be compositionally derived from the negative affix plus the rest.

There are also multiword cues that either have the form of "adverb + core negation cue" construction such as 并不 *bing4bu4* ("not"), 绝不 *jue2bu4* ("absolutely not"), or are compounds that contain false negative affixes but negate some other part of the sentence, for example, 不见得 *bu4jian4de2* ("probably not") and 不怎么 *bu4zen3me* ("not really").

Finally, there are false negatives that could introduce reinforced positive meaning such as double negatives around modal verbs (see (4)), or are fixed expressions that are never translated into negation in English, for example, 对不起 *dui4bu4qi3* ("sorry").

(4)

我	不得不	放弃	这	种	方法	。
wo3	*bu4de2bu4*	*fang4qi4*	*zhe4*	*zhong3*	*fang1fa3*	。
I	not-should-not	abandon	this	CL	method	.

("I am compelled to abandon this method").

Another false negative that is unique to Chinese is the "V+negation marker+V ?"construction in forming a yes-no question (see (5)). The negation marker in this construction introduces modality rather than genuine negation.

(5)

咱们	要	不	要	向	后	退	?
zai2men	*yao4*	*bu4*	*yao4*	*xiang4*	*hou4*	*tui4*	?
we	shall	NOT	shall	towards	back	retreat	?

("Shall we move farther back?")

3.3.2. Scope

The scope of negation should be the part of the sentence that is affected by the reverse of polarity. As the polarity of the cue is not changed, the cue should be left outside the scope (Morante et al., 2011). The equivalent of the "it is not the case" test in Chinese, 并 *bing4* 不 *bu4* 是 *shi4* / 非 *fei1* ("it is not the case"), can be applied to find the longest relevant scope. For example, the paraphrase of the sentence in (6) by adding 并不是 in front of the scope is: 一定 *yi1ding4* 并不是 *bing4bu4shi4* 我 *wo3* 明天 *ming2tian1* 来 *lai2* ("it is definitely not the case that I am coming tomorrow.").

It has been argued that the semantic scope of negation follows the negation cue in Chinese (Li & Thompson, 1989). However, in the context of knowledge representation, we need to also include in the scope model the subject, subject modifiers[2], and time/location adverbial phrases which are usually found before the cue in Chinese, as they provide critical information in the negation instance and are usually annotated as part of the scope in English. Other pre-cue words/phrases such as adverbs of degree and certainty are excluded from the scope annotation as in (6).

---

[2] Notice that quantifiers such as 都 *dou1* ("all"),有些 *you3xie1* ("some") are treated as subject modifiers and are thus included in the scope. However, from the semantic analysis, quantifiers should scope over negation.

(6)

我	明天	一定	不	来	。
wo3	*ming2tian1*	*yi1ding4*	*bu4*	*lai2*	。
I	tomorrow	definitely	NOT	come	。

("I am definitely not coming tomorrow.")

Also, the scope can be discontinuous allowing for the inclusion of recoverable elliptical elements. Compared to English, Chinese exhibits more versatile ellipsis constructions. For example, the object of the previous clause can be the elided subject in the next clause. An example is in (7).

(7)

这里	出现	了	一	位	青年人	，	不	到	三十岁	。
zhe4 li3	*chu1 xian4*	*le*	*yi1*	*wei4*	*qing1nian2 ren2*	，	*bu4*	*dao4*	*san1shi2sui4*	。
here	emerge	PFV	a	CL	young man	,	NOT	reach	thirty-years old	.

("There emerges a young fellow who is under thirty.")

Following the English guidelines, if the main verb is negated, then the whole sentence is in the scope. If the negated verb is in the subordinate clause, then only the subordinate clause is in the scope. Notice that the subordinate clauses in Chinese include both complete clauses that are introduced by subordinate conjunctions such as 如果 *ru2guo3* ("if") and the second verb phrases in serial verbal constructions as in (8)

(8)

我	本	想	找	借口	不	听	他	说	。
wo3	ben3	xiang3	zhao3	jie4kou3	bu4	ting1	ta1	shuo1	。
I	originally	think	find	excuse	NOT	listen	him	say	.

("I was originally thinking of finding excuses not to listen to him.")

Negation that occurs in a coordinate clause will scope over only the clause in which negation occurs plus any recoverable elliptical elements. The coordinate clauses are typically joined by coordinate conjunctions such as 并且 *bing4qie3* ("and"), 而且 *er2qian3* ("and also") in Chinese. An example is in (9).

(9)

那	人	吃了一惊	,	并且	有点	不知所措	。
na4	ren2	chi1leyi1jing1	,	bing4qie3	you3dian3	bu4zhi1suo3cuo4	。
that	person	get shocked	,	and	a little	NOT know what to do	.

("The man looked surprised and a little embarrassed.")

Following the English guidelines, the scope of a negated adjective in a noun phrase is the noun phrase itself as in (10). Also, the negation annotation does not scope over the coordinate adjective 废弃 as it is not affected by the negation.

(10)

那	人	住	在	这些	废弃	不用	的	小房	中	。
na4	ren2	zhu4	zai4	zhe4xie1	fei4qi4	bu4yong4	de	xiao3fang2	zhong1	。
that	person	lives	in	these	deserted	NOT-use	DE	little house	inside	.

("That person lives in these deserted and not-used little houses. ")

If the negated adjective is predicative, that is, when the adjective follows a linking verb such as 是 *shi4* ("is"), 感觉 *gan3jue2* ("feel") and 对待 *dui4dai4* ("treat"), the negation scopes over the entire clause such as in (11). As adjectives are verb-like in Chinese, the copula "to be" can be omitted, so (11) and (12) are both valid constructions in Chinese.

(11)

<u>这样</u>	<u>是</u>	<u>不</u>	<u>公正</u>	<u>的</u>	。
zhe4yang4	*shi4*	*bu4*	*gong1zheng4*	*de*	。
this	BE	NOT	fair	DE	.

("This is unfair.")

(12)

<u>这样</u>	<u>不</u>	<u>公正</u>	。
zhe4yang4	*bu4*	*gong1zheng4*	。
this	NOT	fair	.

("This is not fair.")

Besides adjectives, a relative clause can also be nominalised to modify a noun. Because the head noun is usually the complement of the predicate in the relative clause, the whole noun phrase including the head noun should fall in the scope when the verb in the relative clause is negated as in (13).

(13)

他	是	个	<u>不</u>	<u>爱</u>	<u>出风头</u>	<u>的</u>	<u>人</u>	。
ta1	*shi4*	*ge*	*bu4*	*ai4*	*chu1feng1tou*	*de*	*ren2*	。
he	is	CL	NOT	like	show-off	DE	person	.

("He is not a person who likes to show off.")

There is no negated subject nor negated object in Chinese and Chinese does not allow

negation within the noun phrase. The determiner "no" is translated as the verbal negation marker 无 *wu2* or 没有 *mei2you3* ("do not have/exist") in Chinese as in (14).

（14）

没	有	人	*注意到*	*它们*	。
mei2	*you3*	*ren2*	*zhu4yi4dao4*	*ta1men*	。
NOT	exist	person	notice	them	.

("There is not anyone who noticed them=No one noticed them.")

Finally, we decide to exclude sentence-final modal particles such as 的 *de*, 吗 *ma*, 呢 *ne*, 呀 *ya*, 哇 *wa* from scope annotation because they only express attitude or mood towards the whole sentence rather than affecting or being affected by the negation instance. An example is in (15).

(15)

不	要	等	他	过	了	山	呀	！
bu4	*yao4*	*deng3*	*ta1*	*guo4*	*le*	*shan1*	*ya*	*!*
NOT	shall	wait	he	past	PFV	mountain	MOD-P	!

("Don't wait until he is already past the mountain!")

3.3.3. Event

Following the English guidelines, an event is minimized to only the head of the negated phrase when it is factual. For example in (16), only the head of the negated predicate, 朋友 *peng2you* ("friend"), is marked as the event.

(16)

他的	朋友	也	不	是	我的	朋友	。
ta1de	peng2you	ye3	bu4	shi4	wo3de	peng2you	。
his	friends	also	NOT	BE	my	friends	.

("His friends are (also) not mine.")

The exception here is the set phrases in Chinese which are typically treated as multisyllabic multi-morpheme words. Since the English corpus does not annotate morphemes within a negated event, there is no reason to separate characters in these set phrases when annotating events in Chinese. Therefore, we decide to treat the set phrases as a whole if they are negated (see (17)).

(17)

我	并不	把	和	他	妹妹	的	感情	引以为耻	。
wo3	bing4 bu4	ba3	he3	ta1	mei4 mei4	de	gan3 qing2	yin3yi3wei2 chi3	。
I	NOT	BA	with	his	sister	GEN	feelings	ashamed	.

("I am not ashamed of my feelings towards his sister.")

Another special construction in Chinese is the verb-complement compound. When the compound is negated, we annotate the whole compound as the event if the complement or the verb is bound, or if the meaning of the whole can not be derived compositionally from the verb plus the complement, for example in 看不见 *kan4 bu4 jian4* ("cannot see-appear: cannot see") and 听不到 *ting1 bu4 dao4* ("cannot hear-arrive: cannot hear")。 Otherwise, the event is the verb such as in 说不清楚 *shuo1 bu4 qing1chu3* ("cannot speak clearly").

We follow the principle of translation and do not annotate Chinese coverbs such as 把 *ba3* (object marker) as events when they are negated, because they are never translated into any words or any annotated events in English. Instead, we will mark

the main verb that comes after the coverb as the event. An example is in (18)

(18)

我	父亲	根本	没	把	我	放	在	心	上	。
wo3	fu4 qin	gen1ben3	mei2	ba3	wo3	fang4	zai4	xin1	shang4	。
I	father	essentially	NOT	BA	me	put	at	heart	on	.

("My father essentially did not care about me.")

Following the English annotation guidelines, we do not annotate the copulative verb 是 *shi4* ("to be") and the existential verb 有 *you3* ("there is/exit") when they are negated. Instead, we mark the head of the following noun phrase as the event (see (16)). Notice that if 有 *you3* ("to have/exist") functions as a link verb that introduces a lower verb phrase, the event should be the verb in that verb phrase rather than the noun phrase following the link verb, if the cue precedes 有. This decision as shown in (19) is in correspondence to the translational equivalence principle as the English annotation of the same sentence will also mark the verb as the event as in "**No** <u>one</u> <u>lived</u> here.".

(19)

没	有	人	住	过	这	。
mei2	you3	ren2	zhu4	guo4	zhe4	。
NOT	exist	one	live	EXP	here	.

("No one lived here.")

We will also need to determine the factuality of the sentence before annotating the event. As an event should be a factually negated action or state, we should not annotate events in non-factual situations which are typically associated with modality, future, condition and supposition.

4. Methodology for the projection task

4.1. Sentence alignment

Before aligning the sentences between English and Chinese, we decided to move all the right-aligned inter-quotation attribution phrase in Chinese to the English convention which is left-aligned. That is, we changed the punctuation after the inter-quotation attribution phrase to sentence-final punctuation marks such as a period, and the punctuation before the attribution phrase to commas. For example, the original Chinese sentence in (20) will be changed into (21). (The sentence boundary is indicated by "][”). This step makes sure that Chinese and English sentences can be aligned for the projection task.

(20) ["就在这里好了。"][他说道，"右侧的这些山石是绝妙的屏障。"]
(21) ["就在这里好了，"他说道。]["右侧的这些山石是绝妙的屏障。"]
(" "This will do," said he. "These rocks upon the right make an admirable screen." ")

We used the sentence aligner hunalign (Varga et al., 2007) which was trained from both the HB corpus and an supplementary English-Chinese bilingual dictionary CEDICT (Denisowski, 1999). To increase the accuracy of the sentence alignment, we split the data by chapters before feeding it to hunalign, as the chapter division in two languages always corresponds. The alignment performance for the first chapter is 0.841 in precision and 0.847 in recall. To ensure the minimal effect from sentence alignment errors on the projection results, all the sentence alignment errors that involve negation on either side of the alignment were manually corrected.

