

Annotating Modal Expressions in the Chinese Treebank

Yanyan Cui Ting Chi

Department of Linguistics, Georgetown University
1437 37th Street, NW, Poulton Hall 240
Washington, DC 20057-1051
{yc285, tc374}@georgetown.edu

Abstract

This paper reports an effort to annotate modality in the Penn Chinese Treebank. We introduce the modals and features that were annotated, and describe the phases of our working process. Along with this, we address the issues in the preparation of annotation guidelines, and present the preliminary results of the first pass. Finally, we analyze the types of disagreement, and propose directions to improve consistency.

1 Introduction

Since its release to the public in 2000, the Penn Chinese Treebank (Xia et al.1999) has been annotated with several layers of semantic information such as predicate-argument structures and discourse connectives (Xue 2003, 2005). Our effort, as a part of a larger cross linguistic annotation project, aims to expand this body of work with modal annotation.

Modality is the aspect of meaning that expresses states of affairs beyond the **actual** (Hacquard 2011). Distinguishing between the actual versus modal information is necessary for a wide range of natural language processing (NLP) applications such as sentiment analysis (Wiebe et al. 2005), question answering (Sauri et al 2006), medical information extraction (Mowery et al. 2012), etc.

In recent years, many efforts have been made to create resources of manually annotated modality information. These resources vary greatly in terms of what aspects of modality are annotated and how

the features are marked. The diverse goals and backgrounds of the researchers determined this variety of annotation schemes. Hacquard and Wellwood (2012), for example, annotated the interpretation (root vs. epistemic) of modal words in a range of embedded contexts. Their goal is to answer a particular formal semantic question--whether epistemic modals contribute to sentence meaning, and consequently can be embedded in various environments. Mowery et al. (2012) on the other hand, targets a particular practical problem, namely distinguishing negated, affirmed and uncertain information in medical texts. Their project annotated the polarity (positive vs. negative) of sentences, and the degree of certainty (moderate vs. high) associated with a statement. Hendrickx et al. (2012) did yet another type of work, which is motivated from a theoretical perspective, but tries to facilitate potential NLP research as well. Their scheme not only covers more semantic properties of modality (what is the trigger, what is its target, who is the source of the modality, etc.), but also has a more fine-grained distinction of modal values (eight main values and several sub-values).

The goal of our annotation is similar to that of Hendrickx et al. We aim to create a resource that provides detailed semantic analyses to a set of prototypical modal expressions in Mandarin. The produced corpus will allow for both linguistic studies (e.g. the ranges of constructions a certain modal expression occurs) and various machine learning experiments.

A secondary goal of our project is to test the cross-linguistic adaptability of the schema we adopt, which is originally developed by Rubinstein

et al. (2013). This schema is supposed to be language-independent, and we applied it to Mandarin with minimum modification.

Started in the fall of 2012, we have so far completed the first pass of annotation on 200 files in the Treebank which are articles from Xinhua newswire^[1]. In this process, several goals are achieved:

- We created a first draft of annotation guidelines by modifying the guidelines designed for the parallel English annotation task.
- We calculated agreement measures for different granularities of various annotated features.
- We learned of the difficulties involved in annotation of individual features.

The paper is organized as follows. In section 2, we provide an overview of the project by outlining the modal inventory, the features annotated, and the working process. In section 3, we discuss the issues involved in developing guidelines. Section 4 presents the results of the first pass of annotation, and section 5 discusses the disagreement patterns revealed by these results. Finally, we conclude the paper in section 6.

2 Overview

The annotation is carried out with MMAX2 (Müller & Strube 2006), with a scheme of ten features, and an evolving modal-list. Two annotators are involved, and they are also responsible for the creation of guidelines and quality control.

