Future Reference Sentence Extraction in Support of Future Event Prediction

Yoko Nakajima [1], Michal Ptaszynski [2], Hirotoshi Honma [1], Fumito Masui [2]
[1] National Institute of Technology - Kushiro College, Japan
{yoko, honma}@kushiro-ct.ac.jp
[2] Kitami Institute of Technology, Japan
{ptaszynski, f-masui}@cs.kitami-it.ac.jp

ABSTRACT: *Providing the means for understanding natural language, or some groups of its realizations is one of the main goals of Artificial Intelligence (AI). For example, subfields of AI such as natural language processing (NLP), or sentiment analysis (SA) focus on analyzing of speaker attitudes and emotions expressed in a sentence. A different kind of realization of natural language, which we focus on in this research, is predicting trends of future events from available language data. It is a challenging task, since it focuses on obtaining information on future unfolding of events from the provided instances of language behavior. To make the task feasible we narrowed the scope of our research to detecting and extracting sentences which refer to future events as potentially the most useful in prediction of future unfolding of events. Future reference sentences can contain information on the current state of events, or specific expert knowledge, and thus can be useful in future prediction. We propose a method for automatic extraction of future reference sentences from news corpora using semantic and morphological information. We also perform an experiment in which the sentences extracted with the proposed method are applied to predict future unfolding of a set of specific events.*

Keywords: Future Reference Expressions, Future Prediction, Information Extraction, NLP

Received: 11 March 2017, Revised 9 April 2017, Accepted 19 April 2017

1. Introduction

It is common in everyday life for people to apply their experience and knowledge on past events to predict future events. In such everyday predictions people use widely available resources (newspapers, internet). In this study we focused on sentences referring to the future as potentially most useful in predicting future unfolding of events. Especially people encumbered with large social responsibility, such as planing, strategy management, policy development, especially in large companies or in government, are in constant need of such prediction support tools, since the company's results and profits depend on how accurate their competence in future trend prediction is.

For example, when a company have just found out that a decision depends on the event X, while event Y could happen or not respectively, they can prepare four tactical options A, B, C or D for their company management, depending on the predictions on what would happen in the future:

1. Tactical option 'A' when both events 'X' and 'Y' happen.
2. Tactical option 'B' when the event 'X' happens and the event 'Y' does not happen.
3. Tactical option 'C' when the event 'Y' happens and the event 'X' does not happen.
4. Tactical option 'D' when both events 'X' and 'Y' do not happen.

When people decide to act in a certain way, they usually choose from a range of possible options. In the choice they consider and combine a large amount of information, including one's experiences and other's expertise. Obtaining such information to make accurate future predictions is a challenging task requiring time and labor. A lot of information needs to be processed and an accurate foresight ability needs to be exercised in making the decision. Previous studies have shown that data mining using statistical techniques can support predictions regarding future of events. However, to achieve that, one needs to process large amount of numeric data, which requires professional skills.

In addition, even if people other than an expert obtain statistically reliable information, it is difficult for them to understand its actual meaning or value. Using sentences that refer to the future which represent knowledge similar to that possessed by experts could be easier for laypeople people to understand. Therefore, when there is future event we want to know about and there are given some options, future-referring sentences could support answering such a question. See the example below:

Question: Predict whether in 2020 nursing care fee will be applying AI in Care Planning.

Options:

(a) AI will be applied in Care Planning.

(b) The situation will not change.

(c) Care Planning will be created using new method other than AI.

Example relevant sentences that refer to the future:

- The government plans to revise the nursing care fee in future.
- The government announced that it would develop the legal system and rules for society with AI.
- A model in which people and artificial intelligence work together will be constructed.

Future reference sentences are applicable as a useful knowledge base since they contain various related information, such as background information regarding the event in question, which is also used as the source of knowledge by experts. Therefore, such sentences can be useful for making decisions on future predictions.

Nakajima et al. (2016) [15] proposed a method for automatic extraction of future reference sentences using morphosemantic information and indicated that future reference sentences could be useful in supporting predictions about future events.

In this research[1] we conduct experiments which evaluate the support of future prediction using future reference sen-

[1] *This paper is an extended paper of the workshop "Language Sense on Computers" on* 25th *International Joint Conference on Artificial Intelligence* (2016) [14].

tences and discuss its effectiveness.

The outline of this paper is as follows. In Section 2 we describe previous research related to the prediction of future events. Section 3 describes the proposed method applying automatic extraction of references to future events. Section 4 describes the experiments evaluating the method. Section 5 proposes a discussion of the examined approach and future implications. Finally, Section 6 contains conclusions and sets out plans for future improvement and applications of the proposed method.

2. Previous Research

Linguistically expressed references to the future has been studied by a number of researchers.

Baeza-Yates (2005) [3] investigated five hundred thousand sentences containing future events extracted from one day of Google News (*http://news.google.com/*), and found out that scheduled events occur with high probability and with correlation between the occurrence of an event and its time proximity. It suggests that there is plenty of information of future events. Kanhabua et al. (2011) [10] investigated newspaper articles, and found out that one-third of all sentences contain reference to the future. Kanazawa et al. (2010) [8] extracted implications for future information from the Web using explicit information, such as time expressions. Alonso et al. (2011) [1] indicated that time information included in a document is effective for enhancing information retrieval applications. Kanazawa et al. (2011) [9] extracted unreferenced future time expressions from a large collection of text, and proposed a method for estimating the validity of the prediction by searching for a real-world event corresponding to the one predicted automatically. Jatowt et al. (2013) [7] studied relations between future news in English, Polish and Japanese by using keywords queried on the Web. Popescu and Strapparava (2013) [17] investigated the distribution of terms within the Google Books corpus and noticed significant changes in time as well as established relationship with emotion words.