The aligned sentence pairs in HB was joined with a larger parallel corpus (the training data used for the Edinburgh submission to NIST2012 English<-> Chinese translation task) to form the training data for the word alignment model. To roughly estimate the accuracy of the sentence alignment in the joined corpus, we randomly selected 100 sentence pairs from the data, and found 4 wrongly-aligned sentence pairs. The

accuracy is thus estimated to be around 96%. An additional tokenisation step was taken to change all the instances of the contracted cue "n 't" in the NIST2012 corpus into " n't" to ensure consistency with the HB corpus.

4.2. Word alignment

The word alignment model used in the projection task is the fast_align algorithm, an improved and faster implementation of IBM 2 (Dyer et al., 2013). We used the symmetrical two-way alignment results as the basis for projection.

The negation annotation on the English side was transferred through word alignment to the corresponding Chinese word(s). Since we were transferring the annotation in each sentence pair which may contain multiple natural sentences, the annotation from different sentences in a sentence pair were joined accordingly. If an English word is unaligned, the annotation associated with the English word is lost. If there is one-to-one mapping between a English word and a Chinese word, then the two words will have the same annotation. If an English word is aligned to multiple Chinese words, the annotation will be copied to all of the Chinese words. If multiple English words are aligned to one Chinese word, all the non-repetitive annotations of the English words will be transferred to the Chinese word. For example, if the multiword cue "by no means" is aligned to a single Chinese word 不, the negation annotation of the multiword cue will be transferred only once and 不 will have only one cue span. The outcome of the projection is annotated Chinese sentences in HB in the same CoNLL format.

5. Evaluation metrics for the projection task

Unlike Morante and Blanco (2012), the current study measures cue, event and scope independently, as they were projected independently. There are two types of evaluation metrics used in this study: strict comparison and soft comparison.

The strict comparison metric measures precision and recall on the span level for each sentence pair. Notice that the span could be cue, scope or event. A true positive (TP) requires that the projected span is identical to the span in the gold standard (Morante and Blanco, 2012: Farkas et al., 2010). The overall precision, recall and F1 scores are calculated from the accumulated True positives (TP), false negatives (FN) and false positives (FP) for each sentence pair.

The disadvantage of the strict comparison metric is that it penalizes partial spans more than complete misses. To supplement the strict comparison metric, a token/character-based soft comparison metric was designed. As there could be multiple spans in a sentence pair, we need to first match the spans between projection and gold standard for each sentence pair. Here we calculate F1 for every possible span combination and return the matches with the highest F1. TP, FN and FP are calculated for each determined span match, and are accumulated to calculate the overall precision, recall and F1 measure.

6. Results in the projection task

6.1. Word-based transfer of cues

6.1.1. Strict comparison

The results in Table 6.1 suggest that a large proportion of the errors are false negatives where the projection failed to predict the spans in the gold standard.

Table 6.1. Results from the strict comparison of cue spans

#correct spans (TP)	# spans not in the gold (FP)	# spans not in the projection (FN)	precision	recall	F1
327	573	882	0.363	0.270	0.310

Among the FP and FN, we would like to identify the cases where it is not possible to project the spans because of the language differences i.e. the linguistic errors. The

number of these linguistic errors could be approximated by comparing the number of spans across the projection, the gold standard and the original English annotation for each sentence pair. The results are summarized in Table 6.2.

Table 6.2. Further categories of the errors in strict comparison of cue spans

linguistic errors		alignment errors	
#FP spans	# FN spans	# lost FN spans	# mismatch FN/FP spans
80	335	54	493

Generally speaking, when the gold standard contains fewer spans than the projection, the additional cue spans in the projection are seen as errors from language differences (i.e. linguistic FP spans in table 6.2), as the original English text will not have fewer cue spans than the projection. For example, when a projected sentence contains 3 cue spans and the gold standard contains only 2 spans, the extra cue span in the projection will be unaligned in the gold standard because the original English annotation differs from Chinese by having this extra cue span.

When the number of cue spans in the gold standard is greater than the number of projected cue spans, we need to look at the original English annotation to decide how many spans are due to language differences and how many spans are lost through alignment. Here is an example: the original English annotation contains 3 cue spans which are projected as 2 spans whereas the gold standard contains 4 spans. We could infer from these numbers that there is one FN cue span due to lost in alignment (alignment errors: lost FN spans in Table 6.2), and one FN cue span due to language differences (linguistic FN spans in Table 6.2).

When there is an equal number of cue spans in the sentence pair and thus equal number of FP and FN, we could assume linguistic correspondence in cue annotation between English and Chinese. Although, such an assumption is not always true as the

two languages may coincidentally have the same number of cues with completely different negation events. The assumption is adopted here as a convenient alternative to manually going through errors in approximating the proportion of linguistic errors in the task. FP and FN in this case together with the remaining uncategorised errors in the previous two cases are considered as mismatch spans in the category of alignment errors as shown in Table 6.2.

The results from Table 6.2 show that the errors mainly come from linguistic FN and mismatch spans in alignment. The large number of mismatch alignment errors is within expectation as the strict comparison method counts partially correct spans as mismatches. The large number of linguistic FN suggests that Chinese employs more explicit negation expressions to translate positive or implicitly negative expressions in English. Below are some typical constructions of the Chinese FNs:

Juxtaposed positive + negative constructions:

For example, the English phrase "broken ground" is translated into 坎坷 *kan3ke3* ("broken") 不平 *bu4ping2* ("uneven") 的 *de* 土地 *tu3di4* ("ground")

Emphatic negation:

Sometimes a negation phrase is added to emphasize the previous positive word in Chinese. E.g. "small" is translated as 微不足道 *wei1bu4zu2dao4* ("so small that it is not worth talking about").

False negatives in English:

Chinese does not have equivalent fixed expressions for some of the false negatives in English. For example, "nothing but" is translated as 除了 *chu2le*...什么 也 *shen2me ye3*.., ("save...not").

Measuring distance, time and degree:

The distance and time adjectives and adverbs in English are usually translated into the

form of "cue + antonym" in Chinese. For example, "close" is translated into 不远 *bu4yuan3* ("not far"); "Soon" is translated into 不久 *bu4jiu3* ("not long").

Translating frequency adverbs with implicit negation

Adverbs such as "hardly"and "rarely" can only be translated into negation in Chinese. For example, the English phrase "I hardly expect" is translated as 我 想不到 *wo3 xiang3budao4* ("I did not expect").

6.1.2. Soft comparison

The results from the soft comparison metric are summarized in Table 6.3. Compared with the performance under the strict comparison metric, the F1 measure here is much higher. Also, different from the strict comparison metric, the projected tokens in total outnumber the tokens in the gold standard as there are more FP tokens than FN tokens. To further understand the exact causes of these errors, Table 6.4. groups the errors into fine-grained categories.

Table 6.3. Results from the soft comparison of cue spans

#correct tokens	#FP tokens	# FN tokens	precision	recall	F1
962	1032	702	0.482	0.578	0.526

Table 6.4. Further categories of the errors in soft comparison of cue spans

linguistic errors		alignment errors				
# linguistic FP tokens	# linguistic FN tokens	# lost FN tokens	#FP tokens in partially correct spans	#FN tokens in partially correct spans	#FP tokens in fully missed spans	#FN tokens in fully missed spans
170	382	91	649	63	213	166

The first two columns in Table 6.4 are linguistic errors that come from FP spans and

FN spans. The third column shows FN tokens due to lost in alignment. The fourth and the fifth columns show the number of FP and FN in span pairs where there is at least one correct token. The last two columns show the number of FP and FN in fully missed spans.

The distribution of linguistic FP and FN conforms with the observation from the strict comparison results. There are more linguistic FN spans and thus more linguistic FN tokens. As to the alignment errors, the projection seems to have always over-projected rather than under-projected as there are consistently more FP tokens than FN tokens in both partially correct spans and fully missed spans. The majority of alignment FP occurred when the model has correctly identified part or even all of the gold cue span, but projected more than the gold. This is not surprising as a segmented word would be transferred as a whole even only part of the word is the cue. For example 彻夜不眠 *che4ye4bu4mian2* ("all night no sleep") was transferred all together as the cue, resulting in three extra FP characters. Interestingly, on the other hand, whenever part of the cue was detected, the projection model was able to preserve the rest of the span in most cases as there are only a small number of FN in partially correct spans, suggesting that the word boundaries imposed by the word segmentation algorithm were perhaps useful in preserving the components of the multiword cues.

To examine the cause of the large number of alignment FP tokens, we list the five most frequent FP alignment errors in Table 6.5.

Table 6.5. The 5 most frequent FP alignment errors in soft comparison of cue spans

FP tokens	frequency
有	73
"	67
,	59
么	39
什	39
...	...
total	862

The distribution of the alignment FP is skewed with a long list of tokens that occur only once or twice, suggesting that the model tends to align the English negation cues to a wide range of characters. The main error source is punctuation which is usually surrounding the negation cue. The FP 有 *you3* ("to exit/to have') was typically found in the falsely segmented Verb-Object phrase 没 *mei2* 有 *you3* ("not have/exist") which should be divided into two characters but are sometimes combined together by the segmenter and thus transferred as a whole phrase. The FP tokens 什 *shen2* and 么 *me* come from the translation of the English negation pronoun "nothing" or "none"[3] as in (20).

(20)

什么	也没	有	。
shen2me	*ye3mei2*	*you3*	。
anything	NOT at all	exist	.

("None.")

---

[3] As 什么 *shen2me* ("anything") is neither a bound morpheme nor a construction restricted to negation, we do not include 什么 into the multiword cue.

29

There are also a large number of grammatical and modal particles such as 的 *de* (possession and modification marker) which do not correspond to any particular English word. These particles usually end up with unreliably low word alignment probabilities and might be randomly aligned to English cues.

6.2. Word-based transfer of scopes

6.2.1. Strict comparison

The results from the strict comparison of scopes are presented in Table 6.6. The low F1 here is within expectation as mapping the exact scopes is more difficult than mapping cues and events as a scope usually contains more tokens. We would like to set aside the errors caused by an unequal number of cue spans which are already addressed in the previous section, and focus on the errors specific to scope projection.

Table 6.6. Results from the strict comparison of scope spans

# correct spans (TP)	# spans not in the gold (FP)	# spans not in the projection (FN)	precision	recall	F1
19	864	1142	0.022	0.016	0.019

As shown in table 6.7, apart from linguistic FP and FN cue spans, most of the errors are due to mismatches in alignment rather than due to different number of scopes in the two languages. The cases of linguistic FN scope spans are however worth mentioning here. These errors occur when the English cue span does not specify the scope but the Chinese translation fills in the elided scope for the cue. For example, during a conversation, a speaker replies to a question "What did you do?" with simply "Nothing". The Chinese translation would add the scope by: 没 干 什么 *mei2 gan4 shen2me* ("did not do anything.").

Table 6.7. Further categories of the errors in strict comparison of scope spans

linguistic errors			alignment errors	
# FP cue spans	#FN cue spans	#FN scope spans	# lost FN spans	# mismatch FN/FP spans
97	335	39	1	767

6.2.2. Soft comparison

Table 6.8 and Table 6.9 below show the results from the soft comparison of scope spans and the further categories of the errors.

Table 6.8. Results from the soft comparison of scope spans

#correct tokens	#FP tokens	# FN tokens	precision	recall	F1
6217	3036	6693	0.672	0.482	0.561

Table 6.9 Further categories of errors in the soft comparison of scope spans

linguistic errors		alignment errors				
# linguistic FP tokens	# linguistic FN tokens	# lost FN tokens	# FP tokens in partially correct spans	# FN tokens in partially correct spans	# FP tokens in fully missed spans	# FN tokens in fully missed spans
1021	3158	105	1942	3308	73	122

Again, the F1 value under soft comparison is much better than the F1 under strict comparison. Besides the linguistic errors caused by the unequal number of cue spans, the token-based errors mainly come from FN tokens especially in partially correct spans. It seems that the projection failed to project a large number of characters in Chinese scopes. To further examine the typical errors, we list in Table 6.10 the top 5 most frequent alignment FP and FN tokens in scope projection.