2.1 Working process

Expecting constant modification of guidelines as well as human errors that need to be corrected regularly, we break the process into small rounds. Each round consists of around 200 files and is divided into four phases, (i) preprocessing of files, (ii) a blind double annotation, (iii) an evaluation, and (iv) a revision. In preprocessing, files to be

annotated are prepared, and modals in the modal-list are pre-highlighted for the annotators. In the second phase, the two annotators work independently without discussion. In the next phase, disagreement is measured and analyzed. The inconsistent instances are retrieved and reconsidered. The two annotators work together to resolve the disagreement. In the meantime, guidelines are revised to account for newly encountered issues, while true ambiguities (Rubinstein 2012) will be embraced by keeping both annotations. After revision, the result is re-evaluated then re-revised, until the consistency achieves a pre-decided threshold.

PHASES

Preprocessing

Pre-highlighting modals in the modal-list

Annotation

Blind double annotation: 200 files

Evaluation

Measure Inter-annotator agreement

(using Kappa score)

Analyze disagreement

(employing confusion matrix)

Revision

Correct errors

Resolve disagreement

Revise guidelines

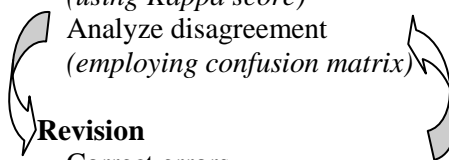


Table 1: working process

2.2 Modal list

The initial modal-list contains 11 entries collected from linguistic literature, most of which are auxiliary verbs. In the first pass, some adverbs are discovered and added to the modal inventory. The updated modal-list is shown in Table 2; the cells containing the acquired modals are shaded.

2.3 Features

Annotators mark not only modality type but also the relation between a modal and various components of the sentence. The annotated features are enumerated in Table 3.

^[1] We completed blind double annotation and evaluation, but have not finished error correction and guidelines revision.

Item	Pinyin	POS	Gloss	#token
可能	keneng	Aux.	possible	74
应该	yinggai	Aux.	should	6
会	hui	Aux	sill	40
可	ke	Aux	can/may	39
必须	bixu	Aux	must	12
得	dei	Aux	have to	8
要	yao	Aux	need to	48
可以	keyi	Aux	may	17
能	neng	Aux	be able to	27
能够	nenggo	Aux	be able to	6
一定	yiding	Adv	definitely	3
将	jiang	Adv	will	143
可望	kewang	Adv	hopefully	5
无望	wuwang	Adv	impossible	0
应	ying	Adv	should	4
total				432

Table 2: modal-list

FEATURE	DESCRIPTION
Modality type	The flavor of modals; e.g. <i>epistemic, deontic</i> , etc
Predicate type	Whether the modal is in its comparative, equative or superlative form
Prejacent	The propositional argument of the modal
Modified element	The NP , AdjP that are modified by the modal
Degree indicator	The element that indicates the degree of the modal
Source	The entity that is responsible for the knowledge, rules etc. that the modal claim is based on; e.g. [_{source} John] believes that Mary might come.
Background	Information that provides the background of the modal statement; see(1)
Environmental attitude	The attitude verb embedding the modal; see (1)
Environmental polarity	Whether the modal is in the scope of a sentential negation
Outscoping quantifier	A quantifier in the syntactic scope of a modal but semantically scopes over it

