Excitatory or Inhibitory: A New Semantic Orientation Extracts Contradiction and Causality from the Web

Chikara Hashimoto[*]   **Kentaro Torisawa**[†]   **Stijn De Saeger**[‡]
Jong-Hoon Oh[§]   **Jun'ichi Kazama**[¶]
National Institute of Information and Communications Technology
Kyoto, 619-0289, JAPAN
{[*]ch, [†]torisawa, [‡]stijn, [§]rovellia, [¶]kazama}@nict.go.jp

Abstract

We propose a new semantic orientation, Excitation, and its automatic acquisition method. Excitation is a semantic property of predicates that classifies them into excitatory, inhibitory and neutral. We show that Excitation is useful for extracting contradiction pairs (e.g., *destroy cancer* ⊥ *develop cancer*) and causality pairs (e.g., *increase in crime* ⇒ *heighten anxiety*). Our experiments show that with automatically acquired Excitation knowledge we can extract one million contradiction pairs and 500,000 causality pairs with about 70% precision from a 600 million page Web corpus. Furthermore, by combining these extracted causality and contradiction pairs, we can generate one million plausible causality hypotheses that are not written in any single sentence in our corpus with reasonable precision.

1 Introduction

Recognizing semantic relations between events in texts is crucial for such NLP tasks as question answering (QA). For example, to answer the question *"What ruined the crops in Japan?"* a QA system must recognize that the sentence *"the Fukushima nuclear power plant caused radioactive pollution and contaminated the crops in Japan"* contains a causal relation and that *contaminate crops* entails *ruin crops* but contradicts *preserve crops*.

To facilitate the acquisition of causality, contradiction, paraphrase and entailment relations between events we propose a new semantic orientation, Excitation, that classifies unary predicates (templates, hereafter) into excitatory, inhibitory and neutral. An excitatory template entails that the main function or

effect of the referent of its argument is activated or enhanced (e.g., *cause X*, *preserve X*), while an inhibitory template entails that it is deactivated or suppressed (e.g., *ruin X*, *contaminate X*, *prevent X*).

Excitation is useful for extracting contradiction; if two templates with similar distributional profiles have opposite Excitation polarities, they tend to be contradictions (e.g., <u>*contaminate crops*</u> and <u>*preserve crops*</u>). With extracted contradictions we can distinguish paraphrases from contradictions among distributionally similar phrases. Furthermore, contradiction in itself is important knowledge for Recognizing Textual Entailment (RTE) (Voorhees, 2008).

Excitation is also a powerful indicator of causality. In the physical world, the activation or deactivation of one thing often causes the activation or deactivation of another. Two excitatory or inhibitory templates that co-occur in some temporal or logical order in the same narrative often describe a causal chain of events, like *"the Fukushima nuclear power plant <u>caused</u> radioactive pollution and <u>contaminated</u> crops in Japan"*.

In this paper we propose both the concept of Excitation and an automatic method for its acquisition. Our method acquires Excitation templates based on certain natural, language independent constraints on narrative structures found in text. We also propose acquisition methods for contradiction and causality relations based on Excitation. Our methods extract one million contradiction pairs with over 70% precision, and 500,000 causality pairs with about 70% precision from a 600 million page Web corpus. Moreover, by combining these extracted causality pairs and contradiction pairs, we generated one million plausible causality hypotheses that were not

619

written in any single sentence in our corpus with reasonable precision. For example, a causality hypothesis *prevent radioactive pollution ⇒ preserve crops* can be generated from an extracted causality *cause radioactive pollution ⇒ contaminate crops*.

We target the Japanese language in this paper.

2 What is Excitation?

Excitation classifies templates into excitatory, inhibitory, and neutral, as explained below.

excitatory templates entail that the function, effect, purpose or role of their argument's referent is activated or enhanced. (e.g., *cause X*, *buy X*, *produce X*, *import X*, *increase X*, *enable X*)

inhibitory templates entail that the function, effect, purpose or role of their argument's referent is deactivated or suppressed. (e.g., *prevent X*, *discard X*, *remedy X*, *decrease X*, *disable X*)

neutral templates are neither excitatory nor inhibitory. (e.g., *consider X*, *proportional to X*, *related to X*, *evaluate X*, *close to X*)

For example, when *fire* fills the *X* slot of *cause X*, it suggests that the effect of *fire* is activated. If *prevent X*'s slot is filled with *flu*, the effect of *flu* is suppressed. In this study, we aim to acquire excitatory and inhibitory templates that are useful for extracting contradiction and causality, though neutral templates are the most frequent in our data (See Section 5.1). Collectively we call excitatory and inhibitory templates *Excitation templates*, and excitatory and inhibitory two opposite *polarities*.

Excitation is independent of the good/bad semantic orientation. (Hatzivassiloglou and McKeown, 1997; Turney, 2002; Rao and Ravichandran, 2009). For example, *sophisticate X* and *complicate X* are both excitatory, but only the former has a positive connotation. Similarly, *remedy X* and *degrade X* are both inhibitory but only the latter is negative.

General Inquirer (Stone et al., 1966) deals with semantic factors some of which were proposed by Osgood et al. (1957). Their 'activity' factor involves binary opposition between 'active' and 'passive.' Notice that activity and Excitation are independent. In General Inquirer, both *accelerate X* and *abolish X* are active, but only the former is excitatory. Both *accept X* and *abate X* are passive, but only the latter is inhibitory. Pustejovsky (1995) proposed telic

and agentive roles, which inspired our excitatory notion, but they have no corresponding notion of inhibitory. Andreevskaia and Bergler (2006) acquired the increase/decrease semantic orientation, which is a subclass of Excitation.

Excitation is inverted if a template's predicate is negated. For example, *preserve X* is excitatory, while *don't preserve X* is inhibitory. We acknowledge that this may seem somewhat counter-intuitive and will address this issue in future work.

3 Excitation Template Acquisition

This section presents our acquisition method of Excitation templates. We introduce constraints in the co-occurrence of templates in text that seem both robust and language independent in Section 3.1. Our method exploits these constraints for the acquisition of Excitation templates. First we construct a template network where nodes are templates and links represent that two connected templates have either SAME or OPPOSITE polarities. Given 46 manually prepared seed templates we calculate the *Excitation value* of each template, a value in range $[-1, 1]$ that is positive if the template is excitatory and negative if it is inhibitory. Technically, our method treats all templates as excitatory or inhibitory, and, upon completion, regards templates with small absolute Excitation values as neutral.