When it comes to predicting the probability of an event to occur in the future, Jatowt et al. (2009) [5] used the rate of incidence of reconstructed news articles over time to forecast recurring events, and proposed a method for supporting human user analysis of occurring future phenomena. Later, Jatowt et al. (2011) [6] proposed a clustering algorithm for detecting future phenomena based on the information extracted from text corpus, and proposed a method of calculating the probability of an event to happen in the future. Aramaki et al. (2011) [2] used an SVM-based classifier on Twitter to perform classification of information related to influenza and tried to predict the spread of the disease by using a truth validation method. In the research of Kanazawa et al. (2011) [9] the proposed method estimated the validity of prediction by automatically calculating cosine similarity between predicted relevant news and searching for the events that actually occurred. Radinsky et al. (2012) [19] proposed the Pundit System for prediction of future events in news based on causal reasoning derived from a similarity measure calculated using different ontologies.

The above findings have lead us to the idea that by using expressions referring to the future included in trend reports (newspaper articles, etc.), we could be able to support future prediction process as one of the activities that people perform everyday. For example, if we take sentences mentioning the future, such as the example below,

"The technique for applying gas occurring in underground waters to generate power is rare and the company is going to sell it worldwide in Europe, China and other countries."

We would be able to determine plausibility of events implied by these sentences. Such a method would be applicable in corporate management, trend foresight, and preventive measures, etc. Also, as indicated by previous research, when applied in real time analysis of Social Networking Services (SNS), such as Twitter or Facebook, it could also become helpful in disaster prevention or handling of disease outbreaks. This way the method would be useful in chance discovery (2003) [22], by e.g., providing hints for a company regarding planning its future investments.

Methods using time referring information, such as "year", "hour", or "tomorrow", has been applied in extracting future information and retrieving relevant documents. It has also been indicated that it is useful to predict future outcomes by using information occurring in present documents. In our research we focused on more sophisticated expressions, namely, morphosemantic sentence patterns.

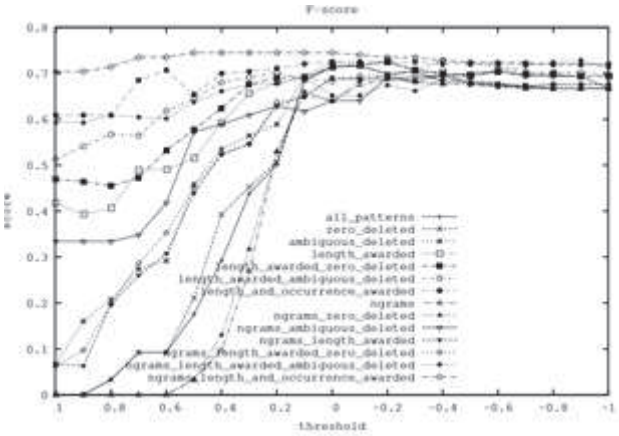

Figure 1. Shows the TC process

3. Method for Automatic Extraction of Future Reference Sentences

In this section, we describe our method for extracting future reference sentences from news corpora. Future reference sentences include both explicit as well as implicit expressions referring to the future. Explicit expressions include e.g., future temporal expressions, or words and phrases referring to the future (e.g. will∼, be expected that ∼, plan to ∼, etc.).

However, many important sentences do not contain such explicit expressions, but the information regarding future outcomes can be encoded as an implicit information. See the example below regarding the future of America's troops dispatch to Afghanistan.

"*He rejoiced that President Obama had reemphasized the need to focus on the War on Terror in Afghanistan, increasing the likelihood of an early withdrawal of U.S. troops from Iraq*."

The sentence does not contain any future referring expressions. Moreover, the sentence is in past tense ("rejoiced", "had reemphasized"), and therefore it is not possible to specify that the sentence refers to the future by using standard methods. Yet, the sentence clearly presents potential future outcomes ("withdrawal of U.S. troops from Iraq") with the use of implicit information.

The method we propose deals with both explicit as well as implicit information, such as the above, and consists of two stages. Firstly, the sentences are represented in a morphosemantic structure [13] (combination of semantic role labeling and morphological information). Secondly, frequent combinations of such patterns are automatically extracted from training data and used in classification.

Morphosemantic patterns (MoPs) are useful for representing languages rich both morphologically and semantically, such as Japanese (language of datasets used in this research). We generated the morphosemantic model using semantic role labeling (SRL) supported with morphological information. SRL provides labels for words and phrases according to their role in the sentence, such as those represented in Table 1.

3.1 Morphosemantic Patterns

In the first stage, all sentences are represented in morphosemantic structure (MS) for further extraction of morphosemantic patterns (MoPs) in the second stage.

The idea of MS has been described widely in linguistics and structural linguistics. For example, Levin and Hovav (1998) [13] distinguish them as one of the two basic types of morphological operations on words, which modify the Lexical Conceptual Structure (LCS), or the semantic representation of a word. As for practical application of the idea,

Example I: Romanized Japanese (RJ): Ashita kare wa kanojo ni tegami o okuru darō./ Glosses: Tomorrow he TOP her DIR letter OBJ send will (TOP: topic particle, DIR: directional particle, OBJ: object particle.) / English translation (E): He will [most probably] send her a letter tomorrow.