Table 3: features

Priority
<ul style="list-style-type: none"> Deontic: the claim is based on rules, standards, social norms, etc. <p>(1a) <i>Xinwen bixu zhenshi.</i> News must real 'News must be real.'</p>
<ul style="list-style-type: none"> Bouletic: the claim is based on desires <p>(1b) You should try this chocolate*.</p>
<ul style="list-style-type: none"> Teleological: the claim is based on one's goal <p>(1c) <i>Zhongguo bixu jinxing gaige, yi zengqiang zishen jingzhengli.</i> China must make reform to improve self competitiveness 'China must make reforms to improve its competitiveness.'</p>
Dynamic (ability_circumstantial)
<ul style="list-style-type: none"> Circumstantial: the claim is based on circumstances <p>(1d) <i>Zai ci jichu shang, jinnian de jingji prep. this basis loc. this year DE. economy zengzhang mubiao wanquan keyi shixian.</i> growth goal completely can realize 'On this basis, this year's goal of economic growth can absolutely be achieved.'</p>
<ul style="list-style-type: none"> Ability: the claim is based on what the agent can do <p>(1e) <i>Zhongguo yi neng shengchan shang wan men</i> China already can produce over 10,000 Cl. <i>shuzi dianhua chengkong jiaohuanji</i> digital telephone SPC exchange 'China can already produce over 10,000 digital telephone SPC exchanges.'</p>
Epistemic
<ul style="list-style-type: none"> the claim is based on's belief of knowledge <p>(1f) <i>Jingguo shidang tiaozheng, dongya jingji</i> through proper adjustments East Asia economy <i>yiding hui jixu xiangqian fazhan</i> definitely will continue forward develop 'With proper adjustment, the economy of East Asia will continue to develop.'</p>

Table 4: examples of modality types

Following the hierarchical classification of modal flavors proposed by Portner (2009), we consider six atomic values for modality type:

* We have not encountered any instance both annotators mark as bouletic. The example is from Portner (2009), p133.

epistemic, circumstantial, ability, deontic, bouletic and teleological. The last three values are subtypes of **priority** modal, while circumstantial and ability are subtypes of **dynamic** modal*. In cases where fine-grained decisions cannot be made, coarser categories are available for selection. For priority modals, in addition to the super-type *priority*, there is an option bouletic_teleological. For non-priority modals, there are epistemic_circumstantial and ability_circumstantial. These collapsed values are created on the basis of a pilot study run on Amazon's Mechanical Turk. (Rubinstein et al 2012).

Table 4 provides descriptions and examples for the atomic values; instances of the coarse classes are shown in (2a-c):

bouletic teleological

(2a) *Su-gang yao kao da yunhe*
 Su-gang need to rely on great canal
yunshu yuanliao he chengpin
 transport raw material and product
 ‘Su-gang Group needs to transport raw materials and products via the Great Canal.’

epistemic circumstantial

(2b) *Shenzhen tequ jinnian guonei*
 Shenzhen special district this year domestic
shengchan zongzhi ke da yiqianyibaisanshiyi
 product gross can reach 11.3 billion
 ‘GDP of Shenzhen this year can reach 11.3 billion (yuan).’

ability circumstantial

(2c) *Wanqi yi-gan jibing wangwang hui*
 advanced hepatitis B disease often can
zhuanhua wei ai
 turn to as cancer
 ‘Advanced stage of hepatitis B can often turn to cancer.’

Modal tokens rarely appear with all ten features. At minimum, *modality type* is specified for each modal, and in most cases, the *prejacent* of modals is marked as well. Other frequent features include *background* and *environmental attitude*. (3) illustrates how these features are annotated.

(3) *Ju shanghai shi ji-wei*
 according to Shanghai-city planning committee
zhuanjia [fenxi yuce], [*yao zai 2000*
 expert analyze estimate want to in 2000
nian shixian ren-jun guonei shengchan
 year realize per capita domestic product
zong-zhi wu qian meiyuan de mubiao,]
 gross five thousand dollar DE goal
[jin-hou san nian hanghai guonei
 from now three year Shanghai domestic
shengchan zong-zhi pingjun nian-zengfu]
 product gross average annual-growth
yao[teleological] [dadao bai-fen-zhi-shi zhi
 need to arrive at 10% to]
bai-fen-zhi-yi
 eleven percent

‘According to the analysis and estimation by the experts from the planning committee of Shanghai city, to achieve a per capita GDP of five thousand dollars by the year 2000 in the following three years, the annual GDP growth of Shanghai needs to be around 10% to 11%.’

