'Fighting' or 'Conflict'? An Approach to Revealing Concepts of Terms in Political Discourse

Linyuan Tang and **Kyo Kageura**

Graduate School of Interdisciplinary Information Studies / Interfaculty Initiative in Information Studies
The University of Tokyo, Tokyo, Japan
linyuan-tang@g.ecc.u-tokyo.ac.jp, kyo@p.u-tokyo.ac.jp

Abstract

Previous work on the epistemology of fact-checking indicated the dilemma between the needs of binary answers for the public and ambiguity of political discussion. Determining concepts represented by terms in political discourse can be considered as a Word-Sense Disambiguation (WSD) task. The analysis of political discourse, however, requires identifying precise concepts of terms from relatively small data. This work attempts to provide a basic framework for revealing concepts of terms in political discourse with explicit contextual information. The framework consists of three parts: 1) extracting important terms, 2) generating concordance for each term with stipulative definitions and explanations, and 3) agglomerating similar information of the term by hierarchical clustering. Utterances made by Prime Minister Abe Shinzo in the Diet of Japan are used to examine our framework. Importantly, we revealed the conceptual inconsistency of the term *Sonritsu-kiki-jitai*. The framework was proved to work, but only for a small number of terms due to lack of explicit contextual information.

1 Introduction

In October 2016, in the process of diet deliberations on assigning Japan's Self-Defense Forces members to U.N. operations in South Sudan, Japanese Prime Minister Abe Shinzo stated that the 'fighting' between the government and rebel forces were not to be considered as a 'military conflict'[1], according to the definition of 'conflict' under Japanese peacekeeping law.

When domain-specific jargons are used, the ambiguity between common usage and domain-specific usage becomes inevitable. In addition, the case above illustrates that the task of political discourse analysis is different from other scientific discourse analysis. In contrast to other scientific domains, terms used by political figures tend to be vague and ambiguous due to their unwillingness to explain their opinions or statements sufficiently clearly to the public. Although social scientists may derive certain implications from the ambiguities, an intentional misuse of terms by a political figure, which is difficult to recognize, could lead the public to misinterpretation.

As a prerequisite for fact-checking, therefore, it is essential to reveal concepts represented by terms in political discourse. As we cannot expect politicians to do this, it is necessary for the public and/or journalists to disambiguate concepts of terms. Automatic processing of political texts, namely word sense disambiguation (WSD), has a potential to assist this process.

The procedure of WSD can be summarized as: 'given a set of words, a techique is applied which makes use of one or more sources of knowledge to associate the most appropriate senses with words in context' (Navigli, 2009). WSD in general relies on knowledge; without knowledge, not only computers but also human beings cannot understand word sense. The unwillingness of political figures to clarify the meaning of their utterances causes at least two difficulties in WSD specific to political discourse.

First, there exist few political domain-specific dictionaries or corpora that could serve as a knowl-

---

[1] As in English, there are also various verbal expressions (*Sentō*, *Shōtotsu*, etc.) of the act of fighting in Japanese. *Sentō* is the term that generally thought to be parallel to 'military conflict' as well as 'military fighting'.

edge base, which is desirable for WSD. Generally, in order to facilitate communication, a dictionary defines standard usage and a corpus exhibits practical usage of terms. On the other hand, political figures almost always create specific usage of terms to escape from common understanding.

Second, every term in political discourse could have a peculiar concept. It is well known in Japan that most utterances made in the Diet are drafted by bureaucrats, and there are always subtle nuances present in bureaucratese. If necessary, political figures would make every single term be independent of its synonyms, hyponyms and hypernyms, even when the terms share a similar surface word form[2]. Therefore, unlike the tasks of document summarization or simplification, when revealing the concept of a term made by political figures, information loss is relatively less acceptable.

The first difficulty could be overcome by creating a domain-specific knowledge base or applying unsupervised disambiguation or word sense induction (WSI) methods. However, knowledge provided by political knowledge bases, which is necessary for further research, could sometimes obstruct the analysis, because concepts of terms can vary across political figures and scenarios. WSI, on the other hand, while it is also necessary for further research, suffers from a more practical problem, i.e., it identifies sense clusters rather than assigning a sense to a word, and 'a significant gap still exists between the results of these techniques and the gold standard of manually compiled word sense dictionaries' (Denkowski, 2009).