Table 6.10. The 5 most frequent FP/FN alignment scope errors under soft comparison

FP tokens	frequency	FN tokens	frequency
，	194	的	335
不	170	我	111
没	82	有	84
。	67	了	81
有	66	这	71
...	...	..	...
total	2015	total	3535

Similar to the alignment FP cue tokens, the major error source of alignment scope FP is punctuation as shown in table 6.10. An explanation for the presence of negation cues 不 *bu4* and 没 *mei2* among the top FP alignment errors is that these cues have been aligned to scope tokens, typically the auxiliary verbs, on the English side. This is possibly due to the fact that there is no fixed translation in Chinese for the English auxiliary verbs ("does", "have", etc.) which are sometimes randomly aligned to Chinese cues.

As to the FN tokens, modal particles such as 的 *de* and 了 *le* are unique to Chinese with no equivalent translation in English. It is expected that these particles are unaligned. The 1ˢᵗ person singular pronoun 我 *wo3* is also a frequent FN token when it is annotated as an elliptical element. As a pro-drop language, Chinese is more tolerant of long-distance ellipsis. The recoverable elliptical subject element 我 for example is usually far away from the cue in Chinese as in (21). As fast_align favors diagonal alignment which assumes the same word order in target and source languages, the long-distance elliptical element is not likely to be aligned to the corresponding English pronoun which is usually found near the cue.

(21)

我	一再地	盘问	她	，	可是	再	问不出	什么	了	。
wo3	yi1zai4 de	pan2wen4	ta1	，	ke3 shi4	zai4	wen4bu4 chu1	shen2me	le	。
I	again	cross-question	her	，	but	again	able to ask	anything	PFV	.

("Again and again I cross-questioned her, but I could never get passed that point.")

6.3. Word-based transfer of events

6.3.1. Strict comparison

The results from the strict comparison of event projection are summarized in Table 6.11. As expected, there are a large number of FN spans due to language differences. As before, we would like to set aside the effect of the unequal number of cue spans, and focus on the specific linguistic errors associated with events. Table 6.12 shows the further categories of the errors. To be specific, we would like to identify cases where a non-factual event in English is translated into a factual event in Chinese and vice versa. The number of the former cases (linguistic FP event spans in Table 6.12) is approximated by counting the span pairs where the projected span contains a cue but no event whereas the gold span contains both a cue and an event. By contrast, the linguistic FN event spans in Table 6.12 refer to the cases where the projected span contains a cue and an event but the event is not annotated in Chinese.

Table 6.11. Results from the strict comparison of event spans

#correct spans (TP)	# spans not in the gold (FP)	# spans not in the projection (FN)	precision	recall	F1
130	399	625	0.0246	0.172	0.202

Table 6.12. Further categories of the errors in strict comparison of event spans

linguistic cue errors		linguistic event errors		alignment errors	
#FP cue spans	#FN cue spans	#FP event spans	#FN event spans	#lost FN spans	#mismatch FN/FP spans
51	223	67	61	60	281

Similar to the results from cue comparison, the error analysis in Table 6.12 suggests that the main errors are mismatch spans and linguistic FN spans caused by cues. The FN and FP event spans will also be discussed in detail when we apply the token-based soft comparison metric in the next section.

6.3.2. Soft comparison

As shown in Table 6.13, the event projection performance under soft comparison yields higher F1 compared with the strict comparison metric, but remains relatively low compared with cue and scope projection. To further examine the errors especially in terms of the cause of the large number of FN tokens, we categorize the errors into linguistic errors as shown in Table 6.14 and alignment errors as shown in Table 6.15.

Table 6.13. Results from the soft comparison of event spans

#correct tokens	#FP tokens	# FN tokens	precision	recall	F1
405	690	789	0.370	0.340	0.354

Table 6.14. Linguistic errors in soft comparison of event spans

linguistic cue errors		linguistic event errors	
# FP tokens	# FN tokens	# FP tokens	# FN tokens
104	311	142	195[4]

---

[4] 195 includes the number of linguistic FN tokens and the lost FN tokens

Table 6.15. Alignment errors in soft comparison of event spans

# lost FN tokens	#FP tokens in partially correct spans	#FN tokens in partially correct spans	#FP tokens in fully missed spans	#FN tokens in fully missed spans
195	177	30	267	250

As shown from the above two tables, apart from the linguistic cue errors, the majority of errors are alignment errors. For one thing, the particularly large number of error tokens in fully missed spans might suggest poor performance of the word alignment model. Similar to the results form cue comparison, the relatively small number of FN tokens in partially correct spans might suggest that the segmented word boundaries were useful.

As to the linguistic errors specific to event projection, unfortunately we cannot distinguish between lost FN tokens and FN tokens that are translated from non-factual to factual events as we do not have access to word alignment gold standard. A general observation from the linguistic event FN tokens is that they occurred when a non-factual modal verb was translated into a lexical verb. For example the English "would" will sometimes translate into 愿意 *yuan4yi4* ("want"). By contrast, linguistic FP event tokens occurred when a lexical or a modal verb expressing "ability" was translated into a non-factual modal verb. Table 6.16 shows a list of 5 most frequent linguistic FP event tokens.

Table 6.16. The 5 most frequent linguistic FP event errors

FP tokens	frequency
能	10
可	9
不	6
会	5
想	5
...	...
total	142

The top two tokens in Table 6.16 come from the Chinese modal verb, 可能 *ke3neng2,* which is always translated into lexical words in English, e.g. "possible" and "possibly". The same is with the Chinese modal verb 会 *hui4* which is sometimes translated as "likely". The FP 不 *bu4* is the by-product of transferring the whole segmented words that contain affixal negation.

The FP 想 *xiang3* ("think") reveals another linguistic difference in event annotation. The two languages use different representations in neg-raising constructions. For example, "I do **n't** *think* it is likely.." has a factual event "think", but the Chinese translation moved the negation cue to negate the verb in the object clause as in (22). Because the object clause was introduced by a supposition verb "think", the event in the object clause was not annotated.

(22)

我	想	不	会	。
wo3	*xiang3*	*bu4*	*hui4*	。
I	think	NOT	likely	。

("I think it is not likely.")

6.4. An additional experiment using character-based alignment

As shown from the error analysis, many of the FP tokens are due to the imposed word boundaries by the segmenter. Also, affixal negation in segmented words is not accessible to the model, resulting in many misaligned and unaligned tokens in the gold standard. On the other hand, word segmentation might have reduced the number of FN especially in partially correct spans. To explore the effect of word segmentation, we conducted another projection experiment using character-based alignment (ie. the input to fast_align is split characters rather than segmented words). The results show that the character-based alignment performed consistently worse than word-based alignment except in the strict comparison of cue spans where it achieved an F measure of 0.339 which is slightly better than the word-based alignment result (F1=0.310). The improvement is mainly because of the decrease in lost FN cue spans. These cue spans are typically affixal negation markers in segmented words and were not correctly aligned to negation cues in English under the word-based alignment. Once the characters were split, these affixal negation cues in Chinese became available to the model and thus were more likely to be aligned to cues in English.

Nevertheless, the general implication from the experiments is that the benefit of having word boundary information from a word segmenter seems to have outweighed its detriment in negation projection.

7.   Implications and future research

To sum up, both the projection performance and the annotation process seem to suggest that there is sometimes no linguistic correspondence in negation expressions between English and Chinese.

Indeed, a simple word alignment model can only correctly project about half of the negation annotation from English to Chinese. A large proportion of the projection errors is due to language differences. The most typical language error throughout the

experiments comes from the fact that the two languages make different choices on when to use explicit and implicit negation. In particular, Chinese exhibits considerably more explicit negation, resulting in a vast number of linguistic FN cue spans. This is in line with the observations made in the annotation guidelines as Chinese has more active negation affixes than English. Another important error source is the inconsistency in the English annotation. For example, the suffix "less" in the word "helpless" was annotated 2 times as a cue but was not annotated for the other 3 cases. We should also be aware that the genre of the corpus is literature (narratives) where translators have more freedom to make non-literal lexical choices compared with genres such as news and scientific writings. It is possible that the linguistic differences we have seen in this study are partly caused by translation style.

As to the alignment errors, the cue and event errors are mainly FP. The model tends to overpredict tokens with low translation probabilities including particles and punctuation marks. For scope comparison, the major alignment errors are FN where the model missed a large number of modal particles and long-distance elliptical elements.

Future research on projecting and translating negation should pay close attention to the use of implicit and explicit negation devices between a language pair. It would also be interesting to compare different versions of the translation to distinguish language differences from the effect of translation variety. It is also advisable to create an annotation scheme that deals with implicit negation in the future. Consistency issues in annotation should also be dealt with for example by employing automatic quality check on the manual annotation. In addition, as the current annotation guidelines were designed by following the principles of linguistic equivalence and translational equivalence to the English annotation, future research could revise the guidelines to follow more faithfully to the linguistic properties in Chinese.

As the projection algorithm used in the current experiments was only a baseline word

alignment model, there are also several directions that future research could take to improve the algorithm. For example: to reduce FP, we could filter out word alignments with low translation probabilities. To reduce FN in scope comparison, we could incorporate syntactic information by mapping constituents rather than words. However, as languages can differ considerably in the way they express negation, it is most likely that we would still need linguistic analysis and annotation from human experts in the target language.

8. Conclusion

The annotation and the projection tasks in this study reveal the different ways English and Chinese express negation. The linguistic differences of the two languages also affect the word alignment probabilities learned from the parallel corpus as Chinese has a large number of untranslated and thus unaligned particles. The situation is complicated by the usually non-literal translation in the genre of narratives.

Since the projection results are not reliable, the manual creation of additional language resources in negation annotation becomes crucial, and the current HB-neg-ch corpus is no doubt an important contribution to cross-lingual natural language understanding of negation, and will benefit a range of NLP tasks including machine translation and information extraction and many more.

Appendix A. Guidelines for negation annotation in Chinese

The current guideline for HB-neg-ch has been created following the order of the English guideline (Morante et al., 2011). When a rule from English can be applied to the Chinese corpus, a Chinese example will be given. When a Chinese construction is not covered by the English guideline, a new rule will be designed. Any errors or anything we do not fully agree with the English guideline will be discussed. Most of the examples listed in this guideline come from the Chinese translation of the Hound of Baskervilles. Whenever it is the case, chapter and sentence numbers of the source sentence are included in the brackets. For example [1.2] refers to the second sentence in the first chapter.

1. cue

As defined in Morante et al. (2011), a cue is "the word that expresses negation". There are in total 45 types of cues and 10 types of core negation cues in the corpus. The core negation cues here refer to the single-morpheme negation markers in Chinese. They are typically one-character word (except for 没有 *mei2you3* "did/have not") that can be combined with bound morphemes to form multiword cues. The core negation cues can be adverbs, auxiliary verbs and prefixes. Unlike English, there are no negation pronouns (e.g. "nothing") and negation prepositions ("without") in Chinese. The following is a list of Chinese core negation cues with examples.

1.1. Core negation cues in Chinese:

1.1.1. 不 *bu*4 ("not"):

不 is the most common cue in Chinese. It is a negation adverb that is able to negate all verbs and adjectives but 有 *you3* ("to have/to exist"). It always comes before the verb/adjective (Chao, 1965: 748). An example is in (23). Notice that for all the annotation examples displayed in this guideline, the cue is marked in bold; the scope is marked with an underline and the event is marked in italics.

(23)

我	不	知道	应该	相信	什么	。
wo3	*bu4*	*zhi1dao4*	*ying1gai1*	*xiang1xin4*	*shen2me*	。
I	NOT	know	should	believe	what	.

("I don't know what I should believe.") [3.72]

1.1.2. 没(有) *Mei2 (you3)* ("did not/have not"):

没(有) *Mei2 (you3)* is a negated auxiliary verb that indicates non-completion (Li & Thompson, 1989: 417). Sometimes the character 有 can be omitted. An example is shown in (24):

(24)

我	没(有)	看见	他	。
wo3	*mei2(you3)*	*kan4jian4*	*ta1*	。
I	did/have NOT	see	him	.