- **modal:** yao, ‘need to’
- **modality type:** priority --> teleological
- **prejacent:** ‘in the following three years, the annual GDP growth of Shanghai needs to be around 10% to 11%’
- **background:** ‘to achieve a GDP of five thousand dollars by the year 2000’
- **environmental attitude:** ‘analyze and estimate’

3 Development of Guidelines

Because this effort is part of a larger cross linguistic annotation, in order to maintain consistency with the other project, we started by applying the guidelines that were originally created for English annotations to the Chinese annotation. It worked well for purely semantic features such as modality type and environmental polarity, but difficulties arise when it comes to the features within the syntax-semantics interface such as span of prejacent.

In cases where no instructions are applicable, we added new specifications. In updating the guidelines, real examples are always included along with the rules. In what follows, we provide

* The term *dynamic* is not used in the schema. Instead the combination *ability_circumstantial* is adopted.

examples of the problems we have encountered and the treatments proposed.

Modals with A-not-A forms: In Chinese, a polarity question can be formed by alternating the main predicate of a sentence with its positive form (full form or the first syllable only) followed by its negative form. This kind of question is called an A-not-A question, and the form of the predicate is A-not-A form (Huang et al. 2009). Take (4) for example:

- (4) *Qingshaonian ke-bu-keyi xiyan*
juvenile may-not-may smoke
'May juveniles smoke or not?'

There are two possible annotations for A-not-A forms. One way is to treat an A-not-A form as one modal. The problem of this proposal is that if it is one markable, then the polarity of then sentence will be neither negative nor positive, and thus is illogical. The other solution is to divide A-not-A into two independent modals. The drawback of this approach is redundancy. Since the positive and negative modal will share the same set of features, the annotation is doubled.

After evaluating both approaches, we adopt the first, i.e. A-not-A is one modal, with a minor modification of the scheme, namely, adding a new value, A-not-A, to the *polarity* feature.

Potential complement: Potential complement construction is marked by the particle *de* (得), which appears “inside the so-called verb-result construction (*dongjie shi*) or verb-direction construction (*dongqu shi*)”. This construction “has a modal interpretation” (Xie 2012). The negation of a potential complement is formed by replacing *de* with the negation word *bu* (不). Compare (5) and (6):

- (5) *Zhangsan ban de qilai na xiang shu*^[1]
Zhangsan lift DE up that box(CL) book
'Zhangsan can lift up the box of books.'
- (6) *Zhangsan ban bu qilai na xiang shu*
Zhangsan lift not up that box(CL) book
'Zhangsan cannot lift up the box of books.'

Because the modal interpretation contributes to the whole construction rather than to individual

components, we treat the whole form “verb *de/bu* complement” as the modal that needs to be annotated. *Bu* is also marked as the item indicating the negative polarity of the sentence.

Relative clauses: The original guidelines for annotating modals in relative clauses specify that the relativizer should be included in the preajacent of the modal; the head noun that the relative clause modifies should not. Consider (7):

- (7) The person [*that we might see*] is John.

Relative clauses have a different structure in Chinese. They precede the head noun, and do not have a relativizer. Instead, there is a structural particle *de* which connects the relative clause to the nominal head.

We decided to exclude the particle *de* from the span of the preajacent, since it is not a part of the relative clause both syntactically and semantically. See (8):

- (8) *Gongsi jue ding jiang ke [yingli baiwan*
company decide *prep.* can profit million
yuan yishang] de yi zheng tao ruanjian dui
yuan above DE one whole set software to
yonghu kaifang.
users open

'The company decided to make open-source the whole set of software which can earn a profit of more than one million yuan.'

Temporal phrases: In many cases, a temporal phrase is contained in the syntactic scope of a modal's preajacent. Take (9) as an illustration:

- (9) [*Shanghai jin-nian quan nian chukou*] *keyi*
shanghai this year all year export may
[*chaoguo yi bai sis hi wu yi meiyuan.*]
exceed 14.5 billion dollar

'This year, the annual export of Shanghai may exceed 14.5 billion dollars.'

When marking the preajacent, we do not separate temporal phrases, whether they are inside or outside the scope of the modal. The reason for this is because singling out temporal phrases will make the preajacent more fragmented than necessary..