No	Surface	Label
1	ashita	[Time-Point]
2	kare ha	[Agent]
3	kanojo ni	[Goal]
4	tegami o	[Object]
5	okuru darou	[State_change]-[Place_change]- [Change_of_place(physical)]

Table 1. An example of a sentence analyzed by ASA

Kroeger (2007) [11] applied MoPs to analyze an Indonesian suffix -kan. Later, Fellbaum et al (2009) [4] applied MoPs to improve links between the synsets in WordNet. More recently, Raffaelli (2013) [20] used MoPs to analyze a lexicon in Croatian, a language rich both morphologically and semantically. In this research we used datasets in Japanese, and applied MoPs for the same reason. Using only one representation, either morphological or semantic, narrows the spectrum of analyzed information.

We generated the morphosemantic model using semantic role labeling with additional morphological information. Below we describe in detail the process of morphosemantic representation of sentences. At first, the sentences from the datasets are analyzed using semantic role labeling (SRL). SRL provides labels for words and phrases according to their role in sentence context. For example, in a sentence "John killed Mary" the labels for words are as follows: John=actor, kill[past]=action, Mary=patient. Thus the semantic representation of the sentence is "[actor][action][patient]".

For semantic role labeling in Japanese we used ASA[2] , a system, developed by Takeuchi et al. (2010) [21], which provides semantic roles for words and generalizes their semantic representation using an originally developed thesaurus. An example of SRL provided by ASA is represented in Table 1.

Moreover, not all words are semantically labeled by ASA. The omitted words include those not present in the thesaurus, as well as grammatical particles, or function words not having a direct influence on the semantic structure of the sentence, but in practice contributing to the overall meaning. For such cases we used a morphological analyzer MeCab[3] in combination with ASA to provide morphological information, such as "Proper Noun", or "Verb". However, in its basic form MeCab provides morphological information for all words separately, which causes compound words to be unnecessarily divided. For example "Japan health policy" is one morphosemantic concept, but in grammatical representation it takes form of "Noun Noun Noun". Therefore as a post-processing procedure we added a set of linguistic rules for specifying compound words in cases where only morphological information was provided.

Moreover, ASA provides labels such as category, semantic role, morphemes, etc. for each sentence element. Therefore in order to normalize and simplify the patterns, we specified the priority of label groups in the following way.

1. Semantic role (Agent, Patient, Object, etc.)
2. Semantic meaning (State_change, etc.)
3. Category (Dog → Living animal → Animated object)
4. In case ASA does not provide any of the above labels, perform compound word clustering for parts of speech (e.g., "International Journal of Computational Linguistics Research" → Adjective Noun Preposition Adjective Noun Noun→ Proper_Noun)

Furthermore, post-processing in the case of no semantic information is organized as follows.

[2] *http://cl.it.okayama-u.ac.jp/study/project/asa*
[3] *http://taku910.github.io/mecab/*

- If a compound word can be specified, output the part-of-speech cluster (point 4 above).
- If it is not a compound word, output part-of-speech for each word.

Below is an example of a sentence generalized with the morphosemantic tagging method applied in this research.

Romanized Japanese: Nihon unagi ga zetsumetsu kigushu ni shitei sare, kanzen yōshoku ni yoru unagi no ryōsan ni kitai ga takamatte iru.

English: As Japanese eel has been specified as an endangered species, the expectations grow towards mass production of eel in full aquaculture.

SRL: [Object][Agent][State_change][Action][Noun][State_change][Object][State_change]

3.2 Future Reference Pattern Extraction

From sentences represented this way we extract frequent MoPs using SPEC [18]. Firstly, we generate ordered non-repeated combinations from all sentence elements. In every n-element sentence there is k-number of combination groups, such as that $1 \leq k \leq n$. All combinations for a'll values of k are generated. Additionally, all non-subsequent elements are separated with a wildcard ("*", asterisk). Pattern lists extracted this way from training set are then used in classification of test and validation set.

SPEC uses all patterns generated this way to extract frequent patterns appearing in a given corpus and calculates their weight. Two features are important in weight calculation. A pattern is the more representative for a corpus when, the longer it is (length k), and the more often it appears in the corpus (occurrence O). Thus the weight can be calculated by

- Awarding length (LA),
- Awarding length and occurrence (LOA),
- Awarding none (normalized weight, NW).

The generated list of frequent patterns can be also further modified. When two collections of sentences of opposite features (such as "future-related vs. non-future-related") is compared, the list will contain patterns that appear uniquely in only one of the sides (e.g., uniquely positive patterns and uniquely negative patterns) or in both (ambiguous patterns). Thus pattern list can be modified by

- Using all patterns (ALL),
- Erasing all ambiguous patterns (AMB),
- Erasing only those ambiguous patterns which appear in the same number on both sides (zero patterns, 0P). Moreover, a list of patterns will contain both the sophisticated patterns (with disjointed elements) as well as more common n-grams. Therefore the system can be trained on a model using
- Patterns (PAT), or
- Only n-grams (NGR).

All combinations of those modification are tested in the experiment.

3.3 Future Reference Sentence Extraction with Morphosemantic Patterns

From three newspaper corpora [4] We collected and annotated a dataset containing equal number of (1) sentences refer-

[4] *Nihon Keizai Newspaper, Asahi Newspaper, Hokkaido Newspaper*.

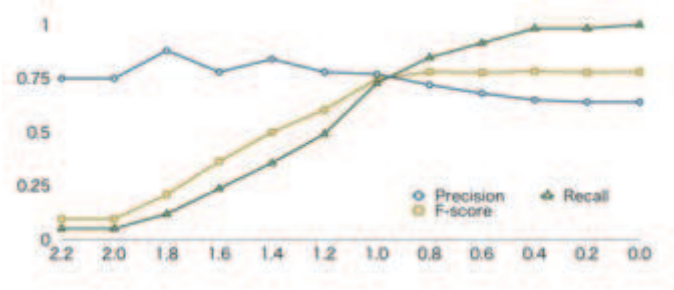

Figure 2. Shows the TC process

ring to future events and (2) other (describing past, or present events). We conducted an evaluation experiment with training dataset containing 130 sentences each, furthermore as the test data we used randomly extracted additional 170 sentences from the news corpora.

