# A Study on the Extraction of Future Trend Information with Future Reference Sentences and its Application in Future Event Prediction


| | |
|---|---|
| Name | Yoko Nakajima |
| Student ID Number | 1371300010 |
| Superviser | Dr. Toshio Eisaka |
| | Dr. Fumito Masui |
| | Dr. Hiroshi Yamada |
| Date | March.2016 |

Manufacturing Engineering of Doctoral Program

Kitami Institute of Technology

# Abstract
(Japanese)

本論文の目的は，人間の知的活動を支援するための自然言語技術を用いた斬新な情報獲得技術を確立することである．

近年，情報通信技術や情報共有技術の発展により，膨大な量の情報を容易に利用できるようになった．これらの膨大な情報の分析や解析を行い，ビジネスや経済活動，災害支援，医療分野など様々な場面で，将来動向を見通したいという需要が高まっている．このような将来動向予測に関する研究では，統計的手法を用いた研究はされているが，自然言語を用いた研究はまだ少なく世界的にも注目され始めている．

これまでの自然言語を用いた将来予測に関する研究の多くは，文に含まれる時間参照情報 (e.g. 来年，明日) や，未来形を表す語句 (e.g. 〜でしょう，〜するつもりだ)，現在よりも先の時点への参照表現 (e.g. 会議は 2017 年 7 月に開催される．) などを手がかりとする手法や因果関係を用いた手法が用いられている．これらの手法では，将来イベントに言及している文であるにもかかわらず，それらの手がかり表現情報は含まれていないが，将来予測をするための重要な将来への言及情報を取得し損ねてしまうことが考えらえる．

また，人間が将来予測をする場合，現在の出来事と合わせて自身がそれまでに蓄積している過去の知識や経験に基づいて予測活動を行っているが，予測するための要素のひとつとして，専門家の見解や信頼性の高い将来へ言及している情報を得ることで，予測がしやすくなり，また，予測精度が向上することがある．

将来に言及する文（以下，将来言及文）は将来イベントの発生を示唆する重要な記述表現であることの他に,「専門家の知識」や「イベント発生までの経緯情報」など，そのイベントについての様々な関連情報を含んでいると考えると，予測活動に重要な知識となり得る．

また，Web 検索を駆使し将来イベントに関する情報を検索する際に，将来言及文をキー文として検索できれば，検索頻度で表示される検索結果以外の重要な情報を

取得できる可能性も考えられる．以上に述べた理由から，将来言及文は人間の将来予測活動において，重要な予測要素の一つになると考えた．

さらに，将来言及文を用いた将来予測支援が有効であるなら，ビジネスや経済活動，災害支援，医療分野などにおいての意思決定を容易にすることが可能となり，将来の計画・対策やアイディア創出などの応用に期待できる．

本研究では，二つのテーマについて有効性を確認した．一つは，将来言及文の表現形式を調査した結果を示し，将来言及文の新しい形態パターン（文章の形態情報と意味論の両方の特性から構成されるフレーム）により，新聞コーパスから将来言及文を自動的に抽出する手法を提案し，従来手法と比較することで有効性を確認した．二つ目は，本手法で分類をした将来言及文を用いて，将来言及文が将来動向予測支援に有効であるかの検証実験を行い，有効性を確認した．

本論文では，将来言及文の自動抽出手法と将来動向予測支援について議論する．

# Abstract
(English)

The purpose of this paper is to establish innovative information acquisition techniques based on natural language technology in order to support human activities.

Thanks to the popularization of the Internet as well as recent advances in information technology, research on language analysis and processing have gained an environment in which a large amount of textual information is available and ready to use in research. Accordingly, many groundbreaking research is being conducted in the fields of information extraction and data mining using neural networks and machine learning on such textual data.

In a variety of fields (e.g. business management, economics, disaster prediction and prevention, healthcare, etc.), there is a growing need for methods and techniques allowing to look into the future and to analyze such amount of data. Such studies, regarding future trend prediction, has often focused on statistical methods. On the other hand, natural language processing (NLP) methods has not been applied in such studies until now. However, applying NLP-based methods in trend prediction has recently started to gain on interest worldwide.

The majority of studies using NLP methods propose to use as a clue for trend prediction - temporal information (e.g. expressions such as "next year", "tomorrow," etc.), future expressions (e.g. " will ", " being to "), or a causal information ( " A happens because of B "). Consequently, regardless of whether a given text refers to future events or not, there is a risk that information crucial to prediction will not be acquired if the texts do not contain helpful representative information. Usually, when someone makes a prediction about the future, it is based upon that person's knowledge and past experience relevant to the event. By adding the opinions of experts and highly reliable information about the future, it is possible to further

improve the accuracy of the prediction.

Future reference sentences are significant in the description of expressions referring to future events. Moreover, future reference sentences can be useful as a significant knowledge base because they include various related information, such as background information regarding the event in question, which is also used as knowledge source by experts.

Furthermore, we assume such sentences could be helpful in obtaining useful information other than word occurrence frequencies, applicable in future trend prediction. For example, it could be useful to retrieve future reference sentences with key phrases when retrieving information for future event prediction from the web.

As mentioned above, we considered future reference sentences as one of the significant factors for future prediction.

Moreover, if future prediction support using future reference sentences is valid, it could also help facilitating decision making process in business, economic activities, disaster prediction and prevention support, and medical field.

Firstly, I investigate future reference sentences in newspapers and Web news. I propose a novel method for extraction of such sentences using automatically obtained patterns consisting of semantic role labels and morphological information. I perform a series of experiments, in which I first extract future reference expressions from sentences using a novel algorithm for automatic extraction of sophisticated sentence patterns. Then I verify the validity of such patterns by applying them in classification of future referring sentences. I use the optimized classifier to retrieve new future-referring sentences from the Web. The results show that it was possible to fully automatically retrieve future sentences with performance significantly higher than state of the art.

Secondly. I perform an experiment for supporting future trend prediction using future reference sentences automatically obtained by using the proposed method. In addition, I perform an experiment using a prototype method for purely automatic

future trend prediction. The results of the verification experiment show that the accuracy was higher than prediction without future reference sentences. The experiments confirmed that future reference sentences are effective in forecasting support, and confirmed the validity of the proposed method.

I discuss the method for automatic extraction of future reference sentences and its application in supporting future trend prediction.

# List of Publications

## (1) Journal Papers

1. Yoko Nakajima, Michal Ptaszynski, Hiroshi Honma, Fumito Masui, *A Method for Extraction of Future Reference Sentences Based on Semantic Role Labeling*, IEICE Transactions, Vol.E99-D, No.2, Feb. 2016. (to appear)

## (2) Oral Presentations

1. Yoko Nakajima, Michal Ptaszynski, Hirotoshi Homma, Fumito Masui:, textittExtracting References to the Future from News using Morphosemantic Patterns, Proceedings of the International Workshop on Chance Discovery, Data Synthesis and Data Market in the 24th International Joint Conference Artificial Intelligence (IJCAI-15), pp.45-52, 2015.07, Buenos Aires, Argentina

2. Yoko Nakajima (Advisors: Fumito Masui, Hiroshi Yamada, Michal Ptaszynski, Hirotoshi Honma), textitAutomatic Extraction of References to Future Events from News Articles Using Semantic and Morphological Information, IJCAI 2015 Doctoral Consortium, p4385-4386, 2015.07, Buenos Aires, Argentina.

3. Yoko Nakajima, Michal Ptaszynski, Hirotoshi Honma, Fumito Masu, textittExtraction of Future Reference Expressions in Trend Information, The 24th Fuzzy, Artificial Intelligence, Neural Networks and Computational Intelligence, (FAN2014), in Kitami Institute of Technology, 2014.

4. Yoko Nakajima, Michal Ptaszynski, Hirotoshi Honma,Fumito Masui, textittInvestigation of Future Reference Expressions in Trend Information, AAAI Spring Symposium, Stanford University, SS-14-01, pp.31-38, 2014.

# Acknowledgment

First of all, I would like to express my sincere gratitude to my principal superviser Associate Professor **Fumito Masui** of Kitami Institute of Technology (KIT) for his academic advice and kind guidance during this work. He has had a profound influence throughout my academic career. He has provided sufficient space for me to expand my interest in this work, leaded me to the intrinsic properties of every problem, and enlightened me with systematic methods to do research. He has given scrupulous care on every difficulty in finalizing the work, and provided precious suggestions in the publications. Without any of them it would probably not be possible to complete this thesis.

I would like to express my sincere gratitude to my super adviser Assistant Professor **Michal Ptaszynski** of KIT for his academic advice and kind guidance and during this work. For encouraging me to dream, for his passion for research, for intriguing me with thoughts and ideas that I never had before, and, eventually, for helping me reach my research goals.

Again, I show my gratitude to my supervisors.

I would like to thanks to the main supervisor: Professor **Toshio Eisaka**, co-supervisor: Professor **Hiroshi Yamada** , Associate Professor **Michiko Watanabe** and Associate Professor **Yasunari Maeda** for taking time from their extremely busy schedules to serve on my dissertation committee.

I would like also to express my sincere thanks to Associate Professor **Hirotoshi Honma** of National Institute of Technology (NIT), Kushiro College, for his helpful support on my research and for his nice friendship.

I wish also to extend my great thank to Mr. **Shinya Kumamoto** for his helpful support on environment of the computer system for my work.

I would like also to express my sincere thanks to Dr. **Nobuo Takahashi** of the president of KIT Dr. **Kouichi Ayuta** of past president of KIT and Dr. **Takeshi**

**Kishinami** of past president of NIT. They give many chances for my research.

Many thanks are to **my friends and colleague of the Text Information Processing Laboratory**.

My special thanks go to my colleague: Dr. **Noriyuki Ootsuk**i of NIT for his crucial help, and for his support during my research.

Also, my special thanks to go to Professor **Edward Lisowski** and Mr. **Pawel Lempa** of Poritechnika Krakowska for taking care of me for my research at PK.

Finally, and most importantly, thanks to my parents for their helpful support.

# Contents

# List of Figures

# List of Tables

# Chapter 1
# Introduction

## 1.1 Background and Goal on This Work

In recent years, obtaining large-scale data from Web pages and newspaper articles has required much less effort. Thus, the amount of research actively developing and discussing the technology to analyze such data has increased rapidly. Large-scale data is of high interest for trend prediction, due to containing large amounts of trend information. Trend information is the data from which one can derive hints about the possible unfolding of certain events. The most common association would be with the prediction of stock trends, but the idea of trend information extends to everyday information, and predicting the outcomes of specific events does not require any special abilities in everyday newspaper readers.