("I did not see him.")

1.1.3. 没 *mei2* ("not")

没 *mei2* ("not") is also used specifically to negate the verb 有 *you3* ("to have/to exist"), meaning "do not have" or "there is not". See (). Sometimes 有 *you3* ("to have/to exist") can be omitted (Li & Thompson, 1989: 416).

(25)

我	没	(有)	理由	。
wo3	*mei2*	*(you3)*	*li3you2*	。
I	NOT	have	reason	.

("I have no reasons.") [11.69]

1.1.4. 未 *wei4* ("did/have not")

未 is the classical form of 没有(1.1.2) with the meaning of "have not". An example is shown in (26)

(26)

一	件	尚	**未**	*得到*	*解释*	<u>的</u>	<u>事实</u>	。
yi1	*jian4*	*shang4*	*wei4*	*de2dao4*	*jie3shi4*	*de*	*shi4shi2*	。
a	CL	yet	have NOT	get	explanation	DE	fact	.

("A fact that has not been explained yet.") [2.96]

1.1.5. 别 *bie2* ("do not")

别 is the negative marker specifically used to form a imperative construction as in (27)

(27)

别	<u>这样</u>	,	华生	。
bie2	*zhe4yang4*	,	*hua2sheng1*	.
NOT	this	,	Watson	.

("Don't be like this, Waston") [4.180]

1.1.6. 莫 *mo4* ("do not/not")

莫 is a classical form of 不 in the set phrase (see (28)) and also can be an imperative negation marker.

(28)

爱	**莫**	*能*	*助*
ai4	*mo4*	*neng2*	*zhu4*
love	not	can	help

("Unable to help though would like to") [5.131]

1.1.7. 勿 *wu4* ("do not")

勿 is a formal form of the imperative negation marker 别 as in (29).

(29)

万	**勿**	将	此	事	告知	其	姊	伊丽莎白	。
wan4	*wu4*	*jiang1*	*ci3*	*shi4*	*gao4 zhi1*	*qi2*	*zi3*	*yi1li4sha 1bai2*	。
make sure	NOT	JIANG	this	matter	tell	your	sister	Elizabeth	.

("Do not tell this matter to your sister Elizabeth.")[2.64]

1.1.8. 非 *fei1* ("is not")

非 is the classical form of 不是 *bu4shi4* ("is not"). 非 is commonly used with the negation adverb 并 *bing4* to negate a whole clause, such as in (30).

(30)

并	非	全	信	都	*能*	读	。
bing4	*fei1*	*quan2*	*xin4*	*dou1*	*neng2*	*du2*	。
BING	NOT	all	letter	all	can	read	.

("It is not the case that the whole letter can be read.") [11.95]

1.1.9. 无 *wu2* ("do not have")

无 is the classical form of 没有 (1.1.3) as in (31). In modern Chinese, 无 is also used as a negation prefix equivalent to "-less". e.g. 无线 *wu2xian4* ("no wire=wireless") (Chao, 1965: 214).

(31)

咱们	无	法	把	他	一直	抬	到	庄园	去	。
zai2men	*wu2*	*fa3*	*ba3*	*ta1*	*yi1zhi2*	*tai2*	*dao4*	*zhuang 1yuan2*	*qu4*	。
We	NOT have	means	BA	him	all the way	carry	to	hall	go	.

("We can't carry him all the way to the Hall.") [12.197]

1.1.10. 否 *fou3* ("not")

否 is a negation prefix in verbs. An example is the verb 否认 *fou3ren4* ("not-acknowledge = deny")

1.1.11. Verbs with implicit negation

Verbs such as "refuse"and "fail" implicitly express negation and were sometimes treated as negation cues in the English guideline. The current annotation will not mark these verbs as cues because (1) these verbs were annotated inconsistently in the English annotation. In fact, only one instance of "fail" was annotated as a cue, and (2) there are no unambiguous criteria of implicit negation.

Notice that the classical forms of negation cues are addressed in this guideline as they are still productive in modern Chinese especially in creating set phrases.

1.2. Affixal negation in Chinese compounds

As an isolating language, Chinese has rare affixes. The definition of an affix is also problematic in Chinese as some common Chinese affixes including negative affixes that occur in fixed locations within a compound can also function as a stand-alone free morphemes in other cases. Here we employ a rather broad definition of negation affixes that refer to local negation cues within a compound, and a broad definition of a compound which consists of multiple compositional characters.

1.2.1. Infix 不 in verb-complement compound

The negative infix 不, which is inserted between the verb and its complement in the verb-complement compound, is the only acknowledged infix in Chinese. The types of complements that can be negated by the infix are result complement, directional complement and potential complement. The infixal negation in verb-complement compound can be seen as negation of the potential form introduced by 得 *de2* ("can") as in (32). A more intuitive interpretation of the infixal negation in Chinese is that the result introduced in the complement is 'unachievable' (Li & Thompson, 1989: 57).

(32)

他	说	得	清楚	。
ta1	*shuo1*	*de*	*qing1chu3*	。
He	say	can	clearly	.

("He can speak clearly")

他	说	不	*清楚*	。
ta1	*shuo1*	*de*	*qing1chu3*	。
He	say	not	clearly	.

("He cannot speak clearly")

1.2.2. prefix Negation cues in compounds

Except 别 *bie2* ("not") and 勿 *wu4* ("not"), all the other core negation cues can function as a prefix in a compound such as 没/无用 *mei2/wu2yong4* ("useless"). Interestingly, some of them are modern prefixes that are created through translation: e.g. 无线 *wu2xian4* ("no wire=wireless"), 非物质 *fei1wu4zhi4* ("not physical=non-physical"), 不道德 *bu4dao4de2* ("immoral").

1.2.3. False affixal negation

Like English, some negative affixes in Chinese do not genuinely introduce negative polarity. However, as Chinese will allow speakers to read meaning from every character (Chao, 1965), the negative meaning in the affixal negation cues in Chinese

compounds is much more active than the negative affixes in English. There are cases such as "disgrace" in English where the negation meaning of "dis" only exists etymologically. By contrast, Chinese characters are more able to preserve their original meaning as they have not changed much historically. Besides a small number of exceptions, Chinese affixal negation cues are always still expressing negation. Therefore, the number of false negatives in Chinese is considerably smaller than in English. The following is a list of these exception cases where we decide not to annotate the negation affix:

Compositionality

Following the English guideline, the meaning of the lexical items in the compound is taken into account to determine whether the negative affix is really a negation cue. When the meaning of the compound cannot be compositionally derived from the negative affix and the rest, we should not mark the cue. For example, 无聊 *wu2liao2* ("boring") in Chinese does not mean "*no* chatting". Also, the Chinese translation of the English example "disappear", 不见 *bu2jian4,* does not mean "not appear". The negation characters in both words should not be annotated as cues.

Obsolescence

When the meaning of the morpheme modified by the negative affix is obsolete, the affix should not be annotated as the negation cue. For example, 然 *ran2* in 不然 *bu4ran2* ("otherwise") and 则 *ze2* in 否则 *fou3ze2* ("otherwise") used to mean "like this" in classical Chinese, but this meaning is not salient any more in Modern Chinese. Therefore, a native speaker's intuition will treat the compound as a whole rather than mark 不 and 否 here as negation cues.

Translation equivalence

Finally, if the negation prefix in a compound is never translated to explicit negation in English, it should not be annotated as a cue. For example, words that introduce

contrary meaning to the negated morpheme (Shou-Hsin, 1978) such as 不少("not few-many"), 无二 ("no another=only one") and words that introduce emphatic degree modification such as 无比 ("not compare - very"), 不已("not stopped - very much") all contain false negative affixes.

1.2.4. Words with multiple meanings

Notice that words with multiple meanings will be annotated differently. For example 不 *bu4* in 不过 *bu4guo4* which has the literal meaning of "not exceeding" (不 超过 *bu4 chao1guo4*) as in (33) should be annotated as the cue. In (34), 不过 means "but", which is completely irrelevant with the literal meaning of "not exceeding". Therefore 不 should be treated as a false negative affix in (34). However, there are intermediate cases that are trickier to decide. In (35), the translation to 不过 is usually "only" and we will not get the same meaning if using the full form of the literal 不过,不 超过 *bu4 chao1guo4* ("not exceeding"). Therefore, we will not annotate 不 in (35) as the cue.

(33)

不	过	一	英里	远	。
bu4	*guo4*	*yi1*	*ying1li3*	*yuan3*	。
NOT	more/exceed	one	mile	away	.

("It's not more than one mile away"). [9.241]

(34)

不过	我	没	去	。
bu4guo4	*wo3*	*mei2*	*qu4*	。
but	I	didn't	go	.

("But I didn't go.")

(35)

这	不过	是	我	的	猜想	。
zhe4	*bu4guo4*	*shi4*	*wo3*	*de*	*cai1xiang3*	。
This	only	is	my	GEN	conjecture	.

("This is only my conjecture.")

1.3. multiword negation cues

1.3.1. Adverb+core cue

Constructions such as 并不 *bing4bu4* ("not"), 绝不 *jue2bu4* ("absolutely not"), 决不 *jue2bu4* ("absolutely not") and 绝无 *jue2wu2* ("absolutely no") should be treated as multiword cues. They are typically a compound that consists of a core negation cue modified by a bound morpheme or a construction that only occurs in negation. The bound morpheme typically expresses a certain degree of emphasis on the negation. Such constructions are comparable with the English multiword cue "by no means". According to Chao (1968: 143)'s definition, a bound morpheme in Chinese "can not be uttered alone and always occurs without pause with another morpheme". 并,绝,决 from the previous examples are all bound morphemes while 并 is a more restricted morpheme as it can only occur in a negation construction. Other examples of negation-restricted multiword cues are ...也 *ye3*+negation marker ("...not at all") and 再也 *zai4ye3*+ negation marker ("no longer/never"). (36) and (37) show that 再也 can not occur in a positive construction[5], and therefore should be treated as part of the negation cue.

---

[5] Ungrammatical constructions are marked with an asterisk

(36)

他	再也没有	回来	。
ta1	*zai4ye3mei2you3*	*hui2lai2*	。
He	never	return	.

("He never returned.") [2.92]

*他	再也	回来	。
ta1	*zai4ye3*	*hui2lai2*	。
He	again	return	.

Notice that we are excluding cases such as 可 *ke3* ("but" or emphasis particle)不 *bu4* ("not"),也 *ye3* ("also")不 *bu4* ("not") as multiword cues. Although 可 and 也 are bound morphemes, they do not form a compound with the cue and they modify the scope introduced by the cue. Instead, we will only annotate 不 *bu4* as the cue in these cases.

1.3.2. False affixal negative cue

Multiword cues can also include bound morphemes that are within the scope of the core negation cue and are restricted to negation constructions. These constructions contain false local affixal negation and negate other parts of the sentence. Examples are 不见得 *bu4jian4de2* ("probably not"), 不用 *bu4yong4* ("need not /without"), 不及 *bu4ji2* ("not reaching"), 不屑 *bu4xie4* ("do not deign to"), 说不上 *shuo1bu4shang4* ("can't say"), 不怎么 *bu4zen3me* ("not really"). A bound morpheme is restricted to negation constructions if the bound morpheme with its specific sense and part of speech cannot be bound with other morphemes to form a similar positive construction. (37) shows 不用 *bu4yong4* ("not need") is a multiword cue as 用 *yong4* ("need") is restricted to negation.

(37).

您	不用	怕	。
nin2	*bu4yong4*	*pa4*	。
you	NOT-need	fear	.

("You need not fear.") [9.287]

*您	用	怕	。
nin2	*yong4*	*pa4*	。
you	need	fear	.