The whole method is a bootstrapping process. Each iteration expands the network and the Excitation value of each template is (re-)calculated.

3.1 Characteristics of Excitation Templates

Our method exploits natural discourse constraints on the possible combinations of (a) the polarity of co-occurring templates, (b) the nouns that fill their argument slots and (c) the connectives that link the templates in a given sentence. Table 1 shows the constraints and Figure 1 shows examples that will be explained shortly. Though our target is Japanese we believe these constraints are universal discourse principles, and as such not language dependent. Examples are given in English for ease of explanation.

We first identify two categories of connectives in our target sentences: AND/THUS-type (e.g., *and*, *thus* and *since*) and BUT-type (e.g., *but* and *though*). Both types suggest a sort of *consistency* or *inconsistency* between predicates. We manually classified 169 frequently used connectives into AND/THUS-

620

(1) He smoked cigarettes, AND/THUS he suffered lung cancer. (Both *smoke X* and *suffer X* are excitatory.)

(2) He quit cigarettes, AND/THUS was immune from lung cancer. (*quit X* and *immune from X* are inhibitory.)

(3) He smoked cigarettes, BUT didn't suffer lung cancer. (*smoke X* is excitatory, *not suffer X* is inhibitory.)

(4) He quit cigarettes, BUT he suffered lung cancer. (*quit X* is inhibitory, but *suffer X* is excitatory.)

(5) He underwent cancer treatment, AND/THUS he could cure the cancer. (*undergo X* is excitatory, *cure X* is inhibitory.)

(6) He underwent cancer treatment, BUT still had cancer. (Both *undergo X* and *have X* are excitatory.)

(7) **Unnatural:** He smoked cigarettes, BUT he suffered lung cancer. (*smoke X* and *suffer X* are excitatory.)

Figure 1: Examples of constraints: (*cigarettes*, *lung cancer*) is PNP and (*cancer treatment*, *cancer*) is NNP.

	PNPs	NNPs	others
AND/THUS	SAME	OPPOSITE	N/A
BUT	OPPOSITE	SAME	N/A

Table 1: Constraint matrix.

and BUT-type (See supplementary materials).

Next we extract sentences from the Web in which two templates co-occur and are joined by one of these connectives, and then classify the noun pairs filling the templates' argument slots into "positively-associated" and "negatively-associated" noun pairs (PNPs and NNPs). Mirroring our definition of Excitation, PNPs are noun pairs in which the referent of the first noun facilitates the emergence of the referent of the second noun. PNPs can range from causally related noun pairs like (*cigarettes*, *lung cancer*) to "material-product" relation pairs like (*semiconductor*, *electronic circuit*). We found that PNPs only fill the argument slots of (a) same Excitation polarity templates connected by AND/THUS-type connectives (examples 1 and 2 in Figure 1), or (b) opposite Excitation polarity templates connected by a BUT-type connectives (examples 3 and 4). Violating such constraints (example 7) seems unnatural. Similarly, NNPs are noun pairs in which the referent of one noun suppresses the emergence of the referent of the other noun. Examples include such "inverse causality" pairs as (*cancer treatment*, *cancer*). NNPs only fill the argument slots of (a) opposite Excitation polarity templates connected by AND/THUS-type connectives (example 5), or (b) same polarity templates connected by a BUT-type connective (example 6).

All these constraints are summarized in Table 1,

which we will call the constraint matrix. According to the constraint matrix, we can know whether two templates' polarities are the same or opposite if we know whether a noun pair filling the two templates' slots is PNP or NNP. Conversely, we can know whether a noun pair is PNP or NNP if we know whether two templates whose slots are filled with the noun pair have the same or opposite polarities. We believe these constraints capture certain universal principles of discourse, since it is difficult in any language to produce natural sounding sentences that violate these constraints. We empirically confirm their validity for Japanese in Section 5.1.

3.2 Bootstrapping Approach to Excitation Template Acquisition

To calculate the Excitation values for the templates, we construct a template network where templates are connected by links indicating polarity agreement between two connected templates (either SAME or OPPOSITE polarity), as determined by the constraint matrix. Excitation values are determined by spreading activation applied to the network, given a small number of manually prepared seed templates.

However, we cannot construct the network unless we know whether each noun pair is PNP or NNP, due to the configuration of the constraint matrix, and currently we have no feasible method to classify all of them into PNPs and NNPs in advance. We therefore adopt a bootstrapping method (Figure 2) that starts from manually prepared excitatory and inhibitory seed templates (Step 1 in Figure 2). Our method begins by extracting noun pairs from the Web that co-occur with two seed templates connected by a AND/THUS- or BUT-type connective, and classifies these noun pairs into PNPs and NNPs based on the constraint matrix (Steps 2 and 3). Next, we automatically extract additional (non-seed) template pairs from the Web that co-occur with these PNPs and NNPs. Links (either SAME or OPPOSITE) between all template pairs are determined by the constraint matrix (Step 4), and we construct a template network from both seed and non-seed template pairs (Step 5).

Our method calculates the Excitation values for all the templates in the network by first assigning Excitation values $+1$ and -1 to the excitatory and inhibitory seed templates, and applies a spreading activation method proposed by Takamura et al. (2005) (Step 6) to the network. This method calcu-

1. Prepare initial seed templates with fixed excitation values (either +1 or −1).
2. Make seed template pairs that are combinations of two seed templates and a connective (either AND/THUS-type or BUT-type).
3. Extract noun pairs that co-occur with one of the seed template pairs from the Web. Classify the noun pairs into PNPs and NNPs based on the constraints matrix. Filter out those noun pairs that appear as both PNP and NNP on the Web or those whose occurrence frequency is less than or equal to F, which is set to 5.
4. Extract additional (non-seed) template pairs that are filled by one of the PNPs or NNPs from the Web. Determine the link type (SAME or OPPOSITE) for each template pair based on the constraint matrix. If a template pair appears on the Web as having both link types, we determine its link type by majority vote.
5. Construct the template network from all the template pairs. Remove from the network those templates whose number of linked templates is less than D, which is set to 5.
6. Apply Takamura et al.'s method to the network and fix the Excitation value of each template.
7. Extract the top- and bottom-ranked $N \times i$ templates from the result of Takamura et al.'s method. N is a constant, which is set to 30. i is the iteration number. They are used as additional seed templates for the next iteration. The top-ranked templates are given Excitation value +1 and the bottom-ranked templates are assigned −1. Go to Step 2.