You (有)-X-modal-Y construction: As illustrated in (10), the modal-Y part expresses certain properties of X, and the verb *you* 'have'

^[1]The example is modified form Xie (2012), (1).

expresses the existence of X. For example, in *you-fa-ke-yi, ke-yi* ‘can-abide’ modifies the preceding noun *fa* ‘law’; and the whole phrase means ‘to have laws to abide by’. We did not annotate the modals in this construction, because there is no settled view about its syntactic analysis: It could be a productive morphological template, or it could be a case of postposed relative clause.

- (10) a. *shi women [you fa ke yi]*
 make we have law can abide
 ‘to make us have laws that we can abide by
 (have laws to abide by)’
- b. *Ta jintian [you gongzuo yao zuo]*
 he today have work need to do
 ‘He has work that he needs to do today.’

4 Results

As described in 2.1, we will have multiple cycles of evaluation and revision to control the quality of the annotation. This section presents the result of the first round of evaluation.

4.1 Measures of agreement

We calculated inter-annotator agreement on four features: *modality type*, *prejacent*, *background*, and *degree modifier*. Other features will be evaluated the next step. Two measures, κ score (Cohen 1960) and percentage of agreed instances, are provided. Also note that the annotated instances vary across features.

FEATURE	% OF AGREED	κ	ANNOTATED INSTANCES*
Modality type	62.3	0.522	406 (253)
Degree modifier	97.8	0.390	12 (3)
Background	84.7	0.349	86 (24)
Prejacent	66.5	N/A	406 (270)

Table 5: agreement: multiple features

We set a minimum threshold of 0.6 (Carletta 1996) for kappa scores for the purpose of quality control. As shown in Table 3, the kappa scores are all lower than the threshold. It indicates that all the four features are hard to annotate, and validates the

*The figures in the parentheses are the number of instances annotated by both annotators.

necessity of cyclic evaluation followed by revision.

Table 4 provides kappa scores on the agreement of modality type. It presents them before and after category collapsing per individual modal. Possible values of modality type are listed in Table 5.

Modal	κ all Cat.	κ Collapsed	#Tokens
dei	1.0	1.0	8
jiang	0.814	0.852	143
bixu	0.406	1.0	12
neng	0.352	0.632	27
hui	0.322	0.399	40
keneng	0.310	1.0	74
ke	0.283	0.633	39
nenggou	0.28	1.0	8
keyi	0.239	0.443	17
yao	0.099	0.289	48
yinggai	-0.256	1.0	6
overall	0.522	0.815	406

Table 6: agreement: modality type

COLLAPSED	ALL CATEGOREIS
<i>priority</i>	priority(p), deontic(d), bouletic(b), bouletic_teleological(b/t)
<i>non-priority</i>	epistemic(e), circumstantial(c), ability(a), ability_circumstantial(a/c), epistemic_circumstantial(a/c),
<i>not set</i>	not_set (n)
<i>to be decided</i>	to_be_decided (tbd)

Table 7: values and labels of modality types

The agreement scores before collapsing modality types is relatively low for most of the modals. The majority of the scores cluster around 0.3, while the extreme scores (=1.0 or < 0) are attested with low-occurrence modals. The kappa scores are generally improved after collapsing. However, some scores are still below the 0.6 threshold. We will discuss these cases in section 5.

Table 8 is the confusion matrix for modality type as marked by both annotators. We point out two obvious differences between Annotator A (columns) and Annotator B (rows):

- Among the priority types, Annotator A prefers *deontic* over *teleological* (64:13), while Annotator B does not have a strong

preference (34:30) between them.

