

Annotating Event Mentions in Text with Modality, Focus, and Source Information

Suguru Matsuyoshi[†], Megumi Eguchi[†], Chitose Sao[†], Koji Murakami[†], Kentaro Inui^{†,‡} and Yuji Matsumoto[†]
[†]Nara Institute of Science and Technology, Japan [‡]Tohoku University, Japan

Introduction

arg. arg. arg.

John claimed that Ann **submitted** the plan to the committee.

An event mention "Ann submitting the plan to the committee"
 + **John's assertion that the event actually happened** Extended modality

The doctor speculated that mercury-based vaccines did **not cause** autism in children.

An event mention "mercury-based vaccines causing autism in children"
 + **the doctor's inference that the event does not happen**

Jim decided to stop **buying** that weekly magazine.

An event mention "Jim buying that weekly magazine"
 + **Jim's volition that the event will not happen in the future**

Abstract:

Many natural language processing tasks require analysis of the polarity, focus of polarity, tense, aspect, mood and source of the event mentions in a text in addition to its predicate-argument structure analysis. We refer to modality, polarity and other associated information as **extended modality**. In this paper, we propose a new annotation scheme for representing the extended modality of event mentions in a sentence. Our extended modality consists of the following seven components: **Source**, **Time**, **Conditional**, **Primary modality type**, **Actuality**, **Evaluation** and **Focus**. We also report on the current progress of our manual annotation of a Japanese corpus of about 50,000 event mentions, showing a reasonably high ratio of inter-annotator agreement.

Proposed scheme

We have the following four desiderata:

1. Information of the extended modality of an event mention should be gathered into one piece, specifically at the core predicate.
2. A system of extended modality should be language-independent.
3. The polarity of event mentions should be divided into two distinct classes: polarity on actuality and polarity from the view of the source's evaluation.
4. Labels in each component of extended modality should not be too fine-grained.

We reviewed the literature about extended modality in Linguistics and Natural Language Processing, and made an annotation scheme of extended modality for event mentions in a sentence. Our extended modality consists of the following seven components:



Component	Description
Source (S)	An agent or an organization that takes an attitude toward an event mention in a sentence.
Time (T)	Relative time of occurrence of a target event on the base of the time when the source took an attitude toward the event mention.
Conditional (C)	Whether a target event mention is a proposition with a condition.
Primary modality type (P)	A primary category that determines the fundamental meaning of an event mention.
Actuality (A)	Degree of certainty toward an event mention in text and transition of certainty toward it.
Evaluation (E)	Polarity toward a target event mention in a sentence from the view of the source's evaluation.
Focus (F)	The focus of negation, inference or interrogative.

Annotated corpus

Based on this scheme, we constructed an annotated corpus of 50,108 event mentions in Japanese.

Sentence and labels
hayaku ie-ni kaeri-tai-to Taro-ga itta. (Taro said he wanted to go home soon.) S=writer_Taro, T=future, C=notConditional, P=wish, A=unknown, E=positive, F=no
touyaku-wa sanhankikan-no kino-wo antei-sa-seru-tameni tsudukeru-noda-souda. (I hear that the medication is continued for regulating functions of three semicircular canals.) S=writer_arbitrary, T=notFuture, C=notConditional, P=assertion, A=certain+, E=neutral, F=no
ashita hare-tara, mizumi-ni sakana-tsuru-ni iko-to omou. (If it is nice out tomorrow, I will go fishing in that lake.) S=writer, T=future, C=hasCondition, P=volition, A=probable+, E=positive, F=no
kangensui-no motsu koka-wo jikkanshi-te-mi-te-kudasai! (Feel for yourself the effects of restoration water!) S=writer, T=future, C=notConditional, P=imperative-direct, A=unknown, E=positive, F=no
Taro-wa sono hamigakiko-wo tsukai hajime-ta-no-desu. (Taro began to use the toothpaste.) S=writer, T=notFuture, C=notConditional, P=assertion, A=certain-→+, E=neutral, F=no
anata-wa sono toki-ni kanojo-ni shinjitsu-wo tsutaeru-beki-dat-ta. (You should have told the truth to her at that time.) S=writer, T=notFuture, C=notConditional, P=assertion, A=certain-, E=positive, F=no
kare-wa kimi-no-tame-ni nokot-ta-no-de-wa-nai. (It was not for you that he stayed.) S=writer, T=notFuture, C=notConditional, P=assertion, A=certain-, E=neutral, F=negation(for you)

Distribution of labels of each component of extended modality in our corpus.

Source	Time	Conditional
writer 49,609	future 5,749	condition 2,395
writer_other 132	notFuture 44,359	hasCondition 509
writer_arbitrary 196		notConditional 47,204
writer_STRING 171		

Primary modality type	Actuality	Evaluation
assertion 45,501	certain+ 40,793	positive 3,021
volition 1,106	certain- 2,693	negative 446
wish 637	certain-→+ 68	neutral 46,641
imperative-direct 581	certain+→- 34	
imperative-indirect 850	probable+ 2,529	
imperative-together 73	probable- 343	
permission 28	probable-→+ 20	
interrogative 1,332	probable+→- 15	
	unknown 3,613	

Focus
negation 21
inference 81
interrogative 356
no 49,650

	S	T	C	P	A	E	F
k statistics	0.69	0.76	0.68	0.66	0.70	0.72	0.75

Related work

	certainty	transition of certainty	evaluation	modality in a narrow sense	polarity	focus	source	time	conditional
Light et al. (2004)	○								
Rubin et al. (2005)	○			○			○	○	
Sauri et al. (2006)	○	○		○	○			○	○
Prasad et al. (2006)	○			○	○	○	○		
Sauri and Pustejovsky (2007)	○				○				
Medlock and Briscoe (2007)	○								
Szarvas et al. (2008)	○				○				
Sauri (2008)	○	○		○	○		○	○	○
Hara and Inui (2008)	○	○		○	○		○	○	○
Kawazoe et al. (2009)	○				○		○		○
Im et al. (2009)	○	○		○	○			○	○
Our work	○	○	○	○	○	○	○	○	○

Future work

- To apply our annotation scheme to different kinds of resources for assessing the scheme and extending our corpus.
- To implement an analyzer of our extended modality with high precision, using machine learning approaches, such as support vector machines and conditional random fields, and the corpus as the training data.

Acknowledgment

This work was partly supported by Japan MEXT Grant-in-Aid for Young Scientists (Start-up) (No. 20800029) and Grant-in-Aid for Scientific Research on Priority Areas (No. 21013036), and by National Institute of Information and Communications Technology Japan.