Figure 2: Bootstrapping for template acquisition.

lates all templates' excitation values by solving the network constraints imposed by the SAME and OPPOSITE links, and the Excitation values of the seed templates (This method is detailed in Section 3.3). In each iteration i, our method selects the $N \times i$ top-ranked and bottom-ranked templates as additional seed templates for the next iteration (N is set to 30) (Step 7). Our method then constructs a new template network using the augmented seed templates and restarts the calculation process. Figure 2 summarizes our bootstrapping process.

Bootstrapping stops after M iterations, with M set to 7 based on our preliminary experiments.

To prepare the initial seed templates we constructed a *maximal* template network that could in theory be created by our bootstrapping method. This maximal network consists of any two templates that co-occur in a sentence with any connective, regardless of their arguments. We manually selected 36 excitatory and 10 inhibitory seed templates from among 114 templates with the most links in the network (See supplementary materials).

3.3 Determining Excitation in the Network

This section details Step 6 of our bootstrapping method, i.e., how Takamura et al.'s method calculates the Excitation value of each template. Their method is based on the spin model in physics, where each electron has a spin of either *up* or *down*. We

chose this method due to the straightforward parallel between the spin model and our Excitation template model. Both models capture the spreading of activation (either spin direction or excitation polarity) between neighboring objects in a network. Determining the optimal algorithm for this task is beyond our current scope, but for the purpose of our experiments we found that Takamura et al.'s method gave satisfactory results.

The spin model defines an energy function on a spin network, and each electron's spin can be estimated by minimizing this function:

$$E(\mathbf{x}, W) = -1/2 \times \Sigma_{ij} w_{ij} x_i x_j$$

Here, $x_i, x_j \in \mathbf{x}$ are spins of electrons i and j, and matrix $W = \{w_{ij}\}$ assigns weights to links between electrons. We regard templates as electrons and Excitation polarities as their spins (*up* and *down* correspond to excitatory and inhibitory). We define the weight w_{ij} of the link between templates i and j as:

$$w_{ij} = \begin{cases} 1/\sqrt{d(i)d(j)} & \text{if SAME}(i,j) \\ -1/\sqrt{d(i)d(j)} & \text{if OPPOSITE}(i,j) \end{cases}$$

Here, $d(i)$ denotes the number of templates linked to i. SAME(i,j) (OPPOSITE(i,j)) indicates a SAME (OPPOSITE) link exists between i and j. We obtain excitation values by minimizing the above energy function. Note that after minimizing E, x_i and x_j tend to get the same polarity when w_{ij} is positive. When w_{ij} is negative, x_i and x_j tend to have opposite polarities. Initially seed templates are given values +1 or −1 depending on whether they are excitatory or inhibitory, and others are given 0.

We used SUPPIN (http://www.lr.pi.titech.ac.jp/~takamura/pubs/SUPPIN-0.01.tar.gz), an implementation of Takamura et al.'s method. Its parameter β is set to the default value (0.75).

4 Knowledge Acquisition by Excitation

This section shows how the concept of Excitation can be used for automatic knowledge acquisition.

4.1 Contradiction Extraction

Our first knowledge acquisition method extracts contradiction pairs like *destroy cancer* ⊥ *develop cancer*, based on our assumption that they often consist of distributionally similar templates that have a sharp contrast in Excitation value. Concretely, we

extract two phrases as a contradiction pair if (a) their templates have opposite Excitation polarities, (b) they share the same argument noun, and (c) the part-of-speech of their predicates is the same. Then the contradiction pairs are ranked by Ct:

$$Ct(p_1, p_2) = |s_1| \times |s_2| \times sim(t_1, t_2)$$

Here p_1 and p_2 are two phrases that satisfy conditions (a), (b) and (c) above, t_1 and t_2 are their respective templates, and $|s_1|$ and $|s_2|$ are the absolute values of t_1 and t_2's excitation values. $sim(t_1, t_2)$ is the distributional similarity proposed by Lin (1998).

Note that "contradiction" here includes what we call "quasi-contradiction." This consists of two phrases such that, if the tendencies of the events they describe get stronger, they eventually become contradictions. For example, the pair *emit smells $\perp$ reduce smells* is not logically contradictory since the two events can happen at the same time. However, they become almost contradictory when their tendencies get stronger (i.e., *emit smells more strongly $\perp$ thoroughly reduce smells*). We believe quasi-contradictions are useful for NLP tasks.

4.2 Causality Extraction

Our second knowledge acquisition method extracts causality pairs like *increase in crime $\Rightarrow$ heighten anxiety* that co-occur with AND/THUS-type connectives in a sentence. The assumption is that if two templates (t_1 and t_2) with a strong Excitation tendency are connected by an AND/THUS-type connective in a sentence, the event described by t_1 and its argument n_1 tends to be a cause of the event described by t_2 and its argument n_2. Here, Excitation strength is expressed by absolute Excitation values. The intuition is that, if the referent of n_1 is strongly activated or suppressed, it tends to have some causal effect on the referent of n_2 in the same sentence.

We focus on extracting causality pairs that co-occur with only "non-causal connectives" like *and*, which are AND/THUS-type connectives that do NOT explicitly signal causality, since causal connectives like *thus* can mask the effectiveness of Excitation. We prepared 139 non-causal connectives (See supplementary materials). We extract two templates such as *increase in X* and *heighten Y* co-occurring with only non-causal connectives, as well as the noun pair that fills the two templates' slots (e.g., (*crime*, *anxiety*)) to obtain causal phrase pairs. In

Japanese, the temporal order between events is usually determined by precedence in the sentence. Cs ranks the obtained causality pairs:

$$Cs(p_1, p_2) = |s_1| \times |s_2|$$

Here p_1 and p_2 are the phrases of causality pair, and $|s_1|$ and $|s_2|$ are absolute Excitation values of p_1's and p_2's templates. As is common in the literature, this notion of causality should be interpreted probabilistically rather than logically, i.e., we interpret causality $A \Rightarrow B$ as "*if A happens, the probability of B increases*". This interpretation is often more useful for NLP tasks than a strict logical interpretation.