In view of the urgent need for an accessible and straightforward approach to practical WSD for political discourse, this ongoing research provides a springboard by introducing a framework to reveal concepts of terms using only explicit contextual information. The method we propose copes with the balance of the needs of knowledge and the attention to the specific usages of terms by creating a concordance that serves as a temporary knowledge base. On the other hand, it deals with precision of concept generation by keeping as much information as possible.

We collected utterances made by Prime Ministers of Japan in the Diet deliberations as target dis-

courses. The concept-revealing framework consists of three parts. First, we applied widely-used tf-idf method to weigh terms and acquired nouns with ranks by their importance. Second, we generated a concordance for each of the important terms in order to collect their stipulative definitions and explanations offered in the document. Thirdly, focusing on the similarity of concepts rather than the quantity of clusters, we agglomerated similar information by hierarchical clustering.

Theoretically, as our approach extracts information from given documents without summarization or simplification, concepts of terms will surely be revealed. Given this, we will show, instead of emphasizing the overall results, an important observation obtained from the concept of *Sonritsu-kiki-jitai*[3] which was identified as one of the most important terms used by Prime Minister Abe Shinzo. Specifically, conceptual inconsistency exists not only between the speaker and the audience, but also in the same speaker's utterances.

2 Related work

The controversy among Uscinski and Butler (2013), Amazeen (2015), and Uscinski (2015) over the epistemology of fact-checking illustrated issues on the methodology of fact-checking.

Uscinski and Butler (2013) made five methodological criticisms against fact-checking methods: selection effects, confounding multiple facts or picking apart a whole, causal claims, predicting the future, and inexplicit selection criteria. These challenges were related to 'the naïve political epistemology at work in the fact-checking branch of journalism' (Uscinski and Butler, 2013).

Amazeen (2015) critized Uscinski and Butler (2013) for their overgeneralization of the selection effects and failure to offer supportive empirical quantification. She also demonstrated that there was a high level of consistency among multiple fact-checkers, and argued that fact-checking is important as long as 'unambiguous practices of deception' continue (Amazeen, 2015).

The rejoinder then from Uscinski (2015) argued that Amazeen's attempt to infer the accuracy of fact-checks failed because of fact-checkers' possible political biases, and she also ignored the distinction between facts and claims. Fact-checking was therefore still a 'continuation of politics by

---

[2]In the example above, according to the government, *Sentō* (fighting) is not defined while *Sentō-kōi* (act of fighting) is defined by the law, so that these two terms have different meanings and only the latter can be used.

[3]An armed attack against foreign country resulting in threatening Japan's survival.

means of journalism' rather than being an 'counterweight to political untruths' (Uscinski, 2015).

Although the discussion was mainly about the epistemological disagreement over so-called "truth" between journalists and social scientists, it did indicate the dilemma between 'the needs of citizens, politicians, and therefore journalists for clear-cut binary answers' (Uscinski, 2015) and ambiguity of most politcal discussion, which suggests the necessity of a novel perspective on fact-checking, focusing on how political figures performed their language rather then what really occurred.

3 Approach

3.1 Dataset

We assembled a corpus of utterances made by prime ministers of Japan at the Plenary Session of the Diet from 1996 to 2016, from the Diet Record[4]. The corpus of 2605 fulltext discourses includes utterances from 11 prime ministers, 47 sessions. We selected utterances of Abe Shinzo, the incumbent Japanese Prime Minister, as our targets. The target utterances include 427 fulltext discourses from 6 sessions (16469 sentences, 9715 types, 492505 tokens). We used the rest of the corpus as supplementary materials to weigh the terms.

3.2 Procedure

3.2.1 Term extraction

We firstly seperated 2605 discourses into 47 documents in accordance with sessions of the Diet (6 target documents for Prime Minister Abe). After data cleansing, we used *ChaSen*[5] (A Japanese morphological analyzer) to convert each document into a bag of its nouns. Nominal verbs were also included.