The test datasets were applied to a text classification task on 10-fold cross validation. Each classified test sentence was given a score calculated as a sum of weights of patterns extracted from training data and found in the input sentence.

Pattern set	Precision	Recall	F-score
10 patterns	0.39	0.49	0.43
10 pattern with only over 3 elements	0.42	0.37	0.40
5 patterns	0.35	0.35	0.35
Optimized (see Fig. 2)	0.76	0.76	0.76
10 word expressions	0.50	0.05	0.10

Table 2. Comparison of results for validation set between different pattern groups of MoPs, optimized MoPs and 10 word expressions

Occ.	Future Reference Patterns	Occ.	Non-future Reference Patterns
43	[Action]*[Object]	4	[Numeric]*[Agent]
42	[Action]*[Action]	4	[Verb]*[Artifact]
26	[Action]*[State change]	5	[Place]*[Agent]
20	[State change]*[Object]	4	[Person]*[Place]
16	[State change]*[State change]	3	[Numeric]*[Agent]*[Action]
	.		.
	.		.
	.		.

Table 3. Examples of extracted morphosemantic patterns (MoPs)

The results were calculated with Precision, Recall and balanced F-measure. We compared F-measures of fourteen classifier versions, shown in Figure 1. We also looked at top scores within the threshold, checked which version got the highest break-even point (BEP) of Precision and Recall, and calculated statistical significance of the results.

The results indicated that the highest overall performance was obtained by the version using pattern list containing all patterns (including ambiguous patterns and n-grams).

We also compared the proposed method with a method of word expression base, which extracts future reference sentences with 10 words unambiguously referring to the future, such as “will” or “be likely to” in English.

In Japanese grammar, the predicate verb “ru-type” (the end of pronunciation is “ru”) is defined as non-past tense, suggesting either present or future unfolding. Among sentences containing the predicate, there are exceptions which represent circumstances or description. It is possible to look at a kind of predicate to distinguish whether a sentence contains future reference. Moreover, aspect and modality can help extract time expressions. However, to classify sentences as future-referring or not using only word expressions is difficult because word expressions are sparse. Nakajima et al. [16] investigated word expressions in 270 future reference sentences and described that expressions used more than twice cover only 45 % of all, while words used only once were 55% of all.

Specifically, the 10 most often appearing words were represented below (in Romanized Japanese and closest English equivalent).

- “ni naru”(become)
- “darou” (will)
- “ga aru”(be likely to)
- “souda”(they say that)
- “ni naru”
- “kamoshirenai”(may)
- “kitai dekiru”(prommising)
- “yoteida”(be going to)
- “keikaku”(plan)
- “kanou+kanousei”(possiblity)

In comparison, the proposed method obtained better results even when only 10 most frequent MoPs were used (Table 2). It is indicated that in this case using only word expressions, the sentences were detected with sufficient precision but were extracted only in a small number. On the other hand, MoPs detected more future reference sentences than using only word expressions. Moreover, Precision grows when more MoPs are applied. Which suggest versatility of MoPs is higher than words expressions when it comes to patterns referring to the future.

Hence, we verified the performance of the fully optimized model (FOM). We retrained the best model using all sentences from the initial dataset and verified the performance by classifying the new validation set. The final overall performance is represented in Figure 2. The obtained break-even point (BEP) was 0.76.

Additionally, analysis of the patterns most frequently used in classification (see Table 3) revealed that MoPs of FRS occur frequently, while for non-FRS the occurrence was lower. This indicates that FRS can be considered as one linguistically consistent group of sentences.

4. Future Prediction Support Experiment

In this section we present a validation experiment for the effectiveness of using future reference sentences (FRS) in the task of supporting predictions regarding future events.

4.1 Experiment Setup

In the experiment for supporting future trend prediction we used the fully optimized model of future reference sentences (FRS) trained on morphosemantic patterns (MoPs) described in Section 3.3. The model was applied to extract new FRS concerning a specific topic, from the available newspaper data. Such sentences are further called future prediction support sentences (FPSS). Future prediction task was performed by a group of thirty laypeople (balanced

Question 3: Predict the stationing status of U.S. troops in Afghanistan at the end of June 2011.
(A) The U.S. troops will be still present and further reinforced comparing to October 2009.
(B) The U.S. troops will be still present on similar level comparing to October 2009.
(C) The U.S. troops will be still present but in decreased number comparing to October 2009.
(D) The U.S. troops will be completely withdrawn.
Answer:
[1st candidate: / 2nd candidate: / 3rd candidate]
Specify which sentence (number ID) from the prepared Future Prediction Support Sentences was most useful in making the above decision:
1st candidate: [ ]
2nd candidate: [ ]
3rd candidate: [ ]

Figure 3. An example of one multiple choice question from the 4th Future Prediction Competence Test with additional question inquiring which of the prepared automatically extracted sentences was most useful

gender distribution, age groups from university students(major in information technology and science) and to their fifties), who were told to read the FPSS and reply to questions asking them to predict the future in 1-2 years from now, or from the starting point of prediction.