4.3 Causality Hypothesis Generation

Our third knowledge acquisition method generates plausible causality hypotheses that are not written in any single sentence using the previously extracted contradiction and causality pairs. We assume that if a causal relation (e.g., *increase in crime $\Rightarrow$ heighten anxiety*) is valid, its inverse (e.g., *decrease in crime $\Rightarrow$ diminish anxiety*) is often valid as well. From a *logical* definition of causation, taking the inverse of an implication obviously does not preserve validity. However, at least under our probabilistic interpretation, taking the inverse of a given causality pair using the extracted contradiction pairs proves to be a viable strategy for generating non-trivial causality hypotheses, as our experiments in Section 5.4 show.

For an extracted causality pair, we generate its inverse as a causality hypothesis by replacing both phrases in the original pair with their contradiction counterparts. For instance, a causality hypothesis *decrease in crime $\Rightarrow$ diminish anxiety* is generated from a causality *increase in crime $\Rightarrow$ heighten anxiety* by two contradictions, *decrease in crime $\perp$ increase in crime* and *diminish anxiety $\perp$ heighten anxiety*. Since we are interested in finding *new* causal hypotheses, we filter out hypotheses whose phrase pair co-occurs in a sentence in our corpus. Remaining causality hypotheses are ranked by Hp.

$$Hp(q_1, q_2) = Ct(p_1, q_1) \times Ct(p_2, q_2) \times Cs'(p_1, p_2)$$

Here, q_1 and q_2 are two phrases of a causality hypothesis. p_1 and p_2 are two phrases of a hypothesis's original causality. That is, $p_1 \perp q_1$ and $p_2 \perp q_2$ are contradiction pairs, and $Ct(p_1, q_1)$ and $Ct(p_2, q_2)$ are their contradiction scores. $Cs'(p_1, p_2)$ is the original causality's causality score. Cs' can be Cs

from Section 4.2, but based on preliminary experiments we found the following score works better:

$$Cs'(p_1, p_2) = |s_1| \times |s_2| \times npfreq(n_1, n_2)$$

$|s_1|$ and $|s_2|$ are absolute Excitation values of p_1's and p_2's templates, whose slots are filled with n_1 and n_2. $npfreq(n_1, n_2)$ is the co-occurrence frequency of (n_1, n_2) with polarity-identical template pairs (if (n_1, n_2) is PNP) or with polarity-opposite template pairs (if (n_1, n_2) is NNP). Thus, $npfreq$ indicates a sort of association strength between two nouns.

5 Experiments

This section shows that our template acquisition method acquired many Excitation templates. Moreover, using only the acquired templates we extracted one million contradiction pairs with more than 70% precision, and 500,000 causality pairs with about 70% precision. Further, using only these extracted contradiction and causality pairs we generated one million causality hypotheses with 57% precision.

In our experiments we removed evaluation samples containing the initial seed templates and examples used for annotation instruction from the evaluation data. Three annotators (not the authors) marked all evaluation samples, which were randomly shuffled so that they could not identify which sample was produced by which method. Information about the predicted labels or ranks was also removed from the evaluation data. Final judgments were made by majority vote between the annotators. They were non-experts without formal training in linguistics or semantics. See supplementary materials for our annotation manuals (translated into English).

We used 600 million Japanese Web pages (Akamine et al., 2010) parsed by KNP (Kawahara and Kurohashi, 2006) as a corpus. We restricted the argument positions of templates to *ha* (topic), *ga* (nominative), *wo* (accusative), *ni* (dative), and *de* (instrumental). We discarded templates appearing fewer than 20 times in compound sentences (regardless of connectives) in our corpus.

5.1 Excitation Template Acquisition

We show that our proposed method for template extraction (PROP$_{tmp}$) successfully acquired many Excitation templates from which we obtained a huge number of contradiction and causality pairs, and that Excitation is a reasonably comprehensible notion

even for non-experts. We also show that PROP$_{tmp}$ outperformed two baselines by a large margin.

The template network constructed by PROP$_{tmp}$ contained 10,825 templates. Among these, the bootstrapping process classified 8,685 templates as excitatory and 2,140 as inhibitory. Note that these candidates in fact also contain neutral templates, as explained at the beginning of Section 3.

Baselines The baseline methods are ALLEXC and SIM. ALLEXC regards all templates that are randomly extracted from the Web as excitatory, since in our data excitatory templates outnumber inhibitory ones. Actually, in our data neutral templates represent the most frequent class, but since our objective is to acquire excitatory and inhibitory templates, a baseline marking all templates as neutral would make little sense. SIM is a distributional similarity baseline that takes as input the same 10,825 templates of PROP$_{tmp}$ above, constructs a network by connecting two templates whose distributional similarity is greater than zero, and regards two connected templates as having the same polarity. The weight of the links between templates is set to their distributional similarity based on Lin (1998). Then SIM is given the same initial seed templates as PROP$_{tmp}$, by which it calculates the Excitation values of templates using Takamura et al.'s method. As a result, SIM assigned positive Excitation values to all templates, and except for the 10 inhibitory initial seed templates no templates were regarded inhibitory.

Evaluation scheme We randomly sampled 100 templates each from PROP$_{tmp}$'s 8,685 excitatory candidates, PROP$_{tmp}$'s 2,140 inhibitory candidates, all the ALLEXC's templates, and all the SIM's templates, i.e., 400 templates in total. To make the annotators' judgements easier, we randomly filled the argument slot of each template with a noun filling its argument slot in our Web corpus. Three annotators labeled each sample (a combination of a template and a noun) as 'excitatory,' 'inhibitory,' 'neutral,' or 'undecided' if they were not sure about its label.

Results for excitatory In the top graph in Figure 3, 'Proposed' shows PROP$_{tmp}$'s precision curve. The curve is drawn from its 100 samples whose X-axis positions represent their ranks. We plot a dot for every 5 samples. Among the 100 samples, 37 were judged as excitatory, 6 as inhibitory, 45 as neutral, and 6 as 'undecided'. For the remaining 6 samples,
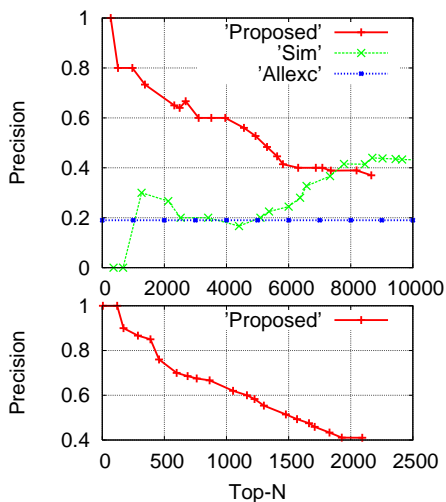
Figure 3: Precision of template acquisition: excitatory (top) and inhibitory (bottom).

the three annotators gave three different labels and the label was not fixed ('split-votes' hereafter). For calculating precision, only the 37 samples labeled excitatory were regarded as correct. PROP_{tmp} out-performed all baselines by a large margin, with an estimated 70% precision for the top 2,000 templates. 'Allexc' and 'Sim' in Figure 3 denote ALLEXC and SIM. Among ALLEXC's 100 samples, 19 were judged as excitatory, 5 as inhibitory, 74 as neutral, and 2 as 'undecided'. SIM's low performance reflects the fact that templates with opposite polarities are sometimes distributionally similar, and as a result get connected by SAME links.