We then ranked nouns to obtain the most important terms in each document. We applied the tf-idf model because it is one of the most popular term-weighting schemes and is empirically useful.

3.2.2 Concordance generation

For each important term in the document, a list of all the instances of the term was generated if the term co-occurred with a stipulation expression such as *to-ha*, *to-tēgi* (both of the phrases represent 'be defined as')[6]. An instance of a term was a sentence which consists of the term and its context. All the instances of a term formed its concordance. The term's concept was constructed with only these instances.

3.2.3 Concept clustering

We converted every entry in the concordance into a vector for calculating the similarity of the term's contextual information. We applied tf-idf model instead of word embedding. Word embeddings contain biases in their geometry that reflect stereotypes present in broader society, and word embeddings not only reflect such stereotypes but can also amplify them (Bolukbasi et al., 2016). On the other hand, tf-idf has no semantic representation of words. In order to cope with potential subtle nuances in the utterances of political figures, a non-semantic representation is preferable to a semantic one. We then generated a hierarchy of clusters of the entries with Ward's method. Even though clustering approaches in WSI are usually non-hierarchical (Navigli, 2009; Denkowski, 2009), the reason for applying a hierarchical clustering instead of a non-hierarchical one is that we focused on the similarity of entries rather than the quantity of concepts.

Finally, by eliminating duplicated entries[7] and combining the remainder manually, we were able to acquire concepts of terms which are constructed with explicit stipulative definitions and explanations offered in documents. The revealed concept was therefore entirely contextual and independent of that which we have already known about.

4 Results

We treated the top 100 out of 8298 (3568 proper nouns) nouns as important terms. Three of them are explicitly defined in Prime Minister Abe's utterances: *Sekkyokuteki-hēwa-syugi* (Proactive pacifism), *Sonritsu-kiki-jitai*, and *Rikken-syugi* (Constitutionalism). *Sekkyokuteki-hēwa-syugi* was ranked as the 17th most important term among all the nouns. The cluster dendrogram of 68 sentences which were in the term's concordance is shown in figure 1. *Sonritsu-kiki-jitai* was ranked as the 24th most important term among all the

---

[4] http://kokkai.ndl.go.jp

[5] http://chasen-legacy.osdn.jp

[6] Besides these two phrases, various expressions are used to define terms. Since there haven't been a comprehensive summary of definition/stipulation expressions yet, we may not be able to extract all the potential defined terms.

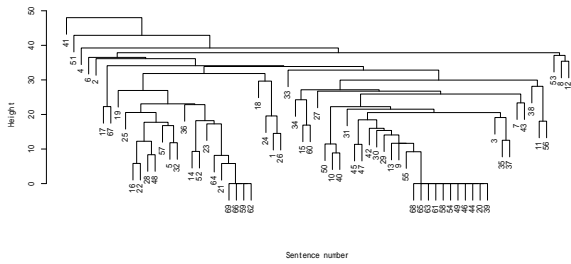[7] Most utterances were repeated twice at the House of Representatives and the House of Councilors.

Figure 1: Cluster dendrogram of *Sekkyokuteki-hēwa-syugi*
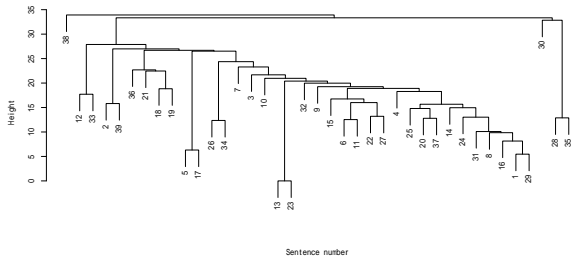

Figure 2: Cluster dendrogram of *Sonritsu-kiki-jitai*

nouns. The clustering result of 39 sentences is shown in figure 2. *Rikken-syugi* was ranked as the 72nd most important term among all the nouns. The clustering result of 22 sentences is shown in figure 3. Few terms were defined in utterances of Prime Ministers (15 terms in total).