For example, if we obtained a hypothetical fact that "the President of the USA is considering paying a state visit to Egypt" and a later one stating that "a revolution has started in Egypt," I could reasonably predict that the President will postpone or cancel the visit. This kind of future prediction is a logical inference which people

experience every day when reading news articles. As another example, if one reads an article in which it is stated that a country is expected to draw up a relaxation of economic law, one could predict that the country's situation could change for the better in the future. Similarly, if one reads an article about releasing a new product, one could predict that, if the product sells well, the finances of companies involved in producing parts for the product will also improve.

Moreover, future reference sentences are included implicit such information.When people predict possible future unfolding of an event in everyday life, they sometimes refer to the opinions of experts. Such experts presents their views based on their experience and expertise. These expert views are usually expressed using sentences referring to the future, containing important information supported by experience and survey. In my research I decided to automatize this process and develop an automatic expert system for the support of future prediction. We assumed that future referring sentences, when efficiently extracted from a credible source, should become useful in the process of future prediction. For example, sentences referring to the future contained in Diet members' reports will suggest trends in politics and in general development of the country's policies for average citizens.

Below are some examples of sentences concerning energy problems that were published in newspapersAll sentences appeared in paper version of Japanese daily newspaper Hokkaido Shinbun. Translation by the authors.

Science and Technology Agency, the Ministry of International Trade and Industry, and Agency of Natural Resources and Energy conferred on the necessity of a new system, and decided to set up a new council. From the comparison of energy sources available in 1992 and 2020, I expect the use of coal and oil to decrease greatly. We aim at practical use of "Space Solar Photovoltaics" to generate electricity with solar energy until the beginning of the 21st century.

All of the above sentences refer to the future. The first sentence claims the country will construct a new energy system. Interestingly, despite sentence is written with

the use of a past tense ("conferred", "decided") the sentence itself refers to future events ("setting up a new council"). The second sentence implies that when a new energy source is developed the use of coal and oil will greatly decrease. The third sentence provides a hint that space solar energy will be put into practical use.

When I read the contents of a newspaper article, I can imagine other events unfolding according to it. As mentioned above, a sentence referring to the future must contain the information relating it to the specific event that may happen in the future. However, the future reference in the sentence does not necessarily have to occur on the surface level (lexis) nor grammar (syntax/morphology). I perform an investigation of such future referring sentences with an assumption that such sentences consist of a variety of patterns which could be represented both morphologically and semantically.

The prediction of the future depends on various information (past events, current information, knowledge, experience, causal relations, etc.). A number of studies approached with natural language processing have been conducted on the prediction of future events with the use of causal relations (Radinsky et al. [1], Sakaji et al. [2] [3] )however, there has been no thorough study of actual future referring sentence and patterns they contain.

Therefore in this work I focused on future reference sentences and thoroughly analyzed expressions frequently appearing in those sentences.

In the study, I clarify effectiveness at future reference sentences and evaluate my proposed method for supporting future trend prediction by experiments.

## 1.2 Previous Research

Linguistically expressed references to the future have been studied by a number of researchers. Baeza-Yates [4] performed a study on about five hundred thousand sentences containing future events extracted from Google News[1] over the course of one day, and concluded that events mentioned in the news as those scheduled to take place, occur with almost perfect probability. A high correlation was also found between the reliability of occurrence and the time proximity of the event. Therefore, information about upcoming events is highly important in predicting future outcomes. Following this discovery, in my research I also chose the news as my data source. This will assure that if I extract the future sentences correctly, the events described in those sentences will have high probability of occurrence in reality.

According to the study of Kanhabua et al. [5], one-third of all newspaper articles contain some reference to the future. This also supports my choice of the news as my data source. In other research, Kanazawa et al. [6] extracted future implications from the Web using explicitly expressed future reference information. Alonso et al. [7] have indicated that time information included in a document enhances the effectiveness of information retrieval applications. Kanazawa et al. [8] focused on extracting unreferenced future time expressions from a large collection of text, and proposed a method for estimating the validity of the prediction by automatically searching for a real-world event corresponding to the predicted one. Jatowt et al. [9] studied the relation between future news written in English, Polish, and Japanese using keywords queried on the Web. Popescu et al. [10] investigated significant changes in the distribution of terms within the Google Books corpus and their relationship with emotion words across a wide time span.

Among the research regarding the retrieval of future information, Kanhabua et

[1] http://news.google.com/

al. [5] proposed a ranking model that takes into consideration the relevance of predictions. In terms of predicting the probability of an event occurring in the future and its relevance, Jatowt et al. [11] developed a model-based clustering algorithm for detecting future phenomena based on information extracted from a text corpus, and proposed a method of calculating the probability of the event happening in the future. In a separate research, Jatowt et al. [12] used the incidence rate of reconstructed news articles over time to forecast recurring events. They presented a technique for supporting the human analysis of future phenomena by applying a method based on the summarization of future information included in documents. Aramaki et al. [13] used Support Vector machine-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. Radinsky et al. [1] proposed the Pundit system for the prediction of future events in news. Their method used causal reasoning derived from a calculated similarity measure based on different existing ontologies. Also, Sakaji et al. [14] [2] proposed the extraction method of causal knowledge from newspaper corpus. they applied their method to economic trends extraction Tanigjchi et al. [15]. However, as their approach is based on causality pairs, rather than specific future-related expressions, it is not able to cope with certain constructions, e.g., sentences containing causality expressions but referring to the past.

The above findings have lead us to the idea that by using expressions referring to the future included in trend reports (newspaper articles, etc.), I could be able to support the future prediction process as one of the activities of people perform everyday. Such a method would be applicable in corporate management, trend foresight, and preventive measures, etc. Also, as indicated in 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 [Ohsawa [16]], since it could for example provide hints for a company planning its

future investments.

The 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. However, although all previous methods have used future time information, none of them used more sophisticated expressions such as morphosemantic sentence patterns. Hence, a method using such expressions would approach the problem of future prediction from a new perspective and could contribute to the research of future information extraction.

## 1.3 Terminology

In this section, here is described terminology for this report.

### 1.3.1 Morphology

Morphologies are motivated by four considerations. The first, the discovery of regularities and redundancies in the lexicon of a language. The second, the need to make explicit the relationship between grammatical features and the affixes whose function it is to express these features. The third, the need to predict the occurrences of words not found in a training corpus. The last, the usefulness of breaking words into parts in order to achieve better models for statistical translation, information retrieval, and other tasks that are sensitive to the meaning of a text.

Thus morphological models offer a level of segmentation that is typically larger than the individual letter, sand smaller than the word.

For example, it show a morphological structure In Japanese below, Japanse: Tarou wa Hanako ni hana o ageta. English: Tarou gave Hanako flowers. morpheme for Japanese : noun particle noun particle noun particle verb auxiliary-verb mark

Morphological Analysis is one of the basic technologies of the natural language processing with a computer. This method uses the knowledge of the grammar of the object language (the rule set of the grammar) and the dictionary (the word list with information on the part of speech) as a source. In addition, it divides the sentence written by the natural language into the row of Morpheme (Minimum unit with the meaning in the language), and distinguishes each part of speech.

### 1.3.2 Semantics

This structure consists of a hierarchy of phrases, the smallest of which are the basic symbols and the largest of which is the sentence. The structure can be described by a tree with one node for each phrase. Basic symbols are represented by values stored at the nodes. The root of the tree represents the sentence. In this work, semantics are labels used verb argument structure thesaurus [Tkeuchi et al. [17]] based on the argument structure data (a semantic role and concept frame) of the predicate to perform natural language processing.

### 1.3.3 N-gram

In the fields of computational linguistics and probability, an $n$-**gram** is a contiguous sequence of $n$ items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words or base pairs according to the application. The $n$-grams typically are collected from a text or speech corpus. When the items are words, $n$-grams may also be called shingles [Broder et al. [18]].

An $n$-gram of size 1 is referred to as a "unigram"; size 2 is a "bigram"; size 3 is a "trigram". Larger sizes are sometimes referred to by the value of $n$, e.g., "four-gram", "five-gram", and so on.

$n$-gram models are widely used in statistical natural language processing. In speech recognition, phonemes and sequences of phonemes are modeled using a $n$-gram distribution. For parsing, words are modeled such that each $n$-gram is composed of $n$ words.

### 1.3.4 $k$-Fold Cross-Validation

$k$-fold cross validation is a common technique for estimating the performance of a classifier. Given a set of $m$ traning examples, a single run of $k$-fold cross validation proceeds as follows:

1. Arrange the training examples in a random order.
2. Divide the training examples into $k$ folds.
3. For $i = 1, \ldots, k$ do
   - Train the classifier using all the examples that do not belong to Fold $i$.
   - Test the classifier on all the examples in Fold $i$.
   - Compute $n_i$, the number of examples in Fold $i$ that were wrongly classified.
4. Return the following estimate to the classifier error:

$$E = \frac{\sum_{i=1}^{k} n_i}{m}$$

To obtain an accurate estimate to the accuracy of a classifier, $k$-fold cross validation is run several times, each with a different random arrangement in Step 1. Let $E_1, E_2, \ldots, E_t$ be the accuracy estimates obtained in $t$ runs. Define:

$$e = \frac{\sum_{j=1}^{t} E_j}{t}, \quad V = \frac{\sum_{j=1}^{t} (E_j - e)^2}{t-1}, \quad \sigma = \sqrt{V}$$

The esimate for the algorithm performance is an error of $e$ with standard-deviation of $\sigma$.

### 1.3.5 Kappa-value

In statistics, **inter-rater reliability**, inter-rater agreement, or concordance is the degree of agreement among raters. It gives a score of how much homogeneity, or consensus, there is in the ratings given by judges. It is useful in refining the tools given to human judges, for example by determining if a particular scale is appropriate for measuring a particular variable. If various raters do not agree, either the scale is defective or the raters need to be re-trained.

**Cohen's kappa** [19], which works for two raters, and **Fleiss' kappa** [20], an adaptation that works for any fixed number of raters, improve upon the joint probability in that they take into account the amount of agreement that could be expected to occur through chance. They suffer from the same problem as the joint-probability in that they treat the data as nominal and assume the ratings have no natural ordering.