Results for inhibitory 'Proposed' in the bottom graph in Figure 3 shows the precision curve drawn from the 100 samples of PROP_{tmp}'s inhibitory candidates. Among the 100 samples, 41 were judged as inhibitory, 15 as excitatory, 32 as neutral, 4 as 'unde-cided', and 8 as 'split-votes'. Only the 41 inhibitory samples were regarded as correct. From the curve we estimate that PROP_{tmp} achieved about 70% precision for the top 500. Note that SIM could not acquire any inhibitory templates, yet we can think of no other reasonable baseline for this task.

Inter-annotator agreement The Fleiss' kappa (Fleiss, 1971) of annotator judgements was 0.48 (moderate agreement (Landis and Koch, 1977)). For training, the annotators were given a one-page anno-tation manual (see supplementary materials), which basically described the same contents in Section 2,

in addition to 14 examples of excitatory, 14 exam-ples of inhibitory, and 6 examples of neutral tem-plates that were manually prepared by the authors. Using the manual and the examples, we instructed all the annotators face-to-face for a few hours. We also made sure the evaluation data did not contain any examples used during instruction.

Observations about argument positions Among the 200 evaluation samples of PROP_{tmp} (for both ex-citatory and inhibitory evaluations), 52 were judged as excitatory, 47 as inhibitory, and 77 as neutral. For the excitatory templates, the numbers of nominative, topic, accusative, dative, and instrumental argument positions are 15, 11, 10, 8, and 8, respectively. For the inhibitory templates, the numbers are 17, 11, 16, 3, and 0. For the neutral templates, the numbers are 8, 23, 17, 21, and 8. Accordingly, we found no no-ticeable bias with regard to their numbers. Likewise, we found no noticeable bias regarding their useful-ness for contradiction and causality acquisition re-ported shortly, too.

Summary PROP_{tmp} works well, as it outperforms the baselines. Its performance demonstrates the va-lidity of our constraint matrix in Table 1. Besides, since our annotators were non-experts but showed moderate agreement, we conclude that Excitation is a reasonably comprehensible notion.

5.2 Contradiction Extraction

This section shows that our proposed method for contradiction extraction (PROP_{cont}) extracted one million contradiction pairs with more than 70% pre-cision, and that Excitation values are useful for con-tradiction ranking. As input for PROP_{cont} we took the top 2,000 excitatory and the top 500 inhibitory templates from the previous experiment (i.e., the other templates were regarded as neutral).

Baselines Our baseline methods are RAND_{cont} and PROP_{cont}-NE. RAND_{cont} randomly combines two phrases, each consisting of a template and a noun that they share. It does not rank its output. PROP_{cont}-NE is the same as PROP_{cont} except that it does not use Excitation values; ranking is based only on $sim(t_1, t_2)$. PROP_{cont}-NE does combine phrases with opposite template polarities, just like PROP_{cont}.

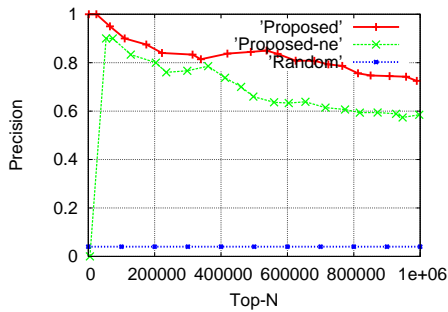**Evaluation scheme** We randomly sampled 200 phrase pairs from the top one million results of each

Figure 4: Precision of contradiction extraction.

PROP$_{cont}$ and PROP$_{cont}$-NE, and 100 samples from the output of RAND$_{cont}$'s output, giving 500 samples. Three annotators labeled whether the samples are contradictions. Fleiss' kappa was 0.78 (substantial agreement).

Results 'Proposed' in Figure 4 shows the precision curve of PROP$_{cont}$. PROP$_{cont}$ achieved an estimated 70% precision for its top one million results. Readers might wonder whether PROP$_{cont}$'s output consists of a small number of template pairs that are filled with many different nouns. If this were the case, PROP$_{cont}$'s performance would be somewhat misleading. However, we found that PROP$_{cont}$'s 200 samples contained 194 different template pairs, suggesting that our method can acquire a large variety of contradiction phrases. 'Proposed-ne' is the precision curve for PROP$_{cont}$-NE. Its precision is more than 10% lower than PROP$_{cont}$ at the top one million results. 'Random' shows that RAND$_{cont}$'s precision is only 4%. Table 2 shows examples of PROP$_{cont}$'s outputs and their English translation. The labels 'Cont,' 'Quasi' and '✖' denote whether a pair is contradictory, quasi-contradictory, or not contradictory. Among PROP$_{cont}$'s 145 samples judged by the annotators as contradiction, 46 were judged as quasi-contradictory by one of the authors. The first ✖ case in Table 2 was caused by the template, X を改善する (*improve X*). It is tricky since it is excitatory when taking arguments like *function*, while it is inhibitory when taking arguments like *disorder*. However, PROP$_{tmp}$ currently cannot distinguish these usages and judged it as inhibitory in our experiments in Section 5.1, though it must be interpreted as excitatory for the ✖ case. The second ✖ case was due to PROP$_{tmp}$'s error; it incorrectly judged the neutral template, X が関係する (*related to X*), as inhibitory.