- Annotator A selects the coarse type *epistemic_circumstantial* much more often than Annotator B (31 vs. 3). In cases where the *epistemic_circumstantial* is chosen by Annotator A, Annotator B tended to mark the type as *circumstantial* (23/31).

	e	a	c	e/c	a/c	d	b	t	b/t	p	n	tbd	S
e	5	0	5	3	0	0	0	0	0	0	1	0	14
a	1	9	0	0	2	0	0	0	0	0	0	0	12
c	6	1	89	23	8	3	0	0	0	0	4	0	134
e/c	0	0	0	1	1	0	0	0	0	0	1	0	3
a/c	0	10	3	3	10	1	0	0	0	0	0	0	27
d	0	0	2	0	0	31	0	1	0	0	0	0	34
b	0	0	0	0	0	0	0	0	0	0	0	0	0
t	0	0	0	0	0	16	1	3	9	0	1	0	30
b/t	0	0	0	0	0	3	0	5	1	0	1	0	10
p	0	0	0	0	0	2	0	2	0	0	1	0	5
n	0	2	18	1	2	8	0	2	0	0	104	0	137
tbd	0	0	0	0	0	0	0	0	0	0	0	0	0
S	12	22	117	31	23	64	1	13	10	0	113	0	406

Table 8: confusion matrix of modality type

5 Discussions

To better understand annotator disagreement, we retrieved the full list of sentences where any of the four features evaluated did not agree among the annotators. Roughly, these instances fall into four classes in terms of the reason for disagreement:

- Human error*: Some features are accidentally overlooked by one of the annotators, yielding disagreement.
- Vague guidelines*: Lack of specification in the guidelines causes divergent annotations. The low κ of the degree modifier feature can mainly be attributed to vagueness.
- Annotator's deviation from guidelines*: Annotators do not always correctly remember the individual instructions in the guidelines, and thus do not always follow

them. This type of disagreement is commonly found in the annotation of background and preajcent.

- High ambiguity of the target word*: we pointed out in the previous section that even with collapsed modality type, the κ scores of some modals are still relatively low. This is because they have a broader spectrum of meaning.

M	non-priority					non-priority					n	K
	e	a	c	e/c	a/c	d	b	t	b/t	p		
yao												.29 .10
hui												.40 .32
keyi												.44 .24
neng												.63 .35
ke												.63 .28
jiang												.85 .81
ying gai												1.0 -.3
yiding												1.0 -.2
neng gou												1.0 .28
ke neng												1.0 .31
bixu												1.0 .41
dei												1.0 1.0
ke wang												1.0 1.0

Table 9: semantic spectrums of modal inventory

Table 9 shows the distribution of modality types annotated for each expression. The cell is shaded if the corresponding type is chosen by at least one annotator. The darker grey cells are the majority types preferred by each annotator. In some cases the types preferred by the two annotators overlap,

but mostly they do not. From Table 9 we can generalize:

a). It is more difficult to achieve high agreement on the annotation of a modal’s flavor when the modal has many possible interpretations. This coincides with naïve intuitions.

b). If an item has both modal and non-modal usages (*yao*, *hui*, *jiang*, *ke*, *neng*), then it is likely that the annotation of the item will arrive at a low kappa score.

Take *yao* and *hui* for illustration. These two words have both a modal usage and some other usages. The word *yao* can be used as an attitude verb meaning ‘want to’. Similarly, *hui* can be used as a pure future marker without any obvious modal content. (11a-b) provides cases where one of the annotators marked the modality type of the target as *not_set*, i.e. not a modal expression. Table 10 summarizes how often the two tokens are marked as non-modal by each annotator.

- (11a) *Oumeng biaoshi yao jinyibu cujin*
 E.U. express YAO further promote
shuangfang zai gelyingyu de jiaoliu
 both parties in each area DE communication
- (i) ‘EU says that (it) is willing to further promote the communication between the two parties in various areas.’
- (ii) ‘EU says that the two parties need to further promote their communication in various areas.’
- (11b) *Ji nian nei, Xianggang hui chuxian*
 several year within, Hong Kong HUI appear
geng-duo de Zhongguo jijin
 more DE China fund
- (i) ‘There will be more funds from China in Hong Kong within several years.’
- (ii) ‘There can be more funds from China in Hong Kong within several years.’