### 1.3.6 Precision, Recall and F-Measure

In pattern recognition and information retrieval with binary classification, **precision** is the fraction of retrieved instances that are relevant, while **recall** is the fraction of relevant instances that are retrieved. Both precision and recall are therefore based on an understanding and measure of relevance.

In a classification task, the precision for a class is the number of true positives (i.e. the number of items correctly labeled as belonging to the positive class) divided by the total number of elements labeled as belonging to the positive class (i.e. the sum of true positives and false positives, which are items incorrectly labeled as belonging to the class). Recall in this context is defined as the number of true positives divided

by the total number of elements that actually belong to the positive class (i.e. the sum of true positives and false negatives, which are items which were not labeled as belonging to the positive class but should have been).

Precision ($P$) is defined as the number of true positives ($T_p$) over the number of true positives plus the number of false positives ($F_p$).

$$P = \frac{T_p}{T_p + F_p}.$$

Recall ($R$) is defined as the number of true positives ($T_p$) over the number of true positives plus the number of false negatives ($F_n$).

$$R = \frac{T_p}{T_p + F_n}.$$

Suppose a program for recognizing dogs in scenes from a video identifies 7 dogs in a scene containing 9 dogs and some cats. If 4 of the identifications are correct, but 3 are actually cats, the program's precision is 4/7 while its recall is 4/9.

**F-measure** is a measure of a test's accuracy. It considers both the precision $P$ and the recall $R$ of the test to compute the score.

$P$ is the number of correct positive results divided by the number of all positive results, and $R$ is the number of correct positive results divided by the number of positive results that should have been returned. The F-measure can be interpreted as a weighted average of the precision and recall, where an F-measure reaches its best value at 1 and worst at 0.

The traditional F-measure is the harmonic mean of precision and recall:

$$F = 2 \times \frac{P \times R}{P + R}.$$

The F-measure is often used in the field of information retrieval for measuring search, document classification, and query classification performance.

### 1.3.7 Argument Structure Analyzer

Developed by Takeuchi et al. [21], the argument structure analyzer (ASA) is a tool for understanding sentence structure based on dependency relationships within the sentence, which it identifies using semantic role labels. These are based on verb argument structure thesauri, and provided by the dependency analysis tool CaboCha[2].

The argument structure contains the dependency relationship between verb and noun, and describes the intersection of their action. For example, in the two sentences, "X arrests Y" and "To catch Y", one can see that the two verbs can be paraphrased the same way, and therefore can grant that they share an argument structure. To capture such a structure, it is necessary to obtain a description of the general relationship between verbs, and identify the dependency relationship within the sentence, the connections within compound nouns and between nouns (mainly the semantic role), as well as any idiomatic phrases. Of these, the clause dependency relationship analysis is performed with CaboCha. Additionally, I use a verb argument structure thesaurus published in previous research (4425 words) to assess the general relationship between verbs. As for idiomatic phrases, I use a dictionary of idioms, made by hand from a previously published corpus of idiomatic phrases. The argument structure assignment system uses all of these language resource tools to assign argument structures according to a rule base. As each sentence is input, CaboCha performs a dependency analysis, and furthermore, identifies the dependency relationship between the compound nouns.

Next, using the idiomatic phrase matching system prepared beforehand, the meaning of the verbs and semantic role of the nouns are taken as examples and assigned to verb and noun dictionaries, respectively. On the sentence "Mary sold the book

[2]http://taku910.github.io/cabocha

to John" , Mary is identified as the actor, the book as the object, and John as the "point person". Further, "to sell" is a verb, and is thus identified as producing a change of state (a change resulting from the subject's decision). The arrow shows the thread of the relationship. That is, whether re-statement of the sentence in other terms is possible, and only depends upon the words used to identify the actor, object, point-person, and state change in a frame.

## 1.4 Structure of the Article

This article is composed of four chapters.

Chapter 1 provides background and an outline of natural language processing research.

Chapter 2 describes a survey of future reference sentence-related expressions in newspaper articles. Also, it elucidates the grounds for the proposed method, and describes future reference sentences extraction and classification methods. It is difficult to discover when a sentence is implicitly referring to the future from future reference sentences morphology alone. Here I describe the morphosemantic method, which combines morphology and semantics to form the core of the methodology implemented in this study. The morphosemantic method assigns morphosemantic labels based on verb argument structure to sentences, and automatically extracts unique morphosemantic patterns from future reference sentences. Using this model to carry out text classification confirmed that the morphological future reference sentences and implicit future reference sentences are sufficiently extractable. Furthermore, I demonstrate the effectiveness of our proposed method through a comparison with the prediction results of another state–of–the–art method that only uses morphology.

Chapter 3 discusses the results of an experiment testing whether morphosemantic pattern structure extracted with a fully-optimized model (FOM) are effective in the prediction of future trends. This experiment demonstrates the effectiveness of using future reference sentences for future prediction by comparing this study's forecasting results with past results from the Language Responsibility Guarantee Corporation's Test of Foresight. We also present the future reference sentences that accurate predictors referred to, and carefully investigate the characteristics of those future reference sentences shown to be effective for future prediction. Furthermore, I discuss the results of the prediction accuracy evaluations of a prototype method that predicts future trends automatically.

Chapter 4 presents the general conclusions of this research. We carry out a comprehensive consideration of the research results proposed and validated in Chapters 2 and 3, and discuss the outcome of a certain set of research results. We also reveal challenges newly derived from the result of the evaluations, and allude to some methods for their solution.

# Chapter 2
# Future Reference Semantic Pattern Extraction Method

In this chapter, I describe our method for extracting semantic patterns from sentences.

## 2.1 Investigation in Future Reference Expressions

We performed a study of expressions which refer to a change in time in general or to the future in particular. The study has been performed by reading through articles from the following newspapers: *the Nihon Keizai Shimbun*[1], *the Asahi Shimbun*[2], *the Hokkaido Shimbun*[3]. I used all newspapers in their both paper and Web version. From the above newspapers I manually extracted from various articles 270 representative sentences which referred to the future. Next, on the sentences I man-

[1]http://www.nikkei.com/
[2]http://www.asahi.com/
[3]http://www.hokkaido-np.co.jp/

ually annotated future expressions. There were 70 time-related expressions and 141 unique future expressions (words, phrases, etc.) that were not time-related. We can assume that these which appear the most often could be said to have a characteristics of being used as future expressions.

Some examples of the expressions are represented in Table 2.1. There are two kinds of future-related expressions. First consists of concrete expressions which include numerical values, such as "year 2013", or "11 o'clock". Second is derived from grammatical information (verb tense, word order, particles, etc.), such as phrases "will [do something]", "the middle of a month", "in the near future", or particles *-ni* ("in, due, till", point of time), *-made* ("until", implied deadline for continuous action), or *-madeni* ("until", implied deadline for single action).

Moreover, if I consider sentences and their different representations (grammatical, semantic) as sets of patterns which occur in a corpus (collection of sentences/documents) I should be able to extract from those sentences new patterns referring to the future. As the basic theory I based my idea on, which considers both word-formation and semantics, was the theory of predicate-argument structure [Bresnan, 2001 [22]]. This theory embraces the synergy between the lexical information of a predicate and their semantic and syntactic properties. In practice this can be realized by representing a sentence using semantic role labels. The proposed method takes advantage of such sentence representation and further extracts implicit future reference patterns, not using hand-crafted lists of explicit future expressions or temporal expressions, as it was in previous methods.

Table2.1: Examples of future- and time-related expressions.

| **Type of expression** | **Number found** | **Examples; Y=year, M=month (usually appearing as numerical values)** |
|---|---|---|
| Time-related expressions | 70 | *Y-Nen M-gatsu kara* ("from month M year Y"), *kongo Y-nenkan ni* ("in next Y years"), *Y-gatsu gejun ni mo* ("late in year Y"), etc. |
| Future expressions | 141 | *mezasu* ("aim to") (11), *hōshin* ("plan to") (12), *mitooshi* ("be certain to") (9), *kentō* ("consider to") (9), *-suru* ("do") (76), *-iru* ("is/to be") (36), etc. |

## 2.2 Morphosemantic Patterns

In the first stage, all sentences included in the datasets (see section 2.5.1), are represented in **morphosemantic patterns** (MoPs).

The idea of MoPs has been described widely in linguistics and structural linguistics. For example, Levin et al. [23] 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, Kroeger [24] applied MoPs to analyze an Indonesian suffix *–kan*. Later Fellbaum et al. [25] applied MoPs to improve links between the synsets in WordNet. More recently, Raffaelli [26] used MoPs to analyze a lexicon in Croatian, a language rich both morphologically and semantically. In this research I used datasets in Japanese, and applied MoPs for the same reason. Using only one representation narrows the spectrum of analyzed information. Moreover, till now there has been no practical application of MoPs to solving real-world problems. In this paper I present the first attempt of this kind.

We generated the morphosemantic model using semantic role labeling with additional morphological information. Below I describe in detail the process of morphosemantic representation of sentences.

## 2.3 Semantic Role Labelling

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 I used **ASA**[4], a system, developed by Takeuchi et al. [21], which provides semantic roles for words and generalizes their semantic representation using an originally developed thesaurus. In particular ASA uses 4400 verbs and around 80 labels from Lexeed (basic word-meaning) database [Takeuchi et al. [27].Examples of labels ASA provides for certain words are represented in Table 2.2. An example of SRL provided by ASA is represented in Table 2.3.