A seemingly similar construction 不便 *bu4bian4* ("not convenient") is analyzed in (38). We should treat 不 *bu4* ("not") and 便 *bian4* ("convenient") separately even when the morpheme 便 is bound, because it is common to keep only the bound morpheme rather than the whole word in constructing a negated expression. The phrase 不便 can be seen as a shorter expression of 不方便 *bu4 fang1bian4* ("not convenient") with no different meanings. Since 方便 *fang1bian4* ("convenient") can occur in a positive construction, we should not treat 不便 as a multiword cue because we cannot ignore the negated event 便 *bian4* ("convenient").

(38)

我	不	便	回答	。
wo3	*bu4*	*bian4*	*hui2da2*	。
I	NOT	convenient	answer	

("It's not convenient for me to answer.") [10.164]

我	方便	回答	。
wo3	*fang1bian4*	*hui2da2*	。
I	convenient	answer	

("It's convenient for me to answer.")

Notice that the positive equivalent of the form 无 *wu2* X ("do not have X") is 有

you3 X ("have X"), and the positive equivalent of the form 非 *fei1* X ("is not X") is 是 *shi4* X ("is X"). Therefore 无法 *wu2fa3* ("have no means") is not a multiword cue, and should be annotated as 无(cue)法(event) because the positive form 有(办)法 *you3 (ban4)fa3* ("have means") is a valid construction in Chinese.

1.3.3.

Another case worth mentioning is the multifunctional 没有 *mei2you3*. When 有 in 没有 means "have" or "exist" as in (39) and (40), 没 should be seen as the negation cue which negates 有, because the verb 有 ("have/exist") can occur in the positive constructions. When 没有 *mei2you3* ("have not/did not") precedes a verb as an auxiliary negation marker as in (41), 有 is considered as part of the cue as it cannot occur in the equivalent positive construction. Therefore，没有 in (41) is a multiword cue.

(39)

我	没	有	理由	。
wo3	*mei2*	*you3*	*li3you2*	。
I	NOT	have	reason	.

("I do not have reasons.") [11.69]

我	有	理由	。
wo3	*you3*	*li3you2*	。
I	have	reason	.

("I have reasons.")

(40)

没	有	人	看到	他	。
mei2	*you3*	*ren2*	*kan4dao4*	*ta1*	。
NOT	exist	person	see	him	.

("No one saw him.")

有	人	看到	他	。
you3	*ren2*	*kan4dao4*	*ta1*	。
exist	person	see	him	.

("Someone saw him.")

(41)

我	没有	*看见*	他	。
wo3	*mei2you3*	*kan4jian4*	*ta1*	。
I	did/have NOT	see	him	.

("I did not see him.")

*我	有	看见	他	。
wo3	*you3*	*kan4jian4*	*ta1*	。
I	-	see	him	.

One thing to be noted that the positive construction in (41) is becoming more acceptable in Mandarin Chinese due to the influence of Cantonese and Taiwanese. 有 here sometimes introduces emphasis on the past event, which is similar to the effect of "did" in English (Chao, 1968:20).

1.3.4.

There are also discontinuous multiword cues in Chinese. An equivalent Chinese cue to English "neither...nor" is 既 *ji4* +core cue... 也 ye3+core cue. The core negation cues in this construction can be any of those listed in 1.1, but has to be the same cue. For example, 既不 *ji4bu4*... 也不 *ye3bu4*. 既 ... 也 ("either...or...") are sometimes omitted in the set phrases: For example: 不骄不躁 *bu4jiao1bu2zao4* ("neither proud nor upset").

1.3.5. A note on defining multiword cues

A multiword cue should have all the properties of a cue. In particular, it should negate some part of the sentence other than the cue itself. Otherwise, it might be just a

compound with a false negative affix. However, to distinguish whether the multiword expression introduces negation is not easy especially when the bound morpheme in the expression introduces modality. A general rule of thumb is to ask whether the modality is attached to the negation of the event or the event itself. For example in (42) and (43), both 不见得 *bu4jian4de2* ("not necessarily") and 说不定 *shuo1bu4ding4* ("not able to say definitely-maybe") express possibility, but 不见得 suggests the possibility of the negated event "not trust" and is semantically similar to a negated modal verb, whereas 说不定 expresses the speaker's positive assumption of the "trust" event. Therefore, we should annotate 不见得 as a cue but not 说不定.

(42)

他们	不见得	信任	我们	。
ta1men	*bu4jian4de2*	*xin4ren4*	*wo3men*	。
they	not necessarily	trust	us	

("They probably don't trust us.") [9.246]

(43)

他们	说不定	信任	我们	。
ta1men	*shuo1bu4ding4*	*xin4ren4*	*wo3men*	。
they	maybe	trust	us	.

("They maybe trust us.")

1.4. False negatives

1.4.1. Double negatives as false cues

The double negatives around the modal verb, for example in 不能不 *bu4neng2bu4,* 不得不 *bu4de2bu4,* usually introduce a reinforced positive meaning of "cannot help doing" or "must" (Chao, 1965:731). These modal verb double negatives are also never translated into double negatives in English.

我	不得不	放弃	这	种	方法	。
wo3	*bu4de2bu4*	*fang4qi4*	*zhe4*	*zhong3*	*fang1fa3*	。
I	not-should-not	abandon	this	CL	method	.

("I am compelled to abandon this method"). [10.1]

Another fixed expression of double negatives is 非得 *fei1dei3*...不可 *bu4ke3* which means "must". Double negatives can also introduce emphasis on the positive meaning in a set phrase. Unlike the discontinuous multiword cues, the false double negatives in the set phrase usually consist of different negation markers. For example: 无往不利 *wu2wang3bu4li4* ("there are no places where victory is not achieved = always successful").

1.4.2. Local false affixal negation cues

See 1.2.3

1.4.3. Fixed expressions

Like English, Chinese has certain fixed expressions or constructions that involve the use of negation markers but do not introduce or translate into negation. For example, 不("not")是("is")... 就("then")是("is") is always translated as "either...or". Another similar construction is 不是别人("not others), 正是 ("is exactly")... which is usually translated as the English false negative "none other than...".

1.4.4. Dialogue checker

Like English, the negation markers in some rhetorical questions serve to confirm rather than to negate. Here are some typical constructions: 你不觉得 *ni3bu4jue2de2*..? (Don't you think...?), 这不就是? *zhe4 bu jiu4 shi4*/这不正是 *zhe4 bu zheng4 shi4*? (Isn't it exactly...).

There are also several adverbial dialogue checkers with the meaning of "no doubt",

for example, 毫无疑问 *hao2wu2yi2wen4* ("without a doubt"), 无疑 *wu2yi2* ("no doubt"). The negation markers in these cases are not annotated as the cue.

1.4.5. Modality

Chinese could have the construction "V+negation marker+V ?" (see (44)) or add the negation marker at the sentence-final position as in (45) to form a a yes-no question. The negation markers in these constructions introduce modality rather than genuine negation, and should not be annotated as cues.

(44)

咱们	要	不	要	向	后	退	?
zai2men	*yao4*	*bu4*	*yao4*	*xiang4*	*hou4*	*tui4*	*?*
we	shall	NOT	shall	towards	back	retreat	?

("Shall we move farther back?") [14.72]

(45)

你	找到	了	没有	?
ni3	*zhao3dao4*	*le*	*mei2you3*	*?*
You	found	PFV	Q-P	?

("Have you got it?")

1.4.6. Fixed pragmatic expressions

The negation markers in the fixed pragmatic expressions are also not marked as cues because the literal negation meaning is not as salient as the expression's pragmatic function in a dialogue. Here are some examples: 对不起 *dui4bu4qi3* ("not able to treat you well - sorry"), 没关系 *mei2guan1xi* ("not a matter - It's all right. ").

2. Scope

The scope of negation should be the part of the sentence that is affected by the change of polarity. As the polarity of the cue is not changed, the cue should be outside the scope (Morante et al., 2011). Also, the scope can be discontinuous. The equivalent of the "it is not the case" test in Chinese, 并 *bing4* 不 *bu4* 是 *shi4* / 非 *fei1* ("it is not the case"), can be applied to find the longest relevant scope. For example, the paraphrase of the sentence in (46) by adding 并不是 in front of the scope is: 一定 *yi1ding4* 并不是 *bing4bu4shi4* 我 *wo3* 明天 *ming2tian1* 来 *lai2* ("it is definitely not the case that I am coming tomorrow.").

It has been argued that the semantic scope of negation follows the negation cue in Chinese (Li & Thompson, 1989). However, in the context of knowledge representation, we need to also include in the scope the subject, subject modifiers[6], and time/location adverbial phrases which are usually found before the cue in Chinese, as they provide critical information in the negation instance and are usually annotated as inside the scope in English. Other pre-cue words/phrases such as adverbs of degree and certainty are excluded in the scope annotation. The annotation in (46) provides an example.

(46)

我	明天	一定	**不**	**来**	。
wo3	*ming2tian1*	*yi1ding4*	*bu4*	*lai2*	。
I	tomorrow	definitely	NOT	come	。

("I am definitely not coming tomorrow.")

In the following analysis, when stating "scope over the entire clause/sentence", it is assumed that sentence-final particles and pre-cue phrases (other than subjects ,subject

---

[6] Notice that quantifiers such as 都 *dou1* ("all"),有些 *you3xie1* ("some") are treated as subject modifiers and are thus included in the scope. However, from the semantic analysis, quantifiers should scope over negation.

modifiers and time/location phrases) are excluded.

2.1. Negated verb:

Following the English guideline, if the verb is negated by 不 *bu4* and 没(有) *mei2(you3)*, the whole clause is under the scope of negation. If the verb happens to be the main verb, the full sentence is in the negation scope.

Chinese has many serial verb constructions in which two or more verb phrases are joined with no relations explicitly marked. There are two subcategories of this type of constructions: coordinate and subordinate verbal constructions.(Chao, 1965: 325).

2.1.1. Coordinate verbal construction

A key criteria for a coordinate verbal construction is that the serial verbs are either reversible or synonyms. The serial verbs can be juxtaposed or be connected with coordinate conjunctions such as 而 *er2* ("and"). If one of the verbs is negated, the negation annotation will neither affect nor scope over the other verbs. For example in (47)

(47)

我	把	他	弃	而	不	顾	了	。
wo3	*ba3*	*ta1*	*qi4*	*er2*	*bu4*	*gu4*	*le*	
I	BA	him	abandon	and	NOT	care	PFV	.

("I abandoned and did not care about him.") [12.172]

2.1.2. Subordinate verbal constructions

In most cases, the second verb phrase functions as a subordinate clause, and the positions of the serial verbs are irreversible. When the second verb is negated, only the verb phrase in which negation occurs is in the scope. For example in (48)

(48)

我	本	想	找	借口	不	听	他	说	。
wo3	*ben3*	*xiang3*	*zhao3*	*jie4kou3*	*bu4*	*ting1*	*ta1*	*shuo1*	。
I	originally	think	find	excuse	NOT	listen	him	say	.

("I was originally thinking of finding excuses not to listen to him.")

In the English annotation, if the lower verb phrase is a negated infinitival complement, the negation scope does not extend to the object which is the controlled subject of the lower clause. If the lower verb phrase is a that-clause, then the subject of the lower phrase is included in the negation scope. See the examples in (49) and (50).

(49) I request you **not** to go.

(50) I request that <u>you should **not** to go</u>.

However, there are no syntactic cues to differentiate the above two constructions in Chinese pivotal constructions. A Chinese pivotal construction (Li and Thompson, 1989: 607) contains a noun phrase that is both the object of the higher verb and the subject of the lower verb. We could only rely on the semantics of the first verb and distinguish pre-pivotal verbs such as 让 *rang4* ("let"), 要 *yao4* ("want") and 许 *xu3* ("allow") that introduce a meaning of "causing", and "think" verbs that take object clauses such as 想 *xiang3* ("think"), 说 *shuo1* ("say") and 认为 *ren4wei2* ("believe"). For pre-pivotal verbs, we decide to treat the lower verb phrase as equivalent to the infinitival clause in English and exclude the subject in the scope if the lower verb phrase is negated such as in (51). By contrast, when being being negated, the object clause that follows the "think" verbs will fall in the scope as a whole including the subject (52).