Rank	Contradiction Pairs	Label
8,767	アンバランスを是正する ⊥ アンバランスを生じさせる *repair imbalance ⊥ become imbalanced*	Cont
103,581	運転を助ける ⊥ 運転を妨げる *assist the driver ⊥ disturb the driver*	Cont
151,338	緊張感が緩和される ⊥ 緊張を伴う *calm tension ⊥ feel tension*	Quasi
184,014	機能を改善する ⊥ 機能を高める *improve function ⊥ boost function*	✖
316,881	円安が止まる ⊥ 円安が進行する *yen depreciation stops ⊥ yen depreciation develops*	Cont
317,028	騒音がひどくなる ⊥ 騒音は減少する *noise gets worse ⊥ noise abates*	Cont
334,642	酸味が増す ⊥ 酸味が消える *a sour taste is augmented ⊥ a sour taste is lost*	Cont
487,496	痛みが発症する ⊥ 痛みを減らす *feel pain ⊥ reduce pain*	Quasi
529,173	アクセスが生ずる ⊥ アクセスを抑制する *access occurs ⊥ curb access*	Cont
555,049	原発をなくす ⊥ 原発を増やす *lose nuclear plants ⊥ augment nuclear plants*	Cont
608,895	放射能が放出される ⊥ 放射能が減る *radioactivity is released ⊥ radioactivity is reduced*	Quasi
638,092	ユーロが下落する ⊥ ユーロが強くなる *Euro falls ⊥ Euro gets strong*	Cont
757,423	シェアを有する ⊥ シェアが低下する *have share (in market) ⊥ share decreases*	Quasi
833,941	活性酸素が放出される ⊥ 活性酸素が関係する *generate active oxygen ⊥ related to active oxygen*	✖
848,331	ガンを破壊する ⊥ ガンを進行させる *destroy cancer ⊥ develop cancer*	Cont
982,980	ウイルスが死滅する ⊥ ウイルスが活性化する *virus becomes extinct ⊥ virus is activated*	Cont

Table 2: Examples of PROP$_{cont}$'s outputs.

Summary PROP$_{cont}$ is a low cost but high performance method, since it acquired one million contradiction pairs with over 70% precision from only the 46 initial seed templates. Besides, Excitation contributes to contradiction ranking since PROP$_{cont}$ outperformed PROP$_{cont}$-NE by a 10% margin for the top one million results. Thus we conclude that our assumption on contradiction extraction is valid.

5.3 Causality Extraction

We show that our method for causality extraction (PROP$_{caus}$) extracted 500,000 causality pairs with about 70% precision, and that Excitation values contribute to the ranking of causal pairs. PROP$_{caus}$ took as input all 10,825 templates classified by PROP$_{tmp}$.

Baselines RAND$_{caus}$ randomly extracts two phrases that co-occur in a sentence with one of the AND/THUS-type connectives, i.e., it uses not only non-causal connectives but also causal ones like *thus*. FREQ is the same as PROP$_{caus}$ except that it ranks its output by the phrase pair co-occurrence frequency rather than Excitation values.
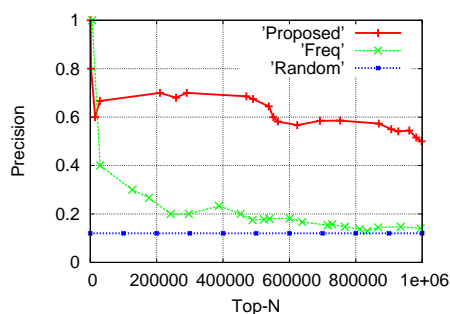
Figure 5: Precision of causality extraction.

Evaluation scheme   We randomly sampled 100 pairs each from the top one million results of PROP$_{caus}$ and FREQ, and all RAND$_{caus}$'s output. The annotators were shown the original sentences from which the samples were extracted.   Fleiss' kappa was 0.68 (substantial agreement).

Results   'Proposed' in Figure 5 is the precision curve for PROP$_{caus}$. From this curve the estimated precision of PROP$_{caus}$ is about 70% around the top 500,000.   Note that PROP$_{caus}$ outperformed FREQ by a large margin, and extracted a large variety of causal pairs since its 100 samples contained 91 different template pairs.   Table 3 shows examples of PROP$_{caus}$'s output along with English translations. The labels '✔' and '✖' denote whether a pair is causality or not.   The ✖ cases in Table 3 were exceptions to our assumption described in Section 4.2; even if two Excitation templates co-occur in a sentence with an AND/THUS-type connective, they sometimes do not constitute causality. Actually, the first ✖ case consists of two phrases that co-occurred in a sentence with a (non-causal) AND/THUS-type connective but described two events that happen as the effects of introducing the RAID storage system; both are caused by the third event. In the second ✖ case, the two phrases co-occurred in a sentence with a (non-causal) AND/THUS-type connective but just described two opposing events.

Summary   PROP$_{caus}$ performs well since it extracted 500,000 causality pairs with about 70% precision. Moreover, Excitation values contribute to causality ranking since PROP$_{caus}$ outperformed FREQ by a large margin. Then we conclude that our assumption on causality extraction is confirmed.

Rank	Causality Pairs	Label
1,036	基礎代謝を高める ⇒ 脂肪燃焼力を高める	✔
	increase basal metabolism ⇒ enhance fat-burning ability	
2,128	学習意欲を高める ⇒ 自己学習を促進する	✔
	increase desire to learn ⇒ facilitate self-learning	
6,471	信頼性を高める ⇒ 容量を増やす	✖
	improve reliability ⇒ increase capacity	
29,638	血中甲状腺ホルモン濃度が高まる ⇒ 新陳代謝が高まる	✔
	circulating thyroid hormone level increases ⇒ improves metabolism	
56,868	輸出が増える ⇒ ＧＤＰが増加する	✔
	exports increase ⇒ GDP grows	
267,364	血行を促進する ⇒ 新陳代謝を助ける	✔
	promote blood circulation ⇒ improve metabolism	
268,670	ＢＳＥが発生する ⇒ 輸入禁止になる	✔
	BSE outbreak occurs ⇒ import ban (on beef) is issued	
290,846	視界が良くなる ⇒ 作業効率が向上する	✔
	improve the view ⇒ improve the efficiency of work	
322,121	大地震が発生する ⇒ メルトダウンを起こす	✔
	giant earthquake occurs ⇒ meltdown is triggered	
532,106	熱効率が良い ⇒ 暖房効果を高める	✔
	good at thermal efficiency ⇒ enhance heating efficiency	
563,462	インフレを起こす ⇒ 円安が進行する	✔
	promote inflation (in Japan) ⇒ yen depreciation develops	
591,175	利益をもたらす ⇒ 不利益をもたらす	✖
	bring profit ⇒ bring detriment	
657,676	体力が落ちる ⇒ 免疫力が下がる	✔
	physical strength declines ⇒ immune system weakens	
676,902	国債先物急落を受ける ⇒ 金利が上昇する	✔
	sharp fall in government bond futures occurs ⇒ interest rates increase	
914,101	誤差が出る ⇒ めいわくをかける	✔
	have a margin of error ⇒ cause trouble	

Table 3: Examples of PROP$_{caus}$'s outputs.