[4]http://cl.it.okayama-u.ac.jp/study/project/asa

Table2.2: An example of semantic representation of words performed by ASA.

| **Surface** | **Semantic (Semantic role, Category, etc.) and grammatical representation** |
|---|---|
| *mezasu* ("aim to") | No change (activity)-action aiming to solve [a problem]-pursuit; Verb; |
| *hōshin* ("plan to") | Other;Noun; |
| *mitooshi* ("be certain to") | Action;Noun; |
| *kentō* ("consider to") | No change (activity)-action aiming to solve [a problem]-act of thinking;Noun; |
| *-suru* ("do") | Change-creation or destruction-creation (physical);Verb; |
| *-iru* ("is/to be") | Verb; |

Table2.3: An example of a sentence analyzed by ASA.

| **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* | [Patient] |
| 4 | *tegami o* | [Object] |
| 5 | *okuru darou* | [State_change]-[Place_change]-[Change_of_place(physical)] |
| 6 | *shūchū suru* | [State change]-[place change]-[change of place (physical)]-[movement towards a goal] |
| 7 | *Hokkaidō ni tai suru* | [Place] |
| 8 | *Mikata ga* | [Other] |
| 9 | *Kawari tsutsu aru* | [State change]-[change] |

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 I used a morphological analyzer MeCab[5] 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. Therefore, there often occurs a situation where a compound word is 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 I added a set of linguistic rules for specifying compound words in cases where only morpho-

[5]http://code.google.com/p/mecab/

logical information is provided. To optimize the method, I used a set of linguistic rules to specify compound words. The heuristic rules were hand crafted on the basis of present state of linguistic research regarding compound words in Japanese [Kobayashi et al. [28],Matsumoto et al. [29]].

Moreover, as it is shown on Table 2.3, some labels provided by ASA are too specific. Therefore in order to normalize and simplify the patterns, I 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 of no analysis by ASA perform compound word clustering for parts of speech (e.g., "International Joint Conference on Artificial Intelligence" → `Adjective Adjective Noun Preposition Adjective Noun` → `Proper_Noun`)

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

- 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 semantic role labeling 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]

## 2.4 Automatic Extraction of Frequent Patterns

Having all sentences represented in morphosemantic structure, I used SPEC (Sentence Pattern Extraction arChitecture), a system for extraction of sentence patterns [Ptaszynski et al. [30]]. SPEC is a system automatically extracting frequent sentence patterns distinguishable for a corpus (a collection of sentences). The patterns are defined in this paper in the following way.

> "A sentence pattern is any frequently occurring ordered non-repeated combination generated from elements of the sentence. When the elements are disjoint, the gap is marked by an asterisk ("*"). Sentence elements are defined as parts of the sentence specified by the process of sentence preprocessing and are consistent with available sentence representation selected by the user (e.g., SRL in Section 2.3)."

According to this definition, the system generates ordered non-repeated combinations from the elements of a sentence.

Firstly, the system generates ordered non-repeated combinations from all sentence elements. In every $n$-element sentence there is $k$-number of combination groups, such as that $1 \leq k \leq n$, where $k$ represents all $k$-element combinations being a subset of $n$. The number of combinations generated for one $k$-element group of combinations is equal to binomial coefficient, represented in equation 2.1. In this procedure the system creates all combinations for all values of $k$ from the range of $\{1, ..., n\}$. Therefore the number of all combinations is equal to the sum of all combinations from all $k$-element groups of combinations, like in the equation 2.2.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{2.1}$$

$$\sum_{k=1}^{n} \binom{n}{k} = \frac{n!}{1!(n-1)!} + \frac{n!}{2!(n-2)!} + ... + \frac{n!}{n!(n-n)!} = 2^n - 1 \tag{2.2}$$

Next, the system specifies whether the elements appear next to each other or are separated by a distance by placing a wildcard ("*", asterisk) between all non-subsequent elements. SPEC uses all patterns generated this way to extract frequent patterns appearing in a given corpus and calculates their weight. The weight can be calculated in several ways. Two features are important in weight calculation. A pattern is the more representative for a corpus when, firstly, the longer the pattern 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 normalized weight $w_j$ is calculated according to equation 2.3. Normalization is performed to make weights fit in range from +1 to -1, and is achieved by subtracting 0.5 from the initial score and multiplying the intermediate product by 2.

$$w_j = \Big(\frac{O_{pos}}{O_{pos} + O_{neg}} - 0.5\Big) * 2 \tag{2.3}$$

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 some that appear more than one time on both sides (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 in both sides (zero patterns, `OP`).

Moreover, a list of patterns will contain both the sophisticated patterns (with disjoint 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.

The SPEC system is trained on bipolar training data (e.g., future reference sentences vs. non-future reference sentences), and generates all patterns. Next, it classifies test data using the generated patterns. The performance of the whole system for classification of sentences into either future related or not is tested using a 10-fold cross validation.

The effectiveness of SPEC is certify applying to extracted patterns from affective sentences by Ptaszynski et al., 2014 [31], to the conversation analysis [Ptaszynski et al., 2014 [32]] and extracting patterns of harmful expressions for cyberbullying detection [Ptaszynski et al., 2015 [33]].

## 2.5 Evaluation Experiment

In this section, I describe experiments to verify whether the future reference pattern extraction method is effective.

### 2.5.1 Dataset Preparation

Firstly, I collected a thousand sentences at random from a corpus containing the following newspapers: Nihon Keizai Shimbun[6], Asahi Shimbun[7], and Hokkaido Shimbun[8].

Next, three people manually judged whether these sentences referred to the future or not. The agreement coefficient (multi-rater kappa-value) was 0.456, which indicates somewhat strong agreement between the annotators. We grouped the annotated sentences into three groups: (1) perfect agreement between all three annotators, (2) ambiguous sentences and (3) other sentences (non future referring sentences). From the collected 1000 sentences the group for which all three annotators agreed contained 130 sentences , the ambiguous sentences group contained 330 sentences and the "other" group contained 540 sentences.

From all collected sentences referring to future events (section 2.1) I randomly selected 130 sentences and manually collected another 130 sentences which did not make any reference to the future (describing past, or present events). Out of those sentences I created two experiment sets. The first one containing 100 sentences, with 50 future-reference sentences and 50 non-future-reference sentences (later called "set50"). The second one containing 260 sentences, also with equal distribution of sentences of the two types (later called "set130"). All sentences were represented in morphosemantic structure according to the procedure described in section 2.2. From

[6]http://www.nikkei.com/
[7]http://www.asahi.com/
[8]http://www.hokkaido-np.co.jp/

the sentences preprocessed this way I extracted pattern lists using the extraction procedure described in section 2.4.

### 2.5.2 Experiment Setup

We designed the experiment as a text classification task with the prepared datasets applied into 10-fold cross validation. The classification was performed as follows. Each test sentence was given a score calculated as a sum of weights of patterns extracted from training data and found in the input sentence (equation 2.4).

$$score = \sum w_j, (1 \geq w_j \geq -1) \tag{2.4}$$

The results were calculated using standard Precision, Recall and balanced F-score. However, if the initial collection of sentences was biased toward one of the sides (e.g., sentences of one kind are in larger number or longer), there will be more patterns of a certain type. Thus, using a rule of thumb in evaluation (e.g., fixed threshold above which a sentence is classified as either future-related or not) does not provide sufficiently objective view on results. Therefore I additionally performed threshold optimization to find which modification of the classifier achieved the highest scores. In the experiment 14 different versions of the classifier are compared under 10-fold cross validation condition. Since the experiment was performed on two datasets, I obtained overall 280 experiment runs. There were several evaluation criteria. Firstly, I looked at top scores within the threshold span. Secondly, I checked which version got the highest break-even point (BEP) of Precision and Recall. Finally, I checked the statistical significance of the results using paired $t$-test.

### 2.5.3 Classification Results

We compared Precision, Recall, and balanced F- score for the classification based on patterns and, additionally, on n-grams alone with semantic role labels.

Experiment results (F-score) for all classifier versions tested on set50 and set130 for models trained on n-grams and patterns are compared separately in Figure 2.1, 2.2, 2.3, and 2.4.Figure 2.1 illustrates the F-score for all classifier versions tested in the experiment on set50 for model trained on patterns. Figure 2.2 illustrates the F-score for all classifier versions tested in the experiment on set130 for model trained on patterns. Figure 2.3 illustrates the F-score for all classifier versions tested in the experiment on set50 for model trained on n-grams. Figure 2.4 illustrates the F-score for all classifier versions tested in the experiment on set130 for model trained on n-grams.

For set50, the F-score was generally around 0.67–0.71 for patterns, and around 0.67–0.70 for n-grams.The F-score for set130 was around 0.67–0.70 for patterns, and 0.67–0.69 for n-grams.The optimal threshold (from the range 1.0 to -1.0, with 0.0 in the middle) was around 0.0 or slightly biased toward 1.0, which means both sides of the training set were balanced or slightly biased toward future-related sentences. Figure 2.5 illustrates the F-score results for set50 when a list of patterns used in classification contained either all patterns with comparison to n-grams only. Figure 2.8 and Figure 2.9 show the Precision and Recall for patterns and n-grams, respectively, for set50. Figure 2.6 illustrates the F-scores result for set50 considering patterns and n-grams for the classifier with length-awarded zero deleted. Figure 2.7 illustrates the F-scores result for set130 considering all patterns and n-grams only.

Furthermore, I compared different versions of the classifier, including those in which the pattern list was modified by deleting either zero patterns or ambiguous patterns. We also verified which method of weight calculation was more effective, the one using normalized weights, or the pattern length-based method. Hence, I also examined the case of length-based weights with zero patterns deleted, and

length-based weights with ambiguous patterns deleted. We performed a t-test on the F-scores given by set50 and set130. The p-value was 0.566 for all patterns. This means that the differences between set 50 and set130 were not statistically significant, which is a positive result, since it proves that the performance of my method does not depend on the amount of learning data. The one-sided t-test value was 0.310, which also does not suggest any significant difference. We will discuss the differences in detail in the Discussion section in this chapter.

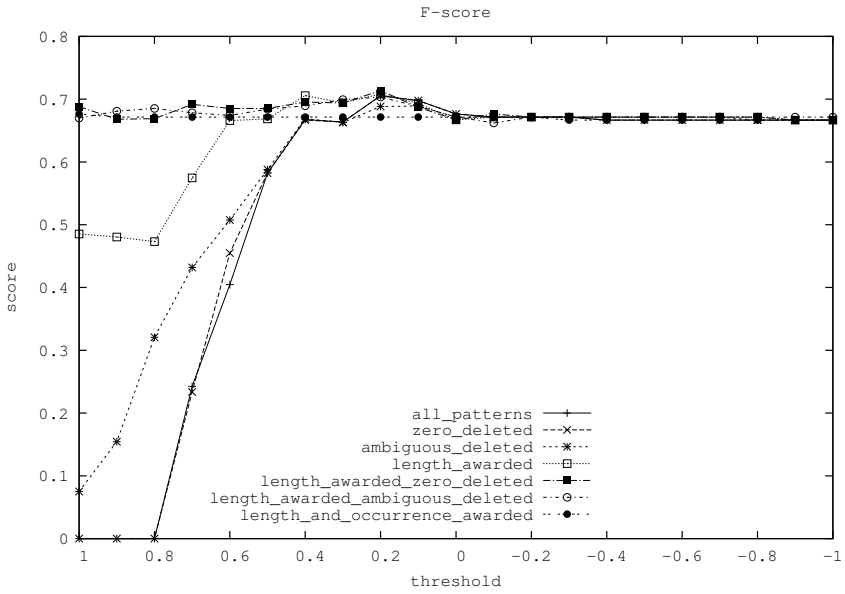


Figure2.1: Results (F-score) for all classifier versions tested in the experiment on set50 for model trained on patterns.