(51)

这	件	事	使	我	不	*能*	去	了
zhe4	jian4	shi4	shi3	wo3	bu4	neng2	qu4	le
This	CL	thing	makes	me	not	can	go	PFV

("This thing makes me unable to go.")

(52)

这些	人们	认为	产权	根本	不	存在	。
zhe4xie1	*ren2men*	*ren4wei2*	*chan3quan2*	*gen1ben3*	*bu4*	*cun2zai4*	。
These	people	think	property rights	essentially	NOT	exist	.

("These people think that property rights essentially do not exist. ") [11.143]

2.2. Negated subject/object

There is no negated subject nor negated object in Chinese as Chinese does not allow negation within the noun phrase. The negation pronoun "nothing" will be translated into negation adverb + verb in Chinese (see (53)). The determiner "no" is translated as the verbal negation marker 无 *wu2* ("do not have/exist") or 没有 *mei2you3* ("do not have/exist") in Chinese as in (54).

(53)

我	什么	也不	亏欠	他	。
wo3	*shen2me*	*ye3bu4*	*qian4*	*ta1*	。
I	any	NOT at all	owe	him	.

("I owe him nothing") [11.22]

(54)

没	有	人	注意到	它们	。
mei2	*you3*	*ren2*	*zhu4yi4dao4*	*ta1men*	。
NOT	exist	person	notice	them	.

("No one noticed them.") [3.8]

2.3. Negated adjectives

Following the English guideline, the scope of a negated adjective in a noun phrase is the noun phrase itself. It is to be noted that the negation does not scope over coordinate adjectives in the same noun phrase. For example in (55), the negation annotation does not scope over the coordinate adjective 废弃 as it is not affected by the negation cue.

(55)

那	人	住	在	这些	废弃	不用	的	小房	中	。
na4	*ren2*	*zhu4*	*zai4*	*zhe4 xie1*	*fei4qi4*	*bu4 yong4*	*de*	*xiao3fang2*	*zhong1*	。
That	person	lives	in	these	deserted	No-use	DE	little house	inside	.

("That person lives in these deserted and not-used little houses. ") [11.123]

If the negated adjective is predicative, that is, when the adjective follows a linking verb such as 是 *shi4* ("is"), 感觉 *gan3jue2* ("feel"), 对待 *dui4dai4* ("treat"), the negation scopes over the entire clause such as in (56). As adjectives are verb-like in Chinese, the copula "to be" can be omitted, and both (56) and (57) are both valid constructions in Chinese.

(56)

这样	是	不	公正	的	。
zhe4yang4	*shi4*	*bu4*	*gong1zheng4*	*de*	。
This	BE	NOT	fair	DE	.

("This is unfair.")

(57)

这样	不	公正	。
zhe4yang4	*bu4*	*gong1zheng4*	。
This	NOT	fair	.

("This is not fair")

2.4. Negated adverbs

Following the English annotation, the scope of a negated adverb which modifies the main verb in a clause should extend to the entire clause. The same principle applies to negation in the adverbial clause as in (58):

(58)

心	里	感到	不安	地	驾车	。
xin1	*li3*	*gan3dao4*	*bu4an1*	*di*	*jia4che1*	。
heart	in	feel	insecure	AD-M	drove	.

("Drove while feeling insecure.") [11.9]

2.5. Negated nominalization

2.5.1. Relative clause

In Chinese, a relative clause construction can be normalized to modify a head noun. Because the head noun is usually the complement of the predicate in the relative clause, the whole noun phrase including the head noun should be in the scope when the verb in the relative clause is negated as in (59).

(59)

他	是	个	不	爱	出风头	的	人	。
ta1	*shi4*	*ge*	*bu4*	*ai4*	*chu1feng1tou*	*de*	*ren2*	。
he	is	CL	NOT	like	show-off	DE	person	.

("He is not a person who likes to show off.")

2.5.2. Nominalisation of an appositive clause to the abstract head noun

There is also a similar nominalisation construction in Chinese that modifies the abstract head noun by specifying the 'content' of the noun. The nominalised appositive clause is itself a complete clause and the negation in the clause does not affect the polarity of the head noun. For example , (60). Therefore, the negation scope in these cases should be within the nominalisation construction excluding the head noun as in (60) (Li and Thompson, 1989: 584).

(60)

未	得	死者	近亲	同意	私	掘	坟墓	的	罪名	。
wei4	*de2*	*si3 zhe2*	*jin4 qin1*	*tong2 yi4*	*si1*	*jue2*	*fen2 mu4*	*de*	*zui4 ming2*	。
NOT	get	the deceased	next of kin	agree	privately	dig	grave	DE	charge	。

"The charge of opening the grave without consent from the next of kin of the deceased." [8.76]

2.5.3. Negation within the 是...的 *shi4....de* ("is...DE") nominalisation construction

The 是...的 construction nominalises what is in between and turns it into a subject complement. We should treat the negated 是...的 construction as a negated predicate. Therefore, the negation scope should include both the nominaliser 的 *de* and the copula 是 *shi* as in (61)

(61)

我	是	不	肯	帮	那些	坏蛋	的	。
wo3	shi4	bu4	ken3	bang1	na4xie	huai4dan4	de	。
I	BE	Not	willing	help	those	rascals	DE	.

("I refused to help those rascals. ") [11.160]

2.6. Chinese particles

2.6.1. Sentence-final particles

Sentence-final modal particles such as 的 *de,* 吗 *ma,* 呢 *ne,* 呀 *ya,* 哇 *wa* only serve to express attitude or mood towards the whole sentence rather than affecting or being affected by the negation instance within the clause. Also, sentence-final particles are analyzed as complementisers that are out of IP (Paul, 2014). Therefore, these particles should not be included in the scope of negation. For example in (62), the sentence-final particle 呀 brings emphasis rather than introducing any lexical meaning.

(62)

不	要	等	他	过	了	山	呀	！
bu4	yao4	deng3	ta1	guo4	le	shan1	ya	！
Not	shall	wait	he	past	PFV	mountain	MOD-P	！

("Don't wait until he is already past the mountain!")

Notice that some of these particles are multifunctional. We will include the sentence-final 的 when it functions as a nominalisation marker as in (61), but we will exclude sentence-final 的 in the other occasions where 的 is simply a situation marker of the whole sentence, introducing a pragmatic meaning of "such is the case." as in (63) (Chao, 1964: 296).

(63)

他	除了	时常	彻夜不眠	之外	，	早晨	总是	起得	很	晚	的	。
ta1	*chu2 shi2 le*	*chang2*	*che4ye4 bu4mian2*	*zhi1 wai4*		*zao3 chen2*	*zong3 shi4*	*qi3de2*	*hen3*	*wan3*	*de*	。
He	save	usually	All-night-not-sleep	apart	，	morning	usually	get up	very	late	MOD-P	.

("He who was usually very late in the mornings, save upon those not infrequent occasions when he was up all night.")   [1.1]

Finally, let us compare the sentence-final particle 了 *le* with the previous particles. 了 is a perfective aspect marker in Chinese, providing time reference to the negation event. As aspectual markers are consistently included into the scope in the English annotation, 了 should be included in the scope as well (see (64)).

(64)

我	问	不	出	什么	东西	了	。
wo3	*wen4*	*bu4*	*chu1*	*shen2me*	*dong1xi*	*le*	
I	ask	not	out	any	thing	PFV	.

("I have already not been able to ask anything (from her) = I cannot ask anything (from her) any more." [11.86]

2.7. Adverbs

As discussed at the beginning of this section, the position an adverb determine whether it is within the scope. Adverbs that come before the negation cue should be excluded. For example, 一定 *yi1ding4* (definitely) in (46) should be outside of the scope, but time and location adverbs should be inside   the scope.

Another case of exception made in the English guideline is "even". They argued that "even" affects the subject and should be included in the scope. Although we do not

fully agree with this decision, for the sake of consistency, we also include the Chinese equivalent of "even": 甚至... 都 *shen4zhi4 ...dou1* into the scope model.

2.8. Compound sentences

Depending on whether the component sentences in a compound sentence are in a coordinate relation or not, we divide compound sentences into coordinate constructions and subordinate constructions.

2.8.1. Coordinate constructions

Following the English guideline, negation that occurs in a coordinate clause will scope over only the clause in which negation occurs plus any recoverable elliptical elements. The coordinate clauses are typically joined by coordinate conjunctions such as 并且 *bing4qie3* ("and"), 而且 *er2qian3* ("and also") in Chinese. An example is provided in (65).

(65)

那	人	吃了一惊	，	并且	有点	不知所措	。
na4	*ren2*	*chi1leyi1jing1*	，	*bing4qie3*	*you3dian3*	*bu4zhi1suo3cuo4*	。
That	person	get shocked	，	and	a little	NOT know what to do	.

("The man looked surprised and a little embarrassed.")

It is also very common in Chinese to juxtapose two clauses together without any explicit marker. For example, the conjunction 并且 in (65) can be omitted.

2.8.2. Conditionals and other subordinate constructions in complex sentences.

Most Chinese linguists agree that there are two types of subordinate constructions: complex sentences and "pregnant" sentences (Chao, 1965). A major difference between the two types of constructions is that a complex sentence consists of two complete sentences whereas the mother clause in the "pregnant" sentence will be incomplete without the child clause. As the "pregnant" constructions have already been discussed in serial verb constructions (see 2.1.1), the current section is mainly

devoted to the complex sentence type of subordinate constructions introduced by subordinate conjunctions such as 如果 *ru2guo3*... 就 *jiu4*... ("if...then..").

The English guideline states that if negation occurs in the subordinate clause, only the subordinate clause excluding the subordinate conjunctions (如果 *ru2 guo3* ("if")....的话 *de hua4,* 就 *jiu4* ("then")... ; 当 *dang1*..的时候 *de shi2hou* ("when")) is in the scope. If the main clause is negated, the subordinate clause will also be included in the scope with the subordinate conjunction. A typical subordinate construction is the conditionals as shown in (66) and (67). In (66), we include the conditional clause into the scope because the condition affects the semantics and the factuality of the negation event in the main clause.

(66)

如果	你	不	去	的话	，	我	就	去	。
ru2guo3	*ni3*	*bu4*	*qu4*	*dehua4*	，	*wo3*	*jiu4*	*qu4*	。
if	you	NOT	go	if	，	I	then	go	.

("If you don't go, I will go then.")

(67)

如果	你	去	的话	，	我	就	不	去	。
ru2guo3	*ni3*	*qu4*	*dehua4*	，	*wo3*	*jiu4*	*bu4*	*qu4*	。
if	you	go	if	，	I	then	NOT	go	.

("If you go, I will not go then.")

Apart from 如果 *ru2 guo3* .... 就 *jiu4...,* there are alternative ways to construct a conditional in Chinese. For example using 不 *bu4* ("not") as in ()

(68)

不	弄清	一切	,		决不	离开
bu4	*nong4qing1*	*yi1qie4*	,		*jue2bu4*	*li2kai1*
(if) NOT	clarify	everything	,		NOT	leave

("I will not leave if I don't clarify everything.") [11.254]

Some conditionals are constructed by 只要 *zhi3yao4*...就 *jiu4* ("as long as") as in (68). The English annotation does not seem to have treated the subordinate clause headed by "as long as" as a conditional statement (sometimes not even as a subordinate clause). However, there is no one-to-one translation between these conditional constructions from English to Chinese. For example, 只要 *zhi3yao4*...就 *jiu4* can be the translation of "if...". It is not possible to follow the English guideline in this case. We will annotate all types of conditional constructions in Chinese consistently as in the example of (68).

(68)

只要	他	怕	打	，	他	就	不	是	什么	鬼怪	。
zhi3 yao4	*ta1*	*pa4*	*da3*	,	*ta1*	*jiu4*	*bu4*	*shi4*	*shen2 me*	*gui3 guai4*	。
as long as	he	fear	(being) hit	,	he	then	NOT	BE	any	ghost	.