Mutually contradictory explanations were found in the concordance of *Sonritsu-kiki-jitai*. Specifically, this term is currently translated to 'an armed attack against foreign country resulting in threatening Japan's survival', and is defined as a situation that 'an armed attack against Japan or a foreign country that is in a close relationship with Japan occurs, and as a result, threatens Japan's survival and poses **a clear danger** to fundamentally overturn people's right to life, liberty and pursuit of happiness' by the Ministry of Foreign Affairs of Japan. The
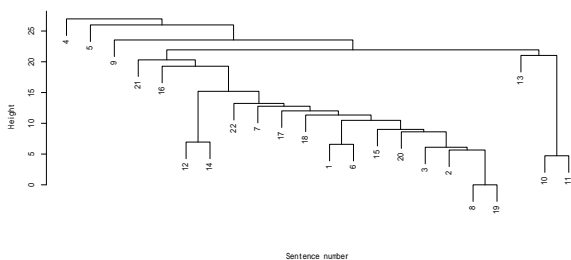

Figure 3: Cluster dendrogram of *Rikken-syugi*

situation is also one of three new conditions by which "use of force" as measures for self-defense is strictly limited[8]. This definition was mentioned two times in Prime Minister Abe's utterances (the 189th session on 26th May and 27th July, 2015). However, it was also mentioned several times that to determine whether a situation is a *Sonritsu-kiki-jitai* requires a **comprehensive analysis** by the government (18th May, 26th May, 29th May, 27th July, 2015; 27th Jan, 2016). Concisely, the concept of *Sonritsu-kiki-jitai* is a 'clear danger' that requires a 'comprehensive analysis' to determine whether it is a clear danger or not. This conceptual inconsistency turns one of the limitations on "use of force" into a mere scrap of paper.

5 Discussion

Political discourse is always vague and ambiguous. Nonetheless, we can still recognize in what manner it is vague and ambiguous. Even though the mission of fact-checking is 'not to measure which candidate "lies most" but rather to provide the public with information about the accuracy of statements' (Amazeen, 2015), in respect of accuracy, the information about how political figures performed their language is as important as the information about what really occurred.

This ongoing work opens a novel perspective on WSD for political discourse as well as fact-checking, by pointing out that a confirmation of concepts of terms which formed discourse is a prerequisite for analyzing formal utterances by political figures.

Our framework makes it possible for the public and/or journalists to recognize the most important terms as well as their stipulative concepts in an objective way. Moreover, we revealed the possibility that conceptual incosistency can also exist in a single term as exemplified by revealed concept of *Sonritsu-kiki-jitai*. This indicated that there is a possibility that a term could be meaningless due to an inherent self-contradiction in its concept.

Due to inadequate explicit information in Prime Minister Abe's utterances, few concepts were revealed. This identified a weakness of our approach, i.e., it relies on how explicitly a speaker stipulated a term. Nonetheless, from another perspective, by focusing on the lack of explicit definitions and explanations of important terms in

---

[8]http://www.mofa.go.jp

discourse, The vagueness and ambiguity of utterances could be evaluated.

Our work is an ongoing research aims at establishing a practical standard for terminological analysis of political discourse. To start with, we provided this framework for revealing concepts of terms in political discourse. It could be technically improved in the following ways: 1) by analyzing the structure of documents' terminology sets and applying suitable term weighting models, we may generate a more applicable term ranking; 2) by discovering patterns of stipulative definition and explanation, we may assemble a more adequate concordance of a term from discourse; and 3) by applying suitable clustering and summarization methods, we may create a better balance between precision and concision.

Acknowledgments

References

Michelle A. Amazeen. 2015. Revisiting the epistemology of fact-checking. *Critical Review* 27(1):1–22.

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. 2016. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, Curran Associates, Inc., pages 4349–4357.

Michael Denkowski. 2009. A survey of techniques for unsupervised word sense induction. *Language & Statistics II Literature Review* pages 1–18.

Roberto Navigli. 2009. Word sense disambiguation: A survey. *ACM Computing Surveys* 4(2):10.

Joseph E. Uscinski. 2015. The epistemology of fact checking (is still naive): Rejoinder to amazeen. *Critical Review* 27(2):243–252.

Joseph E. Uscinski and Ryden W. Butler. 2013. The epistemology of fact checking. *Critical Review* 25(2):162–180.