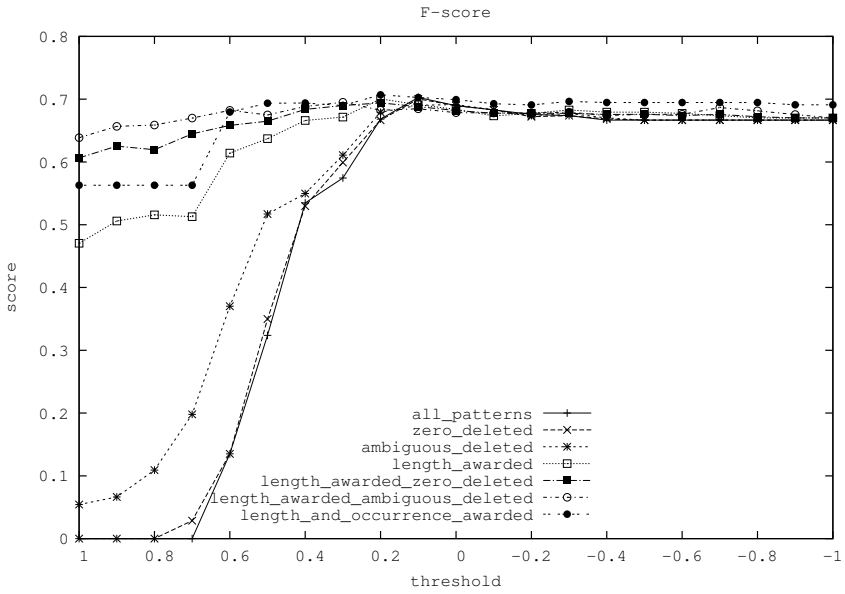


Figure2.2: Results (F-score) for all classifier versions tested in the experiment on set130 for model trained on patterns.

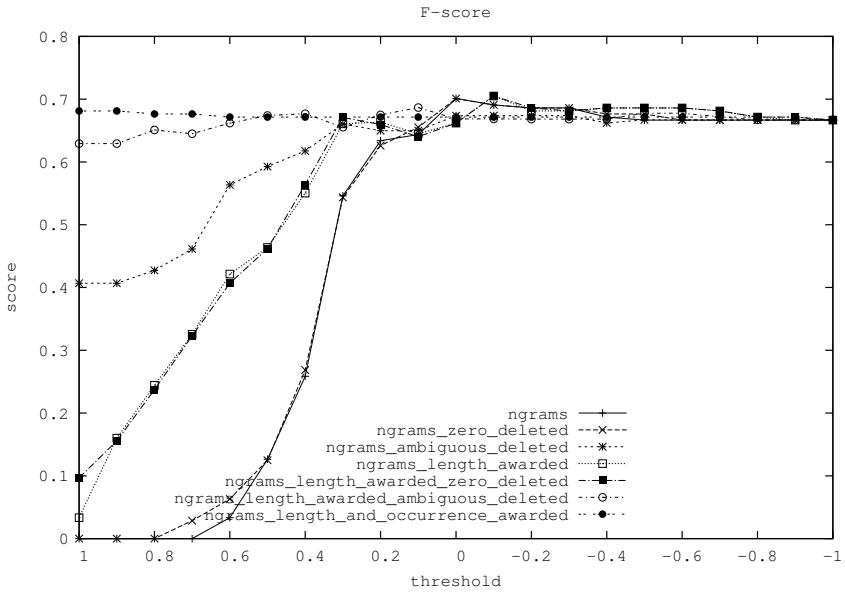


Figure2.3: Results (F-score) for all classifier versions tested in the experiment on set50 for model trained on n-grams.

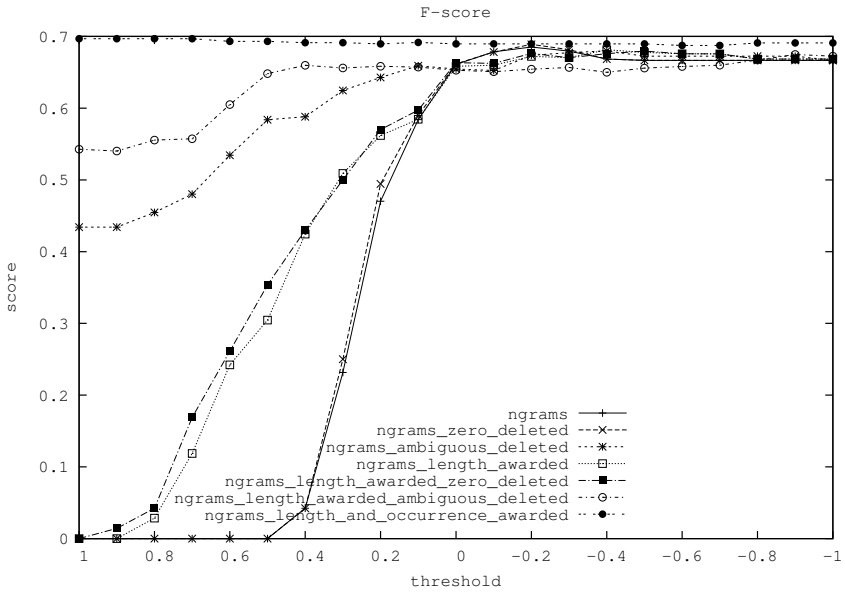


Figure2.4: Results (F-score) for all classifier versions tested in the experiment on set130 for model trained on n-grams.

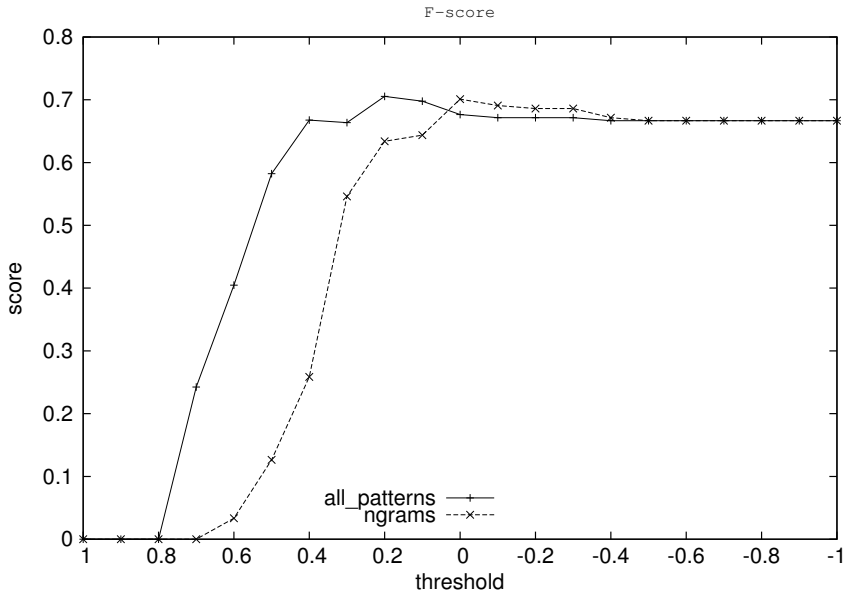


Figure2.5: Comparison of F-scores for set50 for all patterns and n-grams only.

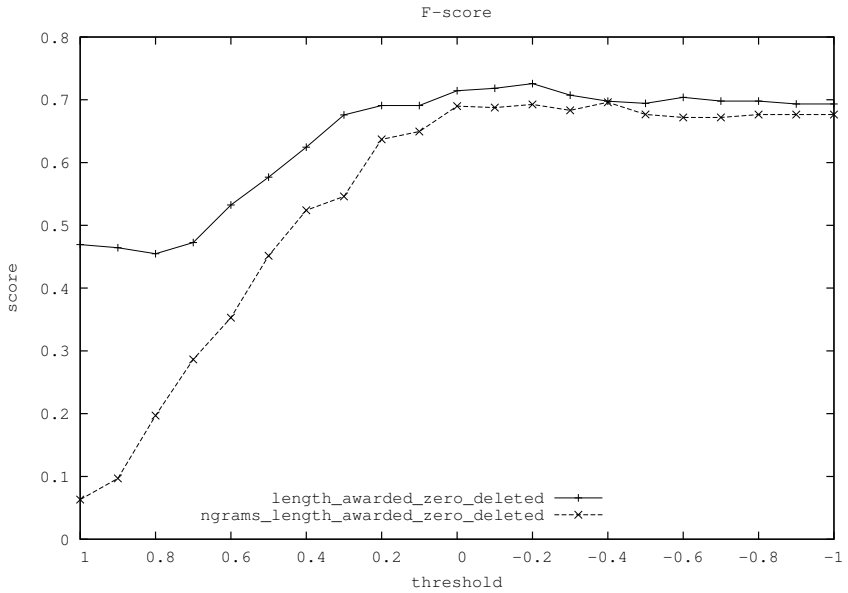


Figure2.6: Comparison of F-scores for set50 for patterns and n-grams for the classifier with length-awarded zero deleted.