No.	Keywords
Q1-1	Participation in regional government by foreigners with permanent resident status | Participation in regional government permanent resident alien
Q1-2	Husband and wife retaining separate family names
Q2	Midterm elections | (Republic Party | Democratic Party) & (United States | America)
Q3	Afghanistan
Q4	Analog broadcasting | Digital broadcasting
Q5	Child allowance
Q6	(Democratic Party | Liberal Party | Ruling Party) & election

Table 4. Keywords for collecting future reference sentences (FRS) each the questions

The questions were taken from the Future Prediction Competence Test (FPCT, Japanese: Senken-ryoku Kentei), released by the Language Responsibility Assurance Association (LRAA, Japanese: Genron Sekinin Hoshōkyōkai)[5], a nonprofit organization focused on supporting people of increased public responsibility (managers, politicians) and people responsible of making decisions influencing civic life. Such people often need to perform public speeches in which they reveal details or opinions regarding future events. In such situations they are obliged to express some contents (e.g., objective facts), while restraining from revealing others, (one's fears towards the future or negative thoughts, disturbing public opinion, etc.). Thus the association helps preparing one's public speeches and responsibility bound presentations.

The FPCT is an examination that measures prediction abilities in humans regarding specific events that are to happen in 1-2 years in the future. It has been initiated in 2006 and from that time it has been performed six times. The test consists of various questions, including multiple choice questions (e.g., "Will US Army contingent in Afghanistan increase or decrease during next year?"), essay questions (e.g., "Describe economic situation of a country after next two years"), and questions that must be answered using numbers (e.g., "What will be the exchange rate of Japanese Yen to US Dollar after two years"), and they are scored after those particular events have come to light.

The questions for the experiment to benchmark our future trend prediction support method were selected from the 4th of the past six FPCTs[6] as it had the largest total number of questions, and respondents, which would assure the

highest possible objectivity of the evaluation. Implemented in 2009,the 4th FPCT contained questions regarding predictions for 2010 and 2011, and the scoring was performed in 2011.Respondents were to choose to answer at least 15 questions from a total of 25 questions in six areas, namely, politics, economics, international events, science and technology, society, and leisure. The test contained a large number of multiple choice questions and several questions requiring predicting specific numbers. There was also a small number of questions requiring a written explanation of the reasoning for the prediction. When participating in the test, respondents can browse any and all materials, and are free to seek the opinions of others in answering the question, but the submission deadline was fixed and set at December 31st, 2009 (end of the year). The scoring is set at 90 total points on prediction questions and 30 total points for descriptive questions, with a total of 120 points.

The prediction support method we developed in this study applies FPSS related to a given question and provide assistance for humans on which answer to choose. Therefore for its evaluation we limited the questions to multiplechoice questions. Questions with two or more (multiple) choices were selected from the 4th FPCT and applied as questions for the experiment. For example, one of questions is shown in Figure 3.

4.2 Data Preparation

In this section we describe the process of data preparation for the experiment. Firstly, a total number of 7 multiple-choice questions are selected from the 4th FPCT. Next, the FPSS are collected from news corpus each questions to consider for answering. The original settings of the FPCT are that participants can refer to any information and take a one year to answer. In this experiment, however, the laypeople only read the provided FPSSs and answered the questions at the time of this experiment.

The FPSS for each questions are collected in the following steps. At first, we extracted from the Mainichi Newspaper's entire 2009 year all sentences related to the questions on the basis of topic keywords (Table 4). In preliminary study we applied several kinds of keywords on each questions, and after careful examination of the sentences, decided to use the noun keywords that appeared in the original questions or options. Where it was possible we added closely related keywords by hand to cover the widest scope of possible search topics that a normal user would perform if they wanted to search for newspaper articles related to a specific topic. In particular, for each question we used the topic keywords as represented in Table:4.

The sentences collected with keywords from the newspaper were analyzed by the proposed method (Section:3) using the fully optimized model trained on MoPs, and sorted in a descending order of the future-relevance score.

This way, the sentences that appeared on top of the list were in the highest probability to be FRS. We retained only those FPSS with scores larger than 0 and presented the highest 30 of them to the subjects in chronological order (date of publication). We decided to present the subjects FPSS in the order they appeared in newspapers instead of descending order of scores so that the subjects could have a better image of how the events unfolded, which would make the prediction more natural. We also decided to limit the number of sentences for the subjects to read to thirty sentences so that the subjects do not get bored or tired easily. However, we keep to represent the rest of the sentences in case the subjects insisted on further reading. Moreover, there are also situations in which the list of initial sentences extracted with topic keywords was less than thirty. In this case, we presented to the subjects all sentences which had a probability of being FRS.

In addition, for questions for which less than thirty FPSSs were extracted in general, we presented all of the sentences that were classified into FRS.

As an example, seven FPSSs for the Question 3 were represented below. The questions were answered directly after reading only the FPSSs. Additionally, the laypeople were asked to report the ID number of the FPSSs they referred to in their answer (or the FPSS that was the most informative and useful in their opinion).

[5] *http://genseki.a.la9.jp/kentei.html*
[6] *http://genseki.a.la9.jp/index.html*

We evaluated their answers based on the original scoring schema, explained as follows. Each of the questions 1, 2 , and 7 were allocated 3 points. Moreover, in questions 2-5 the laypeople were allowed to make up to three candidate choice answers: primary candidate, secondary candidate and tertiary candidate, allocated 3 points, 2 points and 1 point, respectively. Additionally, to make the evaluation more strict and objective, for comparison, we also used a different scoring, based strictly on only one point per question.

Examples (translated in English) of FPSS for Question 3. (ordered chronologically in the form [Month/Day] in 2009) are represented below.

[01/18] Other newspapers are also carrying out the Mainichi Newspaper's three-part feature reportage on trilateral coordination between Japan, Korea, and the US regarding North Korean nuclear arms, cooperation between Japan and Korea on reconstruction aid to Afghanistan, and the establishment of regular meetings or "shuttle diplomacy" between the respective leaders of these countries.