("If he was vulnerable, he was mortal.") [14.102]

The English annotation, in practice, does not include all other types of subordinate clauses introduced by 虽然 *sui1ran2*...但是 *dan4shi4*... ("although..but..."), 因为 *yin1wei2*...所以 *suo3yi3* ("because...therefore..") into the scope when the main clause is negated. For transferability, we follow the English annotation to only annotate conditional constructions and when-subordinate clauses as subordinate complex sentences and annotate all other complex sentences as coordinate constructions.

2.9. Special constructions

2.9.1. 除了 *chu2le* ("except and save")

"Except" and "save" are translated into 除了 *chu2le* in Chinese. Following the translational principle, we will annotate 除了 as a negation cue if it introduces negative polarity to the main clause such as in (63)

2.9.2. Ellipsis:

The elliptical element and the ellipsis in English usually have the same grammatical category for example in (70). Chinese exhibits more versatile ellipsis constructions. For example, the object of the previous clause can be the elided subject in the next clause as in (71)

(70)

I repeat that <u>the lady is</u> his wife and **not** <u>his sister</u> [12.87]

(71)

这里	出现	了	一	位	青年人	，	不	到	三十岁	。
zhe4 li3	chu1 xian4	le	yi1	wei4	qing1nian2 ren2	,	bu4	dao4	san1shi2sui4	。
here	emerge	PFV	a	CL	young man	,	NOT	reach	thirty-years old	.

("There emerges a young fellow who is under thirty.") [1.64]

The elliptical element with a different grammatical category as in (71) is usually translated as a relative pronoun which would be considered within the scope of negation in English. Therefore, such an elliptical element in Chinese should also be included in the scope. In fact, any recoverable elliptical element of negation should be included into the scope regardless of its grammatical category.

2.9.3. Existential sentence

Unlike English, Chinese does not require an existential subject "There is". Optionally, Chinese could place a locative NP as a surface subject as in (72). This NP should be included in the scope of negation (Huang, 1987).

(72)

这儿	没	有	人	。
zhe4er	*mei2*	*you3*	*ren2*	。
Here	not	exist	person	.

("There is no one here.")

2.9.4. Interrogative pronoun

Following the English guideline, we should include interrogative pronouns such as 为什么 *Wei4shen2me* ("why") and 什么 *shen2me* ("what") in the scope annotation. An example is given in (73):

(73)

为什么	不	到	房子	里面	去	拜访	呢	?
wei4 shen2me	*bu4*	*dao4*	*fang2zi*	*li3 mian4*	*qu4*	*bai4 fang3*	*ne*	?
why	not	go-to	house	inside	to	visit	Q-P	?

("Why did (you) not go into the house to visit (him)?") [11.74]

2.9.5. Comparative constructions

(74) shows a typical Chinese comparative construction where X is the thing to be compared and Y is the standard. Typical comparison coverbs are 比 *bi3* ("than") and 象 *xiang4* ("like").

(74) X comparison-coverb Y dimension (Li & Thompson: 564)

The negation scope is determined by applying the 并不是 ("it is not the case") test. The semantic interpretation of the test result suggests that if the negation occurs before the coverb, the entire sentence is under the scope (see (75)). If negation occurs in the dimension, then only X and the dimension are in the scope for example in (76). If the negation occurs in Y, the negation only scopes over Y (see (77)).

(75)

约翰森	先生	年纪	不	比	您	大	。
yue1han4sen1	xian1sheng1	nian2ji4	bu4	bi3	nin2	da4	。
Johnson	Mr.	age	NOT	COMP-V	you	old	.

("Mr. Johnson is not older than you.") [5.11]

(76)

我	的	觉	睡得	比	平常	还要	不	踏实	。
wo3	de	jiao4	shui4 de	bi3	ping2 chang2	hai2 yao4	bu4	tashi	。
I	GEN	slumbers	slept	COMP-V	usual	even	NOT	easy	.

("My slumbers have been lighter.") [8.98]

(77)

我	这样	问	比	事情	无	法	收拾	好	。
wo3	zhe4 yang4	wen4	bi3	shi4 qing	wu3	fa3	shou1 shi	hao3	。
I	like this	ask	COMP-V	matter	NOT-have	means	control	good	.

("It is better that I ask (them) like this than that the matter should pass out of control.") [11.34]

3. Event

Following the English guideline, the event annotation is minimized to only the head of the negated phrase when it is factual. For example in (78), only the head of the negated predicate, 朋友 *peng2you* ("friend"), is marked as the event.

(78)

<u>他的</u>	<u>朋友</u>	也	**不**	<u>是</u>	<u>我的</u>	*<u>朋友</u>*	。
ta1de	peng2you	ye3	bu4	shi4	wo3de	peng2you	。
his	friends	also	NOT	BE	my	friends	.

("His friends are (also) not mine.")   [11.22]

3.1. Negated verb

The event of a negated verb phrase should be the main verb. Following the English annotation, we will not annotate any aspectual markers as an event. For example, the durative aspect marker 在 *zai4* is excluded from the event annotation in (79)

(79)

没	<u>有</u>	<u>人</u>	<u>在</u>	*<u>动</u>*	。
mei2	*you3*	*ren2*	*zai4*	*dong4*	。
Not	exist	one	DUR	move	.

("Nobody is moving")

Sometimes the aspectual marker occurs within the verb phrase. For example in (80), 过 *guo4* as the experiential aspect marker should be excluded from the event annotation.

(80)

他	还	没有	结过婚	啊	！
ta1	*hai2*	*mei2you3*	*jie2guo4hun1*	*a1*	！
He	still	NOT	married-EXP-married	MOD-P	！

("He is not a married man.") [13.216]

3.2. Coverb

English does not typically negate prepositions that modify a verb, for example in (81). Instead, the verb would be marked as the negated event.

(81) I do **n't** *study* at home. *I study not at home.

In Chinese, the negation can modify both the verb and the preposition/coverb phrase as in (82) and the two sentences in (82) translate the same.

(82)

我	不	在	家	里	学习	。
wo3	bu4	zai4	jia1	li3	xue2xi2	。
I	NOT	at	home	in	study	.

("I do not study at home")

我	在	家	里	不	学习	。
wo3	zai4	jia1	li3	bu4	xue2xi2	。
I	at	home	in	NOT	study	.

("I do not study at home.")

The Chinese coverbs are equivalent to English prepositions. The preposition 在 *zai4* in (82) is an instance of coverb in Chinese. A coverb originates from a verb but is functioning as a preposition in Modern Chinese. A coverb typically occurs in the following construction:

Subject + coverb + NP +verb phrase (Li and Thompson, 1989: 356).

Since a coverb is not a real verb, the verb 学习 *xue2xi2* ("stud") is still the main verb in (82). If the main verb is negated, the negation scopes over the entire sentence. If the coverb is negated, what comes after the cue in the clause should be within the scope. Therefore the scope is still 我在家学习 in (82).

Also, we follow the translational equivalence principle and do not mark the coverbs as the events . Instead, we should mark the main verb that comes after as the event. Therefore, the event in both sentences in (82) is 学习. Notice that Several coverbs in Chinese have already been grammaticalized and solely function as grammatical markers. For example 被 *bei4* is a passive voice marker and 把 *ba3* and 对 *dui4* are direct object markers. An example of the coverb 把 is in (83)

(83)

我	父亲	根本	**没**	把	我	放	在	心	上	。
wo3	*fu4qin*	*gen1ben*	*mei2*	*ba3*	*wo3*	*fang4*	*zai4*	*xin1*	*shang4*	。
I	father	essentially	NOT	BA	me	put	at	heart	on	.

("My father essentially did not care about me.") [11.23]

By not annotating these coverbs as the events, the Chinese annotation becomes more comparable to the English annotation, since there is usually no equivalent English translation to these "empty" coverbs.

3.3. Factuality

As the negated event should be a fact, we do not mark the events in the following non-factual constructions.

3.3.1. Imperatives :

Negated imperatives in Chinese are usually introduced by 别 *bie2,* 不要 *bu4 yao4.* Following the English guideline, we do not annotate events in imperatives as in (84) because they are non-factual orders or requests.

(84)

不		要		出声		。
bu4		yao4		chu1sheng1		。
NOT		shall		make a sound		.

("Hush.") [12.130]

3.3.2. Non-factual interrogatives:

Most non-rhetorical yes-no questions are non-factual such as in (85). A speaker asks a yes-no question because he/she is not sure whether a statement is true or false. It is not explained in the English guideline what counts as a factual interrogative. We assume that questions introduced by interrogative pronouns are factual. For example in (86), the event of "not going to the house" is factual because it happened. Therefore, 到 dao4 ("go to") in (86) should be marked as the event.

(85)

那	位	女士	不	在	吗	?
na4	wei4	nv3shi4	bu4	zai4	ma	?
that	CL	lady	NOT	present	Q-P	?

("The lady is not there?") [14.50]

(86)

为什么	不	到	房子	里面	去	拜访	呢	?
wei4sh en2me	bu4	dao4	fang2zi	li3mia n4	qu4	bai4fang 3	ne	?
why	not	go-to	house	inside	to	visit	Q-P	?

("Why did (you) not go into the house to visit (him)?") [11.74]

3.3.3. Conditional constructions:

Following the English guideline, we do not mark events in both the conditional clause and the main clause of a conditional construction (see (66) and (67)).

3.3.4. Modal constructions:

Modality is typically related to non-factuality as it is mainly about possibility and necessity (Li, 2004). Therefore we should not annotate events in clauses introduced by modal verbs or verbs that express modality such as 不见得 *bu4jian4de2* ("probably not"). The exception made in the English guideline is the modal verb "can" when it expresses participant internal ability. The Chinese equivalent of "can", 能 *neng2*, is also multifunctional and can introduce either ability or possibility. We should mark 能 as an event when it means "able to" as in (87). In addition, Chinese has other modal verbs that can express participant-internal ability such as： 能够 *neng2gou4,* 会 *hui4,*可以 *ke3yi3,* 可 *ke3.*

(87)

我	不	*能*	早一点	到	那里	去	。
wo3	bu4	neng2	zao3yi1dian3	dao4	na4li3	qu4	。
I	NOT	can	earlier	to	there	go	.

("I can't go there earlier.")   [11.73]

These modal verbs belong to the group of non-epistemic participant internal modal verbs. These modal verbs are in contrast with the epistemic modality which concerns the speaker's degree of uncertainty and probability towards a truth proposition (Lyons, 1977: 823) and also the participant-external deontic modality defined by Lyons (1977: 823) as the concerns about necessity or possibility of a moral action. The non-epistemic participant internal modal verbs express the participant's ability or need. In addition to the "ability" type of modal verbs specified by the English guideline, we decide to annotate the "need" type of modal verbs as events in Chinese, e.g. 需要 *xu1yao4* ("need") ,要 *yao4* ("want"), partly because it is sensible to treat the group of non-epistemic participant-internal modal verbs consistently, and partly because 需要 and 要 are usually translated into lexical verbs such as "want" in English.

3.3.5. Suppositions or presumption:

To make a judgment on the factuality, we should examine the semantics of the verb that introduces the scope. If the verb suggests the speaker's certainty about the content of the following clause (e.g. 确信 *Que4xin4* ("sure"), 确定 *que4ding4* ("certain"), 知道 *zhi1dao4* ("know")), we should treat the clause as factual and annotate the negated event in the clause. If, however, the verb suggests that the following statement is the speaker's supposition or presumption in the case of 相信 *xiang1xin4* ("believe"), 认为 *ren4wei2* ("believe"), 觉得 *jue2de2* ("think"), 想 *xiang3* ("think"), 害怕 *hai4pa4* ("fear"), etc., the negated event in the statement should not be marked as in (88).

(88)

我	相信	您	决不	愿意	做	一	个	妨碍	别人	的	人	。
wo3	*xiang1 xin4*	*nin2*	*jue2 bu4*	*yuan4 yi4*	*zuo4*	*yi1*	*ge4*	*fang2 ai4 ren*	*bie2 ren*	*de*	*ren2*	。
I	believe	you	NOT	want	be	a	CL	hinder	others	DE	person	.