	Non-modal meaning	# not_set by both	# not_set by A	# not_set by B
yao	'want to'	4	7	12
hui	future marker	8	8	20

Table 10: tokens marked as *not_set*

The effect of modal/non-modal distinction

seems to be more significant than the distinction between different flavors of modality. Yet this observation needs to be tested with a larger data set.

6 Conclusion

In this paper, we described our effort to annotate various aspects of modals in Penn Chinese Treebank, and reported the preliminary results of the first pass of annotation. The results show that it is hard for two annotators to achieve high agreement not only for modality type, but also for preajcent, background, and degree modifier. Therefore, multiple cycles of evaluation and revision are necessary for quality control. In effect, our project shows that, with minor adjustments, it is possible to use one scheme and set of guidelines for cross-linguistic annotation.

References

Carletta, J. 1996. Squibs and Discussions Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, 22(2), 249-254.

Cohen, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20 (1): 37-46.

Hacquard, V. 2011. Modality in *Semantics: An International Handbook of Natural Language Meaning*. HSK 33.2 ed. C. Maienborn, K. von Stechow and P. Portner, 1484-1515. Berlin: Mouton de Gruyter.

Hacquard, V. and Wellwood, A. 2012. Embedding epistemic modals in English: A corpus- based study. *Semantics and Pragmatics* 5:1-29.

Hendrickx, I.; Mendes, A.; and Mencarelli, S. 2012. Modality in text: A proposal for corpus annotation. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*.

Huang, C-T. J.; Li, Y-H. A.; and Li, Y. 2009. *The Syntax of Chinese*. Cambridge University Press.

Kratzer, A. 1981. The notional category of modality. In *In Words, worlds, and contexts*, ed. H.-J. Eikmeyer, H. Rieser, 38-74. Walter de Gruyter.

Müller, C.; and Strube, M. 2006. Multi-level annotation of linguistic data with MMAX2. *Corpus technology and language pedagogy: New resources, new tools, new method*, ed. S. Braun, K. Kohn, J. Mukherjee, 197-214. Peter Lang.

Portner, P. 2009. *Modality*. Oxford University Press.

- Rubinstein, A.; Simonson, D.; Chung, J.; Harner, H.; Katz, G.; and Portner, P. 2012. Developing a methodology for modality type annotations on a large scale. Slides of talk presented at the Modality Workshop, Ottawa University, April.
- Rubinstein, A.; Harner, H.; Krawczyk, E.; Simonson, D.; Katz, G.; and Portner, P. 2013. Toward Fine-grained Annotation of Modality in Text. In *Proceedings of the Tenth International Conference for Computational Semantics (IWCS 2013)*.
- Saurí, R.; Verhagen, M.; and Pustejovsky, J. 2006. Annotating and recognizing event modality in text. In *19th International FLAIRS conference (FLAIRS 2006)*.
- Wiebe, J.; Wilson, T.; and Cardie, C. 2005. Annotating expressions of opinions and emotions in language. *Language Resources and Evaluation*, 37(2-3):165-210.
- Xia, F.; Palmer, M.; Xue, N.; Okurowski, M. E.; Kovarik, J.; Chiou, F. D.; and Marcus, M. 2000. Developing guidelines and ensuring consistency for Chinese text annotation. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC-2000)*.
- Xie, Z. 2012. The modal uses of de and temporal shifting in Mandarin Chinese. *Journal of East Asian Linguistics* 21:387-420.
- Xue, N. and Palmer, M. 2003. Annotating the propositions in the Penn Chinese Treebank. In *Proceedings of the second SIGHAN workshop on Chinese language processing*. Volume 17 (pp. 47-54).
- Xue, N. 2005. Annotating discourse connectives in the chinese treebank. In *Proceedings of the Workshop on Frontiers in Corpus Annotations II: Pie in the Sky* (pp. 84-91).