[01/21] Additionally, it revealed their intention to finish the Iraq War through the gradual withdrawal of US combat troops stationed there, and put full force into the War on Terror in Afghanistan.

[01/22] Substantial negotiations toward realizing the campaign pledge to reduce the number of stationed U.S. troops "within 16 months of inauguration" have begun, aiming for an early formulation of a comprehensive plan that includes sending more U.S. troops to Afghanistan, a key battleground in the War on Terror.

[01/22] Ahmad Saif (29), an engineer in Baghdad, rejoiced hearing that President Obama had reemphasized the need to focus on the War on Terror in Afghanistan, increasing the likelihood of an early withdrawal of U.S. troops from Iraq.

[02/07] At a cabinet-level meeting between Finance and Foreign Ministers of each country, in addition to the steps to be taken on the deterioration of public order in Afghanistan caused by formerly dominant Taliban forces, the agenda featured discussion on water resource development policies in response to the ongoing drought, and negotiations over assistance measures.

[02/26] At the conference, a U.S.-Japan joint investigation into strategies regarding Afghanistan was agreed upon, and a special envoy will be dispatched to the U.S. to settle the details.

[0307] On the 6th, the Russian Ministry of Foreign Affairs made an announcement suggesting that both countries share a stance on the condition in Afghanistan and the War on Terror, and that they are "mildly optimistic" about the results of the Foreign Ministers' talk.

4.3 Experiment Results

Thirty laypeople answered the future prediction task only reading the FPSS. We scored those answers adopting the weight calculation for multiple choice questions under the scoring procedure of the 4th FPCT. The result is shown in Table 5. At first, we summarize the scoring procedure. We set each question is worth up to 3 points with a total of

	correct	answer	accuracy
	average	highest	lowest
Original FPCT	0.3344	0.6111	0.0666
Our experiment	0.3810	0.6190	0.1428

Table 5. Performance of subjects in the future prediction task supported with the proposed method with comparison to original results of the 4th Future Prediction Competence Test

21 possible points. For questions with up to 3 choices possible, we awarded 3 points when the subjects' first choice was correct, 2 points when the second was correct, and 1 point when the third choice was correct. In case of 1 choice

possible, we awarded 3 points for the correct answer.

In the original 4th FPCT, the test was taken by eleven participants. In the performed experiment, the average score of this experiment was 38.10% (see Table 5). In comparison to the original FPCT, the average was 33.44%. Therefore the result of this experiment obtained slightly higher scores than the original FPCT. The questions for the prediction task in fact referred to the past events for the laypeople. Therefore the laypeople might have already known the unfolding of the events and it is possible that this influenced experiment results.

However, when the experiment was conducted, we advised the laypeople that when they answer these questions, they can only read the FPSS and do not apply any other knowledge about the predicted events.

Furthermore, in comparison with the original test results, an improvement of approximately 4.66 percentage-points with average was noticed. This can be considered as the contribution of our method. However, the greatest contribution of our method for future prediction support is the following. Even if we assume that the improvement was not sufficient, and that our subjects performed similarly to the original participants, it must be remembered that our subjects made their decision based only on about thirty specifically extracted FPSSs and were given only short time for decision, whereas the original participants had over one year time for preparing the answer, unlimited access to all available data and receiving help from any other people including experts.

This result improved the lowest score indicates that it is easier to predict future event unfolding with FPSS. It indicates that the future reference sentences are useful in questions regarding prediction of future events for people have no knowledge regarding those events.

Furthermore, the FPCT has an established ranking system based on scores achieved by participants. In the 4th FPCT, if a participant got a final score covering over 60% of a maximum number of points, she was assigned the 1st grade future predictor. Correspondingly, participants were given 2nd grade if their scores covered from 50% to 60% of maximum number of points, and 3rd grade when their scores covered from 40% to 50% of the maximum. This refers to the grade of competence a participant is said to have when it comes to prediction of future unfolding of events. In the original 4th FPCT, 18.18 % of participants earned 1st grade, nobody earned 2nd grade, and 18.18 % of them earned 3rd grade. If the same grading system is applied to the participants of our experiment, 20.00 % of experiment participants would earn the 1st grade, 20.00 % of them would earn the 2nd grade and 13.33 % of them would earn the 3rd grade. This means that the successful participants supported with our method would outnumber the original successful participants of the 4th FPCT.

Hence, we can say that supporting the future prediction task with FPSS for specific topics can be just as efficient as collecting all available information by oneself through one year time.

5. Discussion

In this section, we discuss the effectiveness of FRS for future trend prediction while comparing in detail experiment results with the FPCT.

As shown in Table 5, if we look at the accuracy of the 4th FPCT, the average rate was 33.44%, demonstrating that when people have every means at their disposal, they still only accurately predict the future on around one third of the time. Kurokawa and Kakeya (2009) [12] analyzed trends in the answer results of the 1st FPCT and verified whether the idea of such collective intelligence is useful or not in the context of future prediction. The average rate in the 1st FPCT was 33.17 %. Moreover, Kakeya et al. concluded that the collective intelligence is not effective when it comes to future prediction. This means that predicting future trends is not an easy task for normal people, even when they have access to all available resources.

On the other hand, the average rate of the proposed method was 38.10%, an improvement of 4.66 percentage-points over the average of the original test. Furthermore, a consideration of the certification breakdown from 1st grade to 3rd grade shows that only one third of all FPCTs got a certification, while if the laypeople of our experiment using FPSS,

the participants in the original test, over half of them would get the certification (reaching over 60% of maximum points). Thus, it is evident that when people try to predict future events, FRS not only greatly reduce time and effort of collecting information, but allow achieving above-average predictive accuracy. Therefore we using FRS to support future trend foresight can be considered both effective and efficient.