5.4   Causality Hypothesis Generation

Here we show that our causality hypothesis generation method in Section 4.3 (PROP$_{hyp}$) extracted one million hypotheses with about 57% precision.

This experiment took the top 100,000 results of PROP$_{caus}$ as input, generated hypotheses from them, and randomly selected 100 samples from the top one million hypotheses.   We evaluated only PROP$_{caus}$, since we could not think of any reasonable baseline for this task. Randomly coupling two phrases might be a baseline, but it would perform so poorly that it could not be a reasonable baseline.

The annotators judged each sample in the same way as Section 5.3, except that we presented them with source causality pairs from which hypotheses were generated, as well as the original sentences of these source pairs. Fleiss' kappa was 0.51 (moderate agreement).

As a result, PROP$_{hyp}$ generated one million hypotheses with 57% precision. It generated various kinds of hypotheses, since these 100 samples contained 99 different template pairs.   Table 4 shows some causal hypotheses generated by PROP$_{hyp}$. The source causal pair is shown in parentheses.   The la-

bels '✔' and '✖' denote whether a pair is causality or not. The first ✖ case was due to an error made by

Rank	Causality Hypotheses (and their Origin)	Label
18,886	ストレスが減少する ⇒ 不眠が改善される	✔
	(ストレスが増加する ⇒ 不眠が続く)	✔
	alleviate stress ⇒ remedy insomnia	
	(increase stress ⇒ continue to have insomnia)	
93,781	デフレを阻止する ⇒ 税収が増加する	✔
	(デフレが進む ⇒ 税収が減る)	✔
	halt deflation ⇒ tax revenue increases	
	(deflation is promoted ⇒ tax revenes declines)	
121,163	楽しみが増大する ⇒ ストレスが減少する	✔
	(楽しみが減る ⇒ ストレスが高まる)	✔
	enjoyment increases ⇒ stress decreases	
	(enjoyment decreases ⇒ stress grows)	
205,486	犯罪を減らす ⇒ 不安が無くなる	✔
	(犯罪が増加する ⇒ 不安が高まる)	✔
	decrease in crime ⇒ diminish anxiety	
	(increase in crime ⇒ heighten anxiety)	
253,531	塩素を減らす ⇒ バクテリアは増殖する	✔
	(塩素を発生させる ⇒ バクテリアを死滅させる)	✔
	reduce chlorine ⇒ bacteria grow	
	(generate chlorine ⇒ bacteria extinct)	
450,353	需要が拡大する ⇒ 失業を減少させる	✔
	(需要が減る ⇒ 失業が増える)	✔
	expand demand ⇒ decrease unemployment rate	
	(decrease demand ⇒ increase unemployment rate)	
464,546	消化が悪くなる ⇒ コレステロールを増やす	✖
	(消化を助ける ⇒ コレステロールを減らす)	✖
	(ability of) digestion deteriorates ⇒ cholesterol increases	
	(aid digestion ⇒ decrease cholesterol)	
538,310	疲れを軽減する ⇒ 免疫を増強する	✔
	(疲れがたまる ⇒ 免疫が弱まる)	✔
	relieve fatigue ⇒ improve immunity	
	(feel fatigued ⇒ immunity is weakened)	
789,481	調子があがる ⇒ トラブルを防げる	✔
	(調子が悪くなる ⇒ トラブルが起きる)	✔
	conditions improve ⇒ prevent troubles	
	(conditions become bad ⇒ cause troubles)	
837,850	景気をコントロールする ⇒ 問題を伴う	✖
	(景気が良くなる ⇒ 問題が解消される)	✔
	control economic conditions ⇒ accompany problems	
	(economic conditions improve ⇒ problems are solved)	

Table 4: Examples of causality hypotheses.

our causality extraction method PROP$_{caus}$; the case was erroneous since its original causality was erroneous. The second ✖ case was due to the fact that one of the contradiction phrase pairs used to generate the hypothesis was in fact not contradictory (景気をコントロールする ⊥̸ 景気が良くなる 'control economic conditions ⊥̸ economic conditions improve').

From these results, we conclude that our assumption on causality hypothesis generation is valid.

6 Related Work

While the semantic orientation involving good/bad (or desirable/undesirable) has been extensively stud-

ied (Hatzivassiloglou and McKeown, 1997; Turney, 2002; Rao and Ravichandran, 2009; Velikovich et al., 2010), we believe Excitation represents a genuinely new semantic orientation.

Most previous methods of contradiction extraction require either thesauri like Roget's or WordNet (Harabagiu et al., 2006; Mohammad et al., 2008; de Marneffe et al., 2008) or large training data for supervision (Turney, 2008). In contrast, our method requires only a few seed templates. Lin et al. (2003) used a few "incompatibility" patterns to acquire antonyms, but they did not report their method's performance on the incompatibility identification task.

Many methods for extracting causality or script-like knowledge between events exist (Girju, 2003; Torisawa, 2005; Torisawa, 2006; Abe et al., 2008; Chambers and Jurafsky, 2009; Do et al., 2011; Shibata and Kurohashi, 2011), but none uses a notion similar to Excitation. As we have shown, we expect that Excitation will improve their performance.

Regarding the acquisition of semantic knowledge that is not explicitly written in corpora, Tsuchida et al. (2011) proposed a novel method to generate semantic relation instances as hypotheses using automatically discovered inference rules. We think that automatically generating plausible semantic knowledge that is not written (explicitly) in corpora as hypotheses and augmenting semantic knowledge base is important for the discovery of so-called "unknown unknowns" (Torisawa et al., 2010), among others.

7 Conclusion

We proposed a new semantic orientation, Excitation, and its acquisition method. Our experiments showed that Excitation allows to acquire one million contradiction pairs with over 70% precision, as well as causality pairs and causality hypotheses of the same volume with reasonable precision from the Web. We plan to make all our acquired knowledge resources available to the research community soon (Visit http://www.alagin.jp/index-e.html).