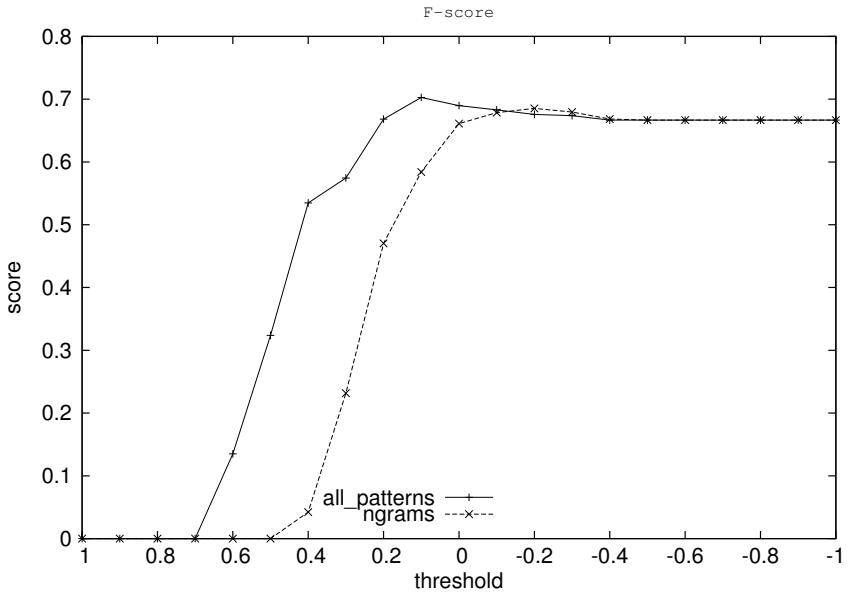


Figure2.7: Comparison of F-scores for set130 for all patterns and n-grams only.

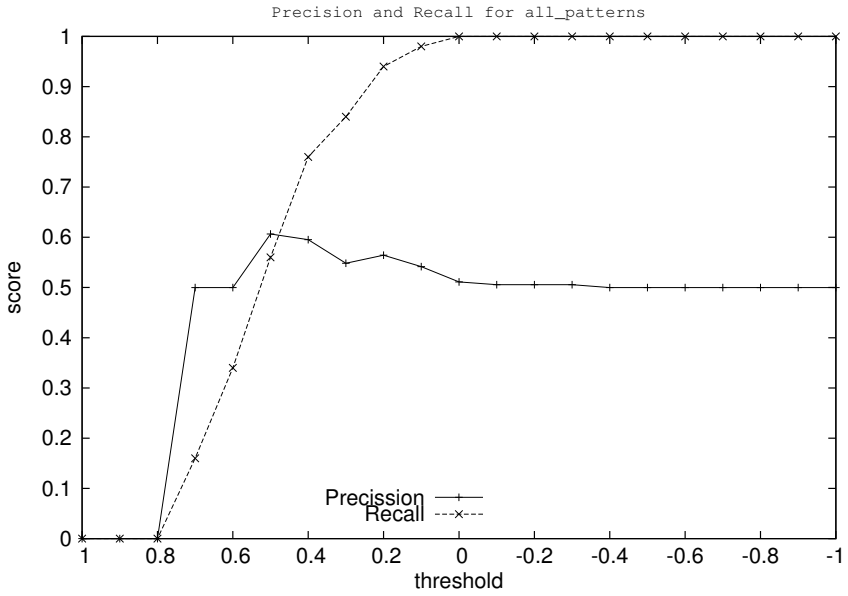


Figure2.8: Precision and Recall for all patterns in set50.

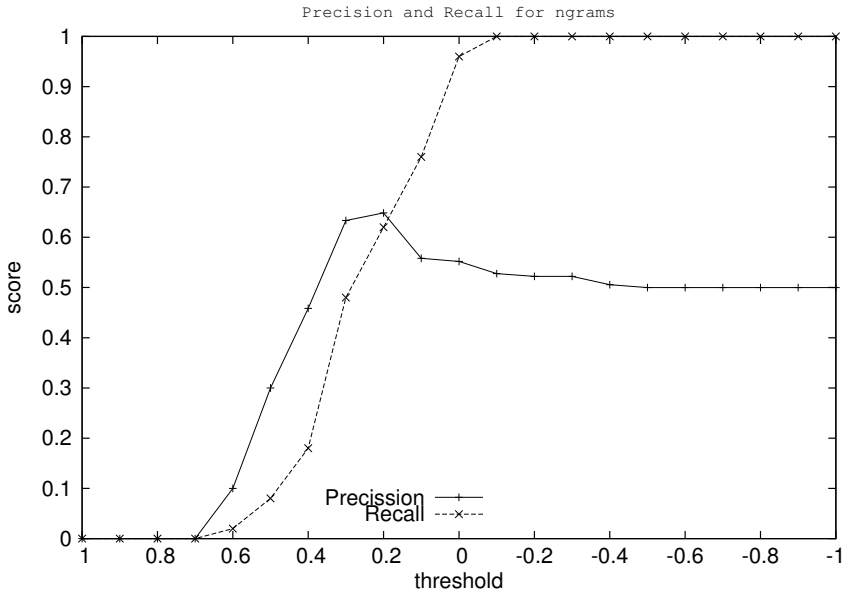


Figure2.9: Precision and Recall for n-grams for set50.

### 2.5.4 Analysis of Most Useful Future Reference Patterns

Besides the automatic classification results, I were also interested in the actual patterns that influenced the results. We extracted the most frequent unique future reference patterns and non-future reference patterns from set50. We obtained 1131 future patterns and 87 non-future patterns. Ten examples of both pattern types are given in Table 2.6.

Semantic role label patterns are grouped according to frequency of their appearance on each side, namely, in future reference sentences or in non-future reference sentences. For better comparison the pattern examples presented in Table 2.6 contain only non-ambiguous patterns which appeared in only in one of the sides.

The asterisk in some patterns means that the elements are disjoint. For example, the pattern [Action]∗[State change] contains two elements, [Action] and [State change], which appeared in the original sentences exactly in this order, and the asterisk indicates that there were other elements between these two. Each sentence pattern can appear either within a sentence, or on its edges (beginning, or end of the sentence). The method used for pattern extraction (SPEC) by the definition, does not make this additional distinction. This is due to the fact that sometimes a sentence can a start with a certain pattern, but in another sentence some words could precede this pattern. Making an additional distinction of sentence edges would force treating patterns which are actually the same as different ones only because of their facing the sentence edge or not. For one pattern this would produce four superficial combinations depending on the position of the beginning and the end of the pattern within the sentence (Edge-Inside, Inside-Edge, Edge-Edge, Inside-Inside). Thus although the four types would in fact represent the same one single pattern, its statistics would become dispersed to the four types.

### 2.5.5 Discussion

In this section, I present a detailed analysis of the results to facilitate better understanding of the extracted future reference patterns.

In general, the pattern-based approach obtained higher scores than the model trained on n-grams-only. This suggests that there are meaningful frequent patterns, more sophisticated than simple n-grams, in sentences referring to the future. In terms of modifying the pattern list and weight calculation, deleting the zero patterns does not appear to influence the results. A larger difference can be seen when all ambiguous patterns are deleted, and only patterns unique to each side are used. Moreover, the pattern length-based weight calculation always yielded better results. The highest scores of F = 0.71 with P = 0.56 and R = 0.98 were obtained using a pattern list with zero-patterns deleted and a length-based weight calculation. The greatest improvement provided by the use of patterns over n-grams is in Recall, which means that there are many valuable patterns omitted in the n-gram-only approach. Precision does not change significantly, oscillating around 0.55–0.60. For some thresholds, n-grams achieved similar or higher Precision. This means that the range 0.55–0.60 is the optimal maximum that could be achieved with the semantic representation used in this study. In the future, I plan to develop a modification that would improve the Precision without reducing Recall.

As well as comparing patterns with n-grams on the baseline classifier, I compared the results for five other cases (modifying the pattern list by deleting zero-patterns, or deleting all ambiguous patterns and modifying the weight calculation according to pattern length). In general, the highest F-score for patterns was 0.71, while for n-grams it was 0.70. (compare Figure 2.1 with 2.3, and 2.2 with 2.4.Although the difference is not that large, patterns usually achieve a high F-score because of superior Recall performance, even close to the threshold of 1.0 (compare Figure 2.5, Figure 2.8, and Figure 2.9). In Figure 2.6, the highest result of F-score was 0.70 for patterns, and 0.69 for n-grams. In this case the highest achieved F-score is

nearly the same between patterns and n-grams. However, patterns achieved better scores for each of the threshold. In case of F-score for set130 (see Figure 2.7), the highest result was also 0.7 for patterns and 0.69 for n-grams. However, the results for patterns are higher mostly within the threshold of 1.0 to 0.0, which confirms the results of set50. Since patterns provide better scores for most of the thresholds, I consider patterns as more effective. To thoroughly verify whether it is always better to use patterns, I need to conduct more experiments. However, from the present data, I can conclude that patterns generally produce better results.

Next, I compared the two datasets, **set50** and **set130**. The comparison in Table 2.5 shows that the results for each dataset did not differ greatly. However, when I look at Figure 2.10, the F-score for the classifier using a pattern list with all ambiguous patterns deleted performs slightly better than the other two (although the differences are not quite statistically significant with $p < 0.06$). Comparing these results to those in Figure 2.11 indicates that the performance is generally better when the pattern length is used to modify the weight calculation. In particular, both modified versions of the classifier (without zero-patterns and without ambiguous patterns) retain high F-scores across the threshold span (from 1.0 to -1.0). The same can be said of the results for set130. Comparing Figure 2.12 and Fig.2.13 also shows that the pattern length-based weight calculation yields better results within the specified threshold. Moreover, it is also advantageous to either exclude zero-patterns or all ambiguous patterns from the pattern list. It is also worth mentioning that the performance of the algorithm as a whole is similar for set50 and set130. In general, larger datasets contain more ambiguities, which can decrease the results. With the proposed approach, the differences in results are generally negligible (compare Figure 2.10 and Figure 2.12) or small (compare Figure 2.11 and Figure 2.13). Therefore it can be said that the method retains its performance regardless of the amount of data.