Next, we considered the FPSS referred to the most useful by the laypeople when answering the questions. As an example, Figure 4 shows the number of FPSS referred to in Question 3. Gray bars indicate the numbers of FPSS referred to by successful laypeople, while white bars indicate the numbers of sentences referred to by laypeople who failed the task of prediction.

The contribution of these FPSS to choosing correct answers can be analyzed by focusing on gray bars. It is possible that differences in prediction accuracy depend on which of the 30 FPSS were referred to. Taking Question 3 as an example, we analyzed both the content of statements that were only referred to in incorrect answers as well as those that also contributed to correct responses. The values on the horizontal axis of Figure 4 correspond to FPSS's numbers. In the experiment, 83.33% of responses to Question 3 were accurate. FPSS consisted of twelve sentences referring to increase of the number of U.S. troops, three neutral sentences and one sentence referring to decrease of the number troops. Rest of FPSS mentioned other countries' support to Afghanistan or referring to past events, etc. Three examples of the FPSS that contributed to the answers for Question 3 were as follows:

(a) The most Useful FRSS Leading to Correct Answer
[$No.10$] "On the other hand, they announced plans to send reinforcements to Afghanistan, demonstrating their intention to shift the central focus of the War on Terror there, while staging a withdrawal from Iraq."

(b) One of Useful FRSS Leading to Both Correct and Incorrect Answer
[$No.22$] "Furthermore, mentioning that 'the United States are emphasizing cooperation in Afghanistan as well,' they indicated their plan to prioritize the establishment of the new Afghan assistance measures."

(c) One of FRSS Leading to Incorrect Answers
[$No.16$] "The prime minister Hatoyama also brought up the policy of improving civilian assistance to Afghanistan as an alternative to the Maritime Self-Defense Force's refueling mission in the Indian Ocean."

All laypeople who referred to the sentence (a) were correct. The sentence (a) contributed the most to leading the correct answer. The sentence (b) contributed to both sides, correct and incorrect, for four laypeople each. Such sentence will lead to correct answer depending on its interpretation. In this case, it would be required to thoroughly understand the ocntext of the sentence to get the correct answer. The sentence (c) states that a Prime Minister of the other country (here: Japan) expressed the will to support in the case of Afghanistan, but it does not state directly about increasing the number of troops, therefore basing ones decision on this sentence might not be effective.

The FPSS as shown in examples (a)-(b) and examples of FPSS for Question 3 in Section 4.2 can be classified into two kinds of FPSS, namely, those referring to the future event directly or indirectly. It could be possible to eliminate the problems described above, regarding sentences (b) and (c), by classifying the FPSS into directly or indirectly mentioning the event before providing them to the users. The sentences referring tho the future indirectly could be useful as a support for the directly referring sentences. Finally, we analyzed the keywords used in collecting FRS from corpora. The examples of FPSS extracted for this question presented in previous section indicate that although all sentences contained the keyword "Afghanistan", some sentences also contained references to U.S. troops (see example [01/22] in Section 4.2), whereas others [02/07] contained the word "Afghanistan" but did not refer to the troops. Therefore, in order to improve the prediction accuracy, it is necessary to devise a better keyword setting for selecting FPSS from newspaper corpora.

6. Conclusions and Future Works

In this paper we conducted a validation experiment to determine whether Future Reference Sentences (FRS) are effective in supporting future trend prediction. In the experiment we applied questions from the official Future Prediction

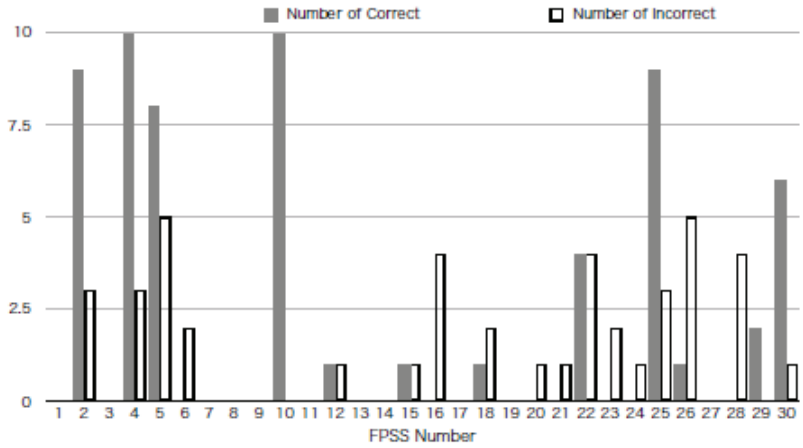

Figure 4. FPSS referred to by the laypeople as most useful with comparison between successful and unsuccessful

Competence Test (FPCT) and, using topic keywords from those questions, gathered newspaper articles from the entire applicable year (2009). Then we extracted Future Prediction Support Sentences (FPSS) from those articles, and had thirty laypeople read these sentences to make predictions regarding unfolding of the events. The results yielded an average of 4.66 percentage point improvement over the results of the participants of the original FPCT. However, the original test allowed respondents to prepare their answers for one whole year and use any available information sources, as well as seek the opinions of others, including experts. On the other hand, the subjects of our experiment replied immediately after reading the provided support materials, which consisted of only thirty FPSS. Therefore, although further experiments are needed, we can say that within the scope of present experiment, the significance of obtained results for prediction support has been sufficiently demonstrated.