We will investigate additional applications of Excitation in future work. For instance, we expect that Excitation and its related semantic knowledge acquired in this study will improve the performance of Why-QA system like the one proposed by Oh et al. (2012).

References

Shuya Abe, Kentaro Inui, and Yuji Matsumoto. 2008. Two-phrased event relation acquisition: Coupling the relation-oriented and argument-oriented approaches. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pages 1–8.

Susumu Akamine, Daisuke Kawahara, Yoshikiyo Kato, Tetsuji Nakagawa, Yutaka I. Leon-Suematsu, Takuya Kawada, Kentaro Inui, Sadao Kurohashi, and Yutaka Kidawara. 2010. Organizing information on the web to support user judgments on information credibility. In *Proceedings of 2010 4th International Universal Communication Symposium Proceedings (IUCS 2010)*, pages 122–129.

Alina Andreevskaia and Sabine Bergler. 2006. Semantic tag extraction from wordnet glosses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*.

Nathanael Chambers and Dan Jurafsky. 2009. Unsupervised learning of narrative schemas and their participants. In *Proceedings of the 47th Annual Meeting of the ACL and the 4th IJCNLP of the AFNLP (ACL-IJCNLP 2009)*, pages 602–610.

Marie-Catherine de Marneffe, Anna Rafferty, and Christopher D. Manning. 2008. Finding contradiction in text. In *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pages 1039–1047.

Quang Xuan Do, Yee Seng Chan, and Dan Roth. 2011. Minimally supervised event causality identification. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing (EMNLP 2011)*, pages 294–303.

Joseph L. Fleiss. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382.

Roxana Girju. 2003. Automatic detection of causal relations for question answering. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL 2003), Workshop on Multilingual Summarization and Question Answering - Machine Learning and Beyond*, pages 76–83.

Sanda Harabagiu, Andrew Hickl, and Finley Lacatusu. 2006. Negation, contrast and contradiction in text processing. In *Proceedings of the 21st National Conference on Artificial Intelligence (AAAI-06)*, pages 755–762.

Vasileios Hatzivassiloglou and Kathleen R. McKeown. 1997. Predicting the semantic orientation of adjectives. In *Proceedings of the 35 Annual Meeting of the Association for Computational Linguistics and the 8the Conference of the European Chapter of the Association of Computational Linguistics*, pages 174–181.

Daisuke Kawahara and Sadao Kurohashi. 2006. A fully-lexicalized probabilistic model for Japanese syntactic and case structure analysis. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT-NAACL2006)*, pages 176–183.

J. Richard Landis and Gary G. Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159–174.

Dekang Lin, Shaojun Zhao Lijuan Qin, and Ming Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence (IJCAI-03)*, pages 1492–1493.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics (COLING-ACL1998)*, pages 768–774.

Saif Mohammad, Bonnie Dorr, and Greame Hirst. 2008. Computing word-pair antonymy. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 982–991.

Jong-Hoon Oh, Kentaro Torisawa, Chikara Hashimoto, Takuya Kawada, Stijn De Saeger, Junichi Kazama, and Yiou Wang. 2012. Why question answering using sentiment analysis and word classes. In *Proceedings of EMNLP-CoNLL 2012: Conference on Empirical Methods in Natural Language Processing and Natural Language Learning*.

Charles E. Osgood, George J. Suci, and Percy H. Tannenbaum. 1957. *The measurement of meaning*. University of Illinois Press.

James Pustejovsky. 1995. *The Generative Lexicon*. MIT Press.

Delip Rao and Deepak Ravichandran. 2009. Semi-supervised polarity lexicon induction. In *Proceedings of the 12th Conference of the European Chapter of the ACL*, pages 675–682.

Tomohide Shibata and Sadao Kurohashi. 2011. Acquiring strongly-related events using predicate-argument co-occurring statistics and case frames. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, pages 1028–1036.

Philip J. Stone, Dexter C. Dunphy, Marshall S. Smith, and Daniel M. Ogilvie. 1966. *The General Inquirer: A Computer Approach to Content Analysis*. MIT Press.

Hiroya Takamura, Takashi Inui, and Manabu Okumura. 2005. Extracting semantic orientation of words using

spin model. In *Proceedings of the 43rd Annual Meeting of the ACL*, pages 133–140.

Kentaro Torisawa, Stijn De Saeger, Jun'ichi Kazama, Asuka Sumida, Daisuke Noguchi, Yasunari Kakizawa, Masaki Murata, Kow Kuroda, and Ichiro Yamada. 2010. Organizing the web's information explosion to discover unknown unknowns. *New Generation Computing (Special Issue on Information Explosion)*, 28(3):217–236.

Kentaro Torisawa. 2005. Automatic acquisition of expressions representing preparation and utilization of an object. In *Proceedings of the Recent Advances in Natural Language Processing (RANLP05)*, pages 556–560.

Kentaro Torisawa. 2006. Acquiring inference rules with temporal constraints by using japanese coordinated sentences and noun-verb co-occurrences. In *Proceedings of the Human Language Technology Conference of the North American Chapter of the ACL (HLT-NAACL2006)*, pages 57–64.

Masaaki Tsuchida, Kentaro Torisawa, Stijn De Saeger, Jong-Hoon Oh, Jun'ichi Kazama, Chikara Hashimoto, and Hayato Ohwada. 2011. Toward finding semantic relations not written in a single sentence: An inference method using auto-discovered rules. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP 2011)*, pages 902–910.

Peter D. Turney. 2002. Thumbs up or thumbs down? semantic orientation applied to unsupervised classification of reviews. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL 2002)*, pages 417–424.

Peter Turney. 2008. A uniform approach to analogies, synonyms, antonyms, and associations. In *Proceedings of the 22nd International Conference on Computational Linguistics (COLING 2008)*, pages 905–912.

Leonid Velikovich, Sasha Blair-Goldensohn, Kerry Hannan, and Ryan McDonald. 2010. The viability of web-derived polarity lexicons. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the ACL*, pages 777–785.

Ellen M. Voorhees. 2008. Contradictions and justifications: Extensions to the textual entailment task. In *Proceedings of the 48th Annual Meeting of the Association of Computational Linguistics: Human Language Technologies (ACL-08: HLT)*, pages 63–71.