Table2.5: Comparison of the best results achieved (Precision, Recall, and F-score) for set50 and set130.

| **classifier version** | **set50** | | | **set130** | | |
|---|---|---|---|---|---|---|
| | Precision | Recall | F-score | Precision | Recall | F-score |
| unmodified pattern list | 0.56 | 0.94 | 0.71 | 0.58 | 0.90 | 0.70 |
| zero deleted | 0.56 | 0.94 | 0.71 | 0.57 | 0.90 | 0.70 |
| ambiguous deleted | 0.55 | 0.92 | 0.69 | 0.56 | 0.91 | 0.69 |
| length awarded | 0.58 | 0.90 | 0.71 | 0.58 | 0.89 | 0.70 |
| length awarded zero deleted | 0.56 | 0.98 | 0.71 | 0.57 | 0.87 | 0.69 |
| length awarded ambiguous deleted | 0.55 | 0.98 | 0.70 | 0.56 | 0.92 | 0.70 |

### 2.5.6 Inquiry into Extracted Future Reference Patterns

Using SPEC I were able to extract frequent patterns from sentences referring to the future and those not referring to the future. Each time a trained pattern was used during the classification, it was also added to a separate list of frequently used patterns. This extraction was performed for each fold in the 10-fold cross-validation. By taking the patterns extracted this way from all tests, and leaving only the frequent ones (used in classification at least two times across all experiment runs), I obtained a refined list of the most valuable patterns (those used most often). We investigated these patterns and the types of sentences in which they were used.

Below I present a number of example sentences used in classification. The information is provided in the following order: Romanized Japanese (transcribed in roman alphabet), English translation, and Semantic representation. The two first examples contain the following pattern: `[Action]*[Object]*[State change]` (pattern in question underlined).

**Ex. 1.** *Iryō, bōsai, enerugī nado de IT no katsuyō wo susumeru tame no senryaku-an wo, seifu no IT senryaku honbu ga 5gatsu gejun ni mo matomeru.* (IT Strategy Headquarters of the government will also put together in late May, the draft strategy for advancing the use of IT for health, disaster prevention, or energy.) [Action]-[Other]-[Other]-[No state change(activity)]-[State change]-[Artifact]-[Object]-[Organization]-[Agent]-[Noun]-[Time]- [State change]

**Ex. 2.** *Tonneru kaitsū ni yori, 1-nichi 50 man-nin wo hakobu koto ga kanō ni naru mitōshi de, seifu wa jūtai kanwa ni tsunagaru to shite iru.* (It is expected that the opening of the tunnel will make it possible to carry 500,000 people a day, which will lead to a reduction in traffic congestion, according to the government.) [Action]-[Time]-[Object]-[State_change]-[Other]-[Noun]-[Action]-[Organization]-[Action]-[Verb]-[State change]

The next examples contain a slightly different pattern, namely `[Object]*[Action]*[State change]`.

**Ex. 3.** *Nesage jisshi wa shinki kanyū-ryō, kihon ryōkin ga 12gatsu tsuitachi kara, tsūwa ryōkin ga 1996nen 3gatsu tsuitachi kara no yotei.* (The price cut implementation is planned to apply to the new subscription fees, for the basic rate plan from December 1, for call charges from March 1, 1996.) [Object]-[Action]-[Agent]-[Numeric]-[Time]-[Action]-[Time]-[Numeric]-[Time]-[State change]

**Ex. 4.** *Kin'yū seisaku wo susumeru ue de no kakuran yōin to shite keishi dekinai, to no mondai ishiki no araware to wa ie, kin'yū-kai ni hamon wo hirogesōda.* (Although they admitted that proceeding with the [new] monetary policy could become a disturbance factor and that it cannot be neglected, which showed an awareness of the problem, it still is likely to spread ripples in the financial world.) [Object]-[State change]-[Reason]-[Action]-[Action]-[Action]-[Agent]-[Place]-[Other]-[State change]

In the above examples, the patterns that were matched comprise those studied in previous research Kanazawa et al. [6], [8], Jatowt et al. [9]. These include time-related expressions ("late May," "from December 1," "from March 1, 1996") and future reference expressions ("is expected," "is planned to," "is likely to").

Next, I examined sentences containing non-future patterns. The following example sentence contains the pattern `[Numeric]*[Action]*[Action]`.

**Ex. 5.** *20man-ji no chōhen shōsetsu kara 2 moji dake wo kopī shite shōbai ni tsukatte mo ihō to wa ienai.* (It cannot be considered illegal to copy only two characters from a two-hundred-thousand-word-novel and use them for commercial purposes.) [Numeric] [Artifact] [Numeric] [State change] [No state change] [No state change] [Action] [Action]

The following example sentence contains the pattern `[Place]*[Place]*[No_state_change (activity)]`.

**Ex. 6.** *Nagata-ku wa Hanshin Daishinsai de ōkina higai wo uketa chiiki de, koko de wa Betonamu no hito ga kazu ōku hataraite iru.* (Nagata Ward, one of the areas that were greatly affected by the Great Hanshin Earthquake, is a place where many people from Vietnam are working.) [Place] [Organization] [adjective] [Other] [No state change(state)] [Object] [Place] [Agent] [Adjective] [No state change(action)]

The following example sentence contains the pattern `[Time]*[Noun]*[Role]`.

**Ex. 7.** *Sakunen 6gatsu, Kaifu ga Jimintō to tamoto wo wakatte aite jin'ei (gen Shinshintō) ni kumi shita toki mo, rinen to meibun ga hakkiri shinakatta.* (June last year, when Kaifu parted company with the Liberal Democratic Party and joined an opponent camp (now called New Frontier Party), their ideas and causes were unclear.) [Time] [Numeric] [Person] [Organization] [Noun] [State change] [Noun] [Organization] [Verb] [Role] [Place] [No state change(state)]

Example 5 contains the phrase *to wa ienai* ("it cannot be said/considered that"), which is labeled as an [Action] by ASA. This label is frequently used in future referring sentences, but this sentence is not classified as future-related. As for Example 7, although it contains time-related expressions ("June last year"), the use of sophisticated patterns that take the wider context into account allows correct disambiguation in this case. Furthermore, although this pattern contains a time-related expression, it is not listed as a future reference pattern. Thus, the presence of time-related information alone does not influence the classification. Instead, other elements of the pattern, such as the appropriate tense together with time-related expressions, constitute the pattern being distinguished as referring to the future.

Many future reference patterns had a high occurrence frequency (see Table 2.6), which means the sentences contain many of those patterns. Therefore, I can say that in general, "the future" has high linguistic expressiveness. For non-future reference patterns, the occurrence frequency was low, which suggests a large number of patterns, each used only once (thus, they were not included in the list of frequently used patterns). Because of this variety of patterns, there are no particularly distinctive patterns for sentences that are not referring to the future.

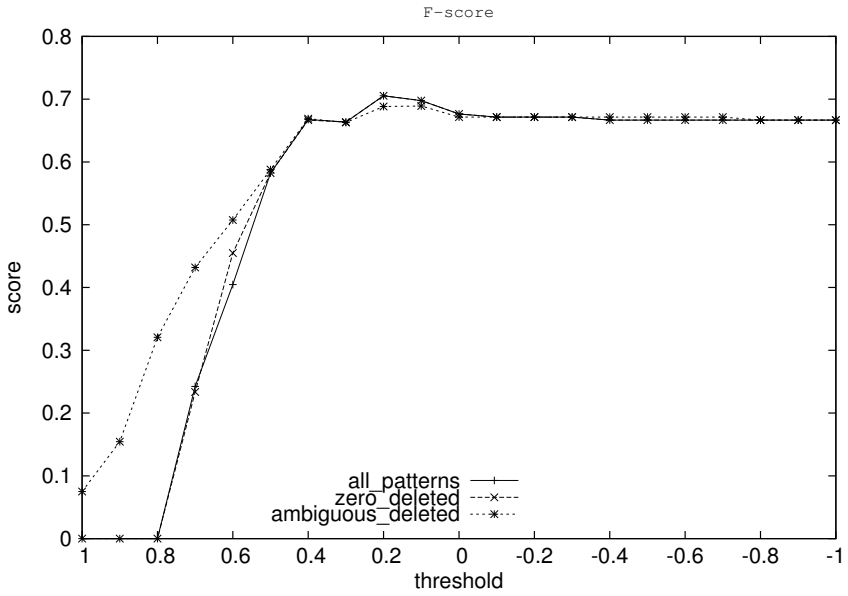


Figure2.10: F-scores for the classifier with three different versions of pattern list modification for set50.

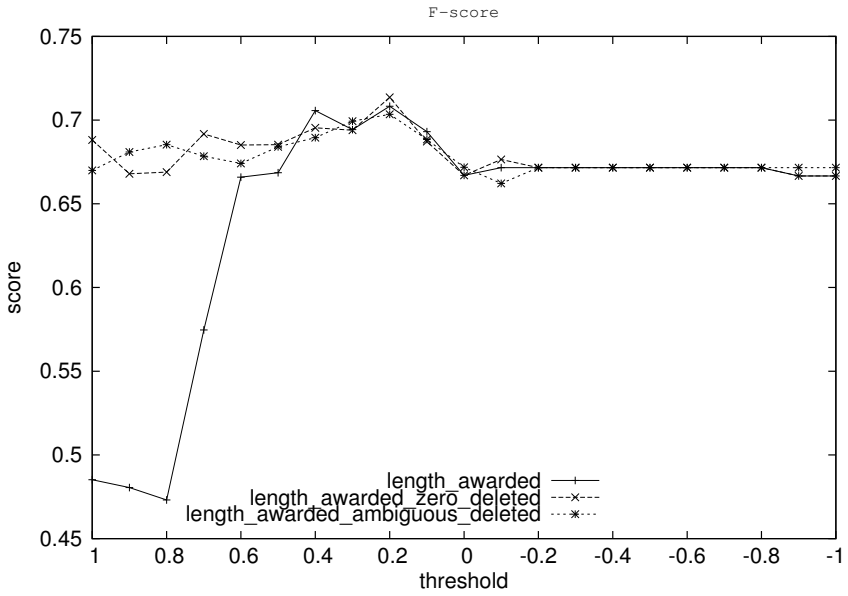


Figure2.11: F-scores for length based weight calculation for set50.