("I believe that you do not wish to be a spoil-sport.") [9.44]

3.3.6. Future tense:

As a language with no morphological tense marker, Chinese employs various linguistic devices to indicate future tense.

3.3.6.1. Deontic modal verbs

Notice that deontic modal verbs are intrinsically futuristic. Deontic modality, as Lyons (1977: 823) describes, "is concerned with the necessity or possibility of acts performed by morally responsible agents". The core semantic concepts of deontic modality are obligation and permission which are both associated with future actions as in (89):

(89)

对		他	不	应该	原谅		。
dui4		*ta1*	*bu4*	*ying1gai1*	*yuan2liang4*		。
object marker		him	NOT	should	forgive		.

("He should not be forgiven.") [9.246]

3.3.6.2. Time adverbs

Phrases such as 明天 *ming2tian1* ("tomorrow"), 下周 *xia4zhou1* ("next week"), 将 *jiang1* ("going to") , 以后 *yi3hou4* ("later on") can also introduce future actions.

我	以后	再也不	提	这	件	事	了	。
wo3	*yi3hou4*	*zai4ye3bu4*	*ti2*	*zhe4*	*jian4*	*shi4*	*le*	。
I	later on	never	mention	this	CL	matter	PFV	.

("I will **not** mention this matter again.")[7.103]

3.3.7. Neutralized negation

An additional case of non-factuality which has not been mentioned in the original English guideline is when the negation is neutralized because it occurs in the scope of a higher negation cue as in (90).

(90)

没	有	一	天	他	不	到	庄园	来	看	我们	。
mei2	*you3*	*yi4*	*tian1*	*ta1*	*bu4*	*dao4*	*zhuang1yuan2*	*lai2*	*kan4*	*wo3men*	。
NOT	exist	one	day	he	NOT	to	Hall	come	see	us	.

("Not a day has passed that he has not called at the hall to see us.") [10.]

3.4. Copulative and existential constructions

Following the English annotation guideline, we do not annotate the copulative verb 是 *shi4* ("to be") and the existential verb 有 *you3* ("there is/exit") as events when they are negated. Instead, we mark the head of the following noun phrase as the event (see 78).

Notice that if 有 *you3* ("to have/exist") functions as a link verb that introduces a lower verb phrase, the event should be the verb of that verb phrase rather than the noun phrase following the link verb. This decision is in correspondence to the translational equivalence principle. For example in (91), the annotation of the English translation is "**No** <u>one</u> *<u>lived</u>* <u>here</u>." with the verb "lived" marked as the event.

(91)

没	有	人	住	过	这	。
mei2	*you3*	*ren2*	*zhu4*	*guo4*	*zhe4*	。
Not	exist	one	live	EXP	here	.

("No one lived here.")

3.5. Negated subject complement

The event of a negated subject complement is the head of the complement. If the head of the subject complement is a coordinate construction, then the whole construction should be marked as the event as in (92).

(92)

我	需要	的	不	是	传说	和	谣言	。
wo3	*xu1yao4*	*de*	*bu4*	*shi4*	*chuan2shuo1*	*he2*	*yao2yan2*	
I	need	GEN	NOT	BE	legends	and	rumours	.

("What I need are not legends and rumors.") [12.277]

3.6. Negated clause

If the whole clause is negated, we shall mark the main verb of the clause as the event as in (93):

(93)

并非	全	信	都	能	读	。
bing4fei1	*quan2*	*xin4*	*dou1*	*neng2*	*du2*	。
NOT	all	letter	DOU	can	read	.

("The letter was not all legible.")   [11.95]

3.7. Verb-complement compound

We annotate the whole compound as the event if the complement or the verb is bound, or if the meaning of the whole can not be derived compositionally from the verb plus the complement. For example, the following verb-complement compounds will be annotated as a whole when being negated: 看见 *kan4jian4* ("see-appear: see"), 听到 *ting1dao4* ("hear-arrive: hear"), 看出来 *kan4chu1lai2* ("see-out: see"), 弄明白 *nong4ming2bai* ("make clear: understand"). Otherwise, the event is the main verb such as in 说不清楚 *shuo1 bu4 qing1chu3* ("cannot speak clearly").

There are two reasons for this decision: (1) As discussed in 1.2.1, infixal negation in verb-complement compound is in fact negation of the potential form and introduce the meaning of "cannot" that directly negates the main verb. (2). Such an analysis will more closely follow the English annotation as English usually translates the free-standing complement as an adverb in a compositional verb-complement compound but uses a verb to translate a bound verb-complement compound in Chinese. However, we should bear in mind that some Chinese linguistic analysis treats the complement as the main event of negation. (Li and Thompson, 1989).

3.8. Comparative constructions mark "more than" as the event.

Following the English annotation guideline, the event of a negated comparative construction is the dimension construction 更 *geng4* +adj ("more + adj.") rather than the comparative coverb 比 *bi3*. An example can be found in (94).

(94)

我	从没有	比	这	更	快活	过	。
wo3	*cong2mei2you3*	*bi3*	*zhe4*	*geng4*	*kuai4huo*	*guo4*	。
I	NEVER	COMP-V	this	more	glad	EXP	.

("I never was more glad than this.") [12.9]

3.9. Negated set phrases

If a set phrase is negated by the cue, the phrase should be treated as the event as a whole. An example is shown in (95). A set phrase in Chinese is typically treated as a multisyllabic multi-morpheme word. Since the English annotation does not annotate morphemes within a negated event, there is no reason to separate the characters in the set phrases in event annotation in Chinese.

(95)

我	并不	把	和	他	妹妹	的	感情	引以为耻	。
wo3	*bing4bu4*	*ba3*	*he3*	*ta1*	*mei4mei4*	*de*	*gan3qing2*	*yin3yi3wei2chi3*	。
I	NOT	BA	with	his	sister	DE	feelings	ashamed	.

("I am not ashamed of my feelings towards his sister.") [9.115]

3.10. Negated compounds

If the negated event in a sentence contains bound morphemes, the whole compound should be the event including any affixes. Here are some example annotations: 不可爱*bu4ke3ai4* ("not love-able=not lovely"), 不好吃 *bu4hao3chi1* ("not good-eat= not delicious").

Notice that 可 *ke3* ("can") usually functions as a non-epistemic participant-internal modal verb rather than a prefix in four-character set phrases. For example, the negated event in 无可估量*wu2ke3gu1liang4* ("not able to estimate") should be 可 *ke3* ("can") alone.

Appendix B. Abbreviation

AD-M   adverb marker e.g. 地 *di*

BA   把 *ba3* (object marker)

CL   classifier

COMP-V   comparative coverb

CONJ   conjunction

DE   的 *de*

DUR   durative aspect e.g. 在 *zai4*

EXP   experiential aspect e.g. 过 *guo4*

GEN   genitive e.g. 的 *de*

JIANG   将 *jiang1* (object marker)

MOD-P   modal particles e.g. 啊 *a1*

PFV   perfective aspect e.g. 了 *le*

Q-P   question particle e.g. 吗 *ma*

(The abbreviations are adapted from Li & Thompson (1989))

References

Aziz, W. (n.d.). Chinese word segmentation. Retrieved August 17, 2016, from http://wilkeraziz.github.io/work/2014/11/04/chinese-word-segmentation.html

Blanco, E., & Moldovan, D. I. (2011, March). Some Issues on Detecting Negation from Text. In FLAIRS Conference (pp. 228-233).

Brooke, J., Tofiloski, M., & Taboada, M. (2009, September). Cross-Linguistic Sentiment Analysis: From English to Spanish. In RANLP (pp. 50-54).

Chao, Y. R. (1965). A grammar of spoken Chinese. Univ of California Press.

Chapman, W. W., Bridewell, W., Hanbury, P., Cooper, G. F., & Buchanan, B. G. (2001). A simple algorithm for identifying negated findings and diseases in discharge summaries. Journal of biomedical informatics, 34(5), 301-310.

Denisowski, P. (1999). Cedict (chinese-english dictionary) project.

Dyer, C., Chahuneau, V., & Smith, N. A. (2013). A simple, fast, and effective reparameterization of ibm model 2. Association for Computational Linguistics.

Fancellu, F., & Webber, B. L. (2014, April). Applying the semantics of negation to SMT through n-best list re-ranking. In EACL (pp. 598-606).

Farkas, R., Vincze, V., Móra, G., Csirik, J., & Szarvas, G. (2010, July). The CoNLL-2010 shared task: learning to detect hedges and their scope in natural language text. In Proceedings of the Fourteenth Conference on Computational Natural Language Learning---Shared Task (pp. 1-12). Association for Computational Linguistics.

Huang, C.-T. James. 1987. Existential sentences in Chinese and (in)definiteness. In The Representation of (In)definiteness, eds. Eric J. Reuland and Alice G.B. ter Meulen, 226-253. Cambridge, MA: MIT Press.

Hwa, R., Resnik, P., Weinberg, A., & Kolak, O. (2002, July). Evaluating translational correspondence using annotation projection. In Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (pp. 392-399). Association for Computational Linguistics.

Li, R. (2004). Modality in English and Chinese: A typological perspective. Universal-Publishers.

Li, C. N., & Thompson, S. A. (1989). Mandarin Chinese: A functional reference

grammar. Univ of California Press.

Lyons, J. (1977) Semantics. 2nd vol. Cambridge: Cambridge University Press

Mihalcea, R., Banea, C., & Wiebe, J. M. (2007). Learning multilingual subjective language via cross-lingual projections. UNT Digital Library. http://digital.library.unt.edu/ark:/67531/metadc30991/. Accessed August 20, 2016.

Morante, R., & Blanco, E. (2012, June). * SEM 2012 shared task: Resolving the scope and focus of negation. In Proceedings of the First Joint Conference on Lexical and Computational Semantics-Volume 1: Proceedings of the main conference and the shared task, and Volume 2: Proceedings of the Sixth International Workshop on Semantic Evaluation (pp. 265-274). Association for Computational Linguistics.

Morante, R., & Daelemans, W. (2012). ConanDoyle-neg: Annotation of negation in Conan Doyle stories. In Proceedings of the Eighth International Conference on Language Resources and Evaluation, Istanbul.

Morante, R., Schrauwen, S., & Daelemans, W. (2011). Annotation of negation cues and their scope Guidelines v1. Computational linguistics and psycholinguistics technical report series, CTRS-003.

Padó, S., & Lapata, M. (2009). Cross-lingual annotation projection for semantic roles. Journal of Artificial Intelligence Research, 36(1), 307-340.

Paul, W. (2014). Why particles are not particular: Sentence‐final particles in Chinese as heads of a split CP. Studia Linguistica, 68(1), 77-115.

Skeppstedt, M. (2011). Negation detection in Swedish clinical text: An adaption of NegEx to Swedish. Journal of biomedical semantics, 2(3), 1.

Shou-Hsin, T. (1978). Negation in Chinese: Mandarin and Amoy. Journal of the American Oriental Society, 50-60.

Tseng, H., Chang, P., Andrew, G., Jurafsky, D., & Manning, C. (2005, October). A conditional random field word segmenter for sighan bakeoff 2005. In Proceedings of the fourth SIGHAN workshop on Chinese language Processing (Vol. 171).

Varga, D., Halácsy, P., Kornai, A., Nagy, V., Németh, L., & Trón, V. (2007). Parallel corpora for medium density languages. AMSTERDAM STUDIES IN THE THEORY AND HISTORY OF LINGUISTIC SCIENCE SERIES 4, 292, 247.

Vincze, V., Szarvas, G., Farkas, R., Móra, G., & Csirik, J. (2008). The BioScope corpus: biomedical texts annotated for uncertainty, negation and their scopes. BMC

bioinformatics, 9(11), 1.

Wetzel, D., & Bond, F. (2012, July). Enriching parallel corpora for statistical machine translation with semantic negation rephrasing. In Proceedings of the Sixth Workshop on Syntax, Semantics and Structure in Statistical Translation (pp. 20-29). Association for Computational Linguistics.

Zhou, B. Z. Q. Z. G. (2015). Negation and speculation identification in Chinese language. In Proceeding of the Annual ACL Conference 2015.