In the experiment, only separate FRS were extracted for support from whole articles. However, the experiment showed that if such sentences are extracted with accurately set topic keywords they yield very detailed information sufficient to make the prediction. Furthermore, we believe that a combination of the proposed method with inference from statistical data would further increase the potential for obtaining information useful for future trend predictions.

In the future, we plan to use this method with other corpora to conduct experiments on real-world problems, such as company management support or economic trend prediction. Carrying out a chronological analysis of FRS and the addition of other technologies, such as sentiment analysis, could lead to the discovery of new knowledge. We also plan to take part in the next FPCT to check if the proposed method could allow obtaining a certificate in the task of future prediction.

References

[1] Alonso, Omar., Strotgen, Jannik., Baeza-Yates, Ricardo., Gertz, Michael. (2011). Temporal information retrieval: Challenges and opportunities. TWAW, 11. 1-18.

[2] Aramaki, Eiji., Maskawa, Sachiko., Morita, Mizuki. (2011). Twitter catches the flu: Detecting influenza epidemics using Twitter *In*: Proceedings of the Conference on Empirical Methods in Natural Language Processing, p. 1568-1576.

[3] Baeza-Yates, R. (2005). Searching the future. *In*: SIGIR Workshop on MF/IR.

[4] Fellbaum, Christiane., Osherson, Anne., Clark, Peter E.. (2009). Putting semantics into wordnet's morphosemantic links. Human Language Technology. Challenges of the Information Society. 5603. 350-358.

[5] Jatowt, Adam., Kanazawa, Kensuke., Oyama, Satoshi., Tanaka, Katsumi.(2009). Supporting analysis of future related information in news archives and the web. *In*: Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries, p. 115-124.

[6] Jatowt Adam., Au Yeung, Ching man. (2011). Extracting collective expectations about the future from large text collections. *In*: Proceedings of the 20th ACM international conference on Information and knowledge management, p. 1259-1264.

[7] Jatowt, Adom., Kawai, Hideki., Kanazawa, Kensuke., Tanaka, Katsumi., Kunieda, Kazuo., Yamada, Keiji. (2013). Multilingual, longitudinal analysis of future-related information on the web. *In*: Proceedings of the 4th International Conference on Culture and Computing 2013, p. 27-32. IEEE Press.

[8] Kanazawa, Kensuke., Jatowt, Adam., Oyama, Satoshi., Tanaka, Katsumi. (2010). Exracting explicit and implicit future related information from the web(o) (in Japanese). *In*: DEIM Forum 2010, p. paper ID: A9-1.

[9] Kanazawa, Kensuke., Jatowt, Adam., Tanaka, Katsumi. (2011). Improving retrieval of future-related information in text collections. *In*: IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), p. 278-283.

[10] Kanhabua, Nattiya., Blanco, Roi., Matthews, Michael. (2011). Ranking related news predictions. *In*: Proceedings of the 34th International ACM SIGIR Conference on Research and development in Information Retrieval, p. 755-764.

[11] Kroeger, Paul. (2007). Morphosyntactic vs. morphosemantic functions of indonesian -kan. In A. Zaenen, J. Simpson, T. H. King, J. Grimshaw, J. Maling, and C. Manning, editors, Architectures, Rules, and Preferences: Variations on Themes of Joan Bresnan, p. 229-251. Stanford, CA: CSLI Publications.

[12] Tomoya Kurokawa, Hideki Kakeya. (2009). Analysis of a tendency to answer in foresight official approval (in Japanese). *In*: The Fifth media information inspection arts and sciences meeting, p. 50-55, 2009.

[13] Levin, Beth., Hovav, Malka Rappaport. (1998). Morphology and lexical semantics. In eds, editor, The handbook of morphology, p. 248-271. Oxford: Blackwell.

[14] Nakajima, Yoko., Ptaszynski, Michal., Honma, Hirotoshi., Masui, Fumito. (2016). Application of future sentence reference extraction in support of future event prediction. *In*: Proceedings of Workshop on Language Sense on Computer (IJCA 2016), p. 81-88.

[15] Nakajima, Yoko., Ptaszynski, Michal., Masui, Fumito., Honma, Hirotoshi. (2016). Automatic extraction of future references from news using morphosemantic patterns with application to future trend prediction. ACM Summer Newsletter. *AI Matters*, 2 (4) 13-15.

[16] Nakajima, Yoko., Ptaszynski, Michal., Masui, Fumito., Honma, Hirotoshi. (2016). A method for extraction of future reference sentences based on semantic role labeling. *IEICE Transactions on Information and Systems* 524.

[17] Popescu, Octavian., Strapparava, Carlo. (2013). Behind the times: Detecting epoch changes using large corpora. *In*: International Joint Conference on Natural Language Processing. p. 347-355.

[18]Ptaszynski, Michal., Rzepka, Rafal., Araki, Kenji., Momouchi, Yoshio. (2011). Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion. *International Journal of Computational Linguistics Research* 2 (1) 24-36.

[19] Radinsky, Kira., Davidovich, Sagie., Markovitch, Shaul. (2012). Learning causality for news events prediction. *In*: Proceedings of the 21st International Conference on World Wide Web, p. 909-918. ACM.

[20] Ida Raffaelli. (2013). The model of morphosemantic patterns in the description of lexical architecture. *Lingue elinguaggio* 12 (1) 47-72.

[21] Takeuchi, Koichi.,Tsuchiyama, Suguru., Moriya, Masato., Moriyasu, Yuuki. (2010). Construction of argument structure analyzer toward searching same situations and actions. IEICE Technical Report. 109 (390) 1-6.

[22] Ohsawa, Y., McBurney, P. (2003). Chance Discovery. Springer.