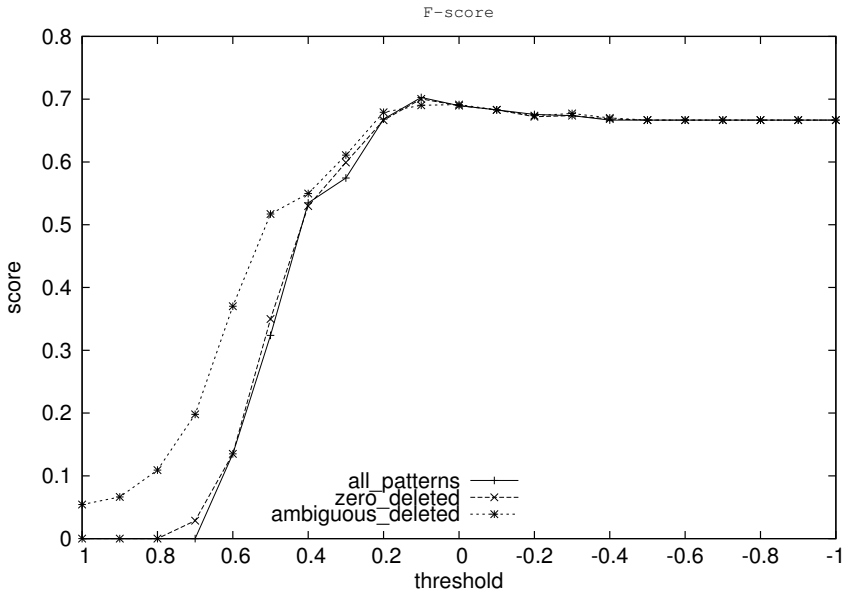


Figure2.12: F-scores for the classifier with three different versions of pattern list modification for set130.

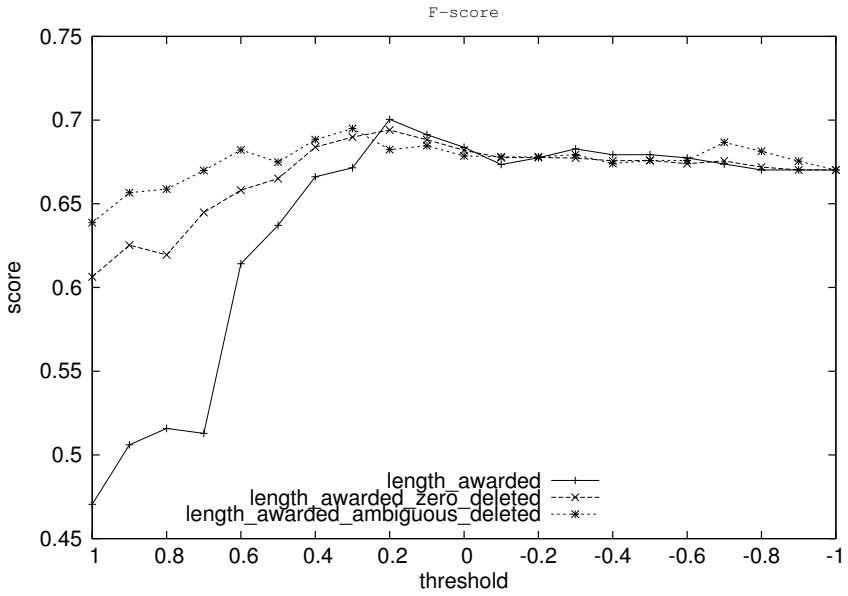


Figure2.13: F-scores for length based weight calculation for set130.

## 2.6 Method Validation

In this section, I present an additional experiment to validate the effectiveness of the proposed method in the extraction of future reference sentences.

### 2.6.1 Performance Change for Small Pattern Sets

At first I performed estimation of the effectiveness of morphosemantic patterns in future reference sentence classification. Firstly, I collected the following additional new validation set, unrelated to the initial datasets. From one year (1996) of *Mainichi Shinbun* newspaper I extracted 170 sentences from articles appearing on first three pages of each edition, and articles from the topics "economy", "international events" and "energy."

We manually annotated these sentences as either future or non-future related with five annotators: one expert annotator and four laypeople. Each sentence was annotated by one expert- and two layperson-annotators. I decided to leave the sentences for which there was an agreement between at least one layperson annotator and the expert. In result 59% (exactly 100 sentences) were left as the validation set.

Next, I classified these newly obtained sentences using the most frequent patterns (first 10 of them are represented in Table 2.6) generated in previous experiment. In particular, I performed pattern matching on the new sentences with the following sets:

A: first 10 patterns,
B: adding 5 patterns longer than three elements to set A,
C: subtracting 5 patterns from the tail of set A (to discard less frequent patterns shorter than three elements),
D: using only first 10 patterns containing more than three elements (differently to Set A which contains also frequent but shorter patterns).

Table2.6: Examples of extracted morphosemantic patterns.

| Occ. | Future Reference Patterns | Occ. | Non-future Reference Patterns |
|---|---|---|---|
| 26 | [Action]*[State change] | 5 | [Place]*[Agent] |
| 43 | [Action]*[Object] | 4 | [Numeric]*[Agent] |
| 42 | [Action]*[Action] | 4 | [Verb]*[Artifact] |
| 20 | [State change]*[Object] | 4 | [Person]*[Place] |
| 16 | [State change]*[State change] | 3 | [Numeric]*[Agent]*[Action] |
| 15 | [Action]*[Object]*[State change] | 3 | [Adjective]*[State change]*[State change] |
| 15 | [Action]*[State change]*[No state change (activity)] | 3 | [Place]*[Place]*[No state change (activity)] |
| 14 | [Object]*[Action]*[State change] | 3 | [Place]*[State change]*[Place] |
| 13 | [Object]*[Action]*[Object] | 3 | [Time]*[State change]*[Artifact] |
| 12 | [State change]*[Action]*[State change] | 2 | [Noun]*[Person]*[Noun]*[State change] |
| | ⋮ | | ⋮ |

Once performance reached plateau (F-score = 0.43), increasing the number of patterns made little difference. The performance of pattern set C was poor since only a few patterns are used. The Precision of pattern set D is slightly higher than that of the other sets. This indicates it could be more effective to use frequent morphosemantic patterns containing more than three elements, even when the number of applied patterns is small. From the above, I conclude that it would be more effective to use patterns consisting of a few (two or three) elements if the focus of the extraction was on Recall, whereas it would be more effective to use patterns consisting of three or more elements if the focus was on Precision.

The scores in this experiment were lower than in the evaluation experiment. However, I were able to extract future reference sentences with approximately 40% of Precision using only ten patterns, a score not far below the one achieved in the evaluation experiment (in which a total of 1102 patterns was used). This suggests that the performance could be also further improved when morphosemantic patterns are narrowed to those appearing in specific genre of events (only “economy”, or only “energy”).

Table2.7: Comparison of results for validation set between different pattern groups and the state-of-the-art.

| **Pattern set** | **Precision** | **Recall** | **F-score** |
|---|---|---|---|
| 10 patterns | 0.39 | **0.49** | **0.43** |
| 15 patterns | 0.38 | **0.49** | **0.43** |
| 5 patterns | 0.35 | 0.35 | 0.35 |
| 10 pattern with only over 3 elements | **0.42** | 0.37 | 0.40 |
| Optimized (see Figure 2.14) | **0.76** | **0.76** | **0.76** |
| Jatowt, 2011 [11] | **0.50** | 0.05 | 0.10 |

### 2.6.2 Compariosn with State-of-the-Art

We also compared my experimental results with those reported by [11]. In their experiment they extracted future reference sentences with 10 words and phrases unambiguously referring to the future, such as temporal expressions like "will," "may," "be likely to", etc. I translated those phrases into Japanese and applied to the new validation dataset of 170 sentences. The results of were low with P = 0.50, R = 0.05, F = 0.10. Although the Precision seems higher than the one described in section 2.6.1, my method extracted correctly much more future referring sentences with only 10 morphosemantic patterns. This indicates that the proposed method is valid. The reason for the low score obtained by the method of [11] on my validation dataset, despite its showing better performance previously could be explained by the differences in the approach. [11] used future-related patterns well known in linguistics, and searched for future sentences on the Internet which contains sufficient amount of data for extraction with even minimal number of seed words. I on the other hand trained my method automatically without providing any linguistic knowledge on a corpus from which I automatically extracted sophisticated morphosemantic patterns.

### 2.6.3 Performance of Fully Optimized Model

Finally, I verified the performance of the fully optimized model. The results of evaluation experiment (section 2.5) indicated that the model with the highest overall performance was the one using pattern list containing all patterns (including both ambiguous-, zero-patterns and n-grams) with weights modified by awarding pattern length. We re-trained the above model using all sentences from set130 and verified the performance by classifying the new validation set of 100 sentences.

As the evaluation metrics I used standard Precision, Recall and F-score. The scores of sentences oscillated from -0.01 to 2.27. The stronger was morphosemantic similarity to the training data the higher was the score. I also verified the performance for each threshold, beginning form 0.0 and checked every 0.2, up-till 2.2. The overall performance is represented in Figure 2.14. The highest reached Precision was 0.89, at Recall=0.13 with F-score=0.22. The highest reached F-score was 0.78 with Precision=0.65 and Recall=0.98 around the threshold of 0.4. Finally, break-even point (BEP) was at 0.76, which indicates that the proposed method trained on automatically extracted morphosemantic future reference patterns is sufficiently capable to classify future reference sentences.

Figure2.14: The results (F-score, Precision and Recall) for classification of future reference sentences in the test data.

Apart from the automatic classification results, I were also interested in the actual patterns that influenced the results. In Figure 2.15, Figure 2.16 I present detailed analysis of two sentences which obtained high scores in the experiment with first four patterns mapped on the sentences to facilitate better understanding of the future-referring morphosemantic patterns.

Examples(Example1 and Example2) of two sentences which obtained high scores in the experiment with their morphosemantic structure and extracted morphosemantic patterns. Each example contains in order: Score, Romanized Japanese [RJ], English Translation [E], Morphosemantic structure [MS], Morphosemantic future-reference patterns found in this sentence [MoPs]; for each example sentence, three examples of patterns from the list they contain (MoPs) are underlined, double underlined, overlined, or highlighted in gray.

1. Score=2.27

**RJ** *Dōsha wa kore made, Shigen Enerugī-Chōni taishi , dō hatsudensho no heisa , kaitai ni tsuite hōshin o setsumei shite kitaga, kaitai ni tsuite no hōteki kisei wanai tame, dōchō mo kaitai no kettei o shitatameru koto ni nari-sōda.*

**E** So far the company has been describing to the Agency for Natural Resources and Energy the policy for either closure or dismantling of the plant, and since there are no legal regulations found for dismantling, it is most likely that the agency will also lean to the decision of dismantling.

**MS** [Agent] [Other] [Organization] [Action] [State-change] [State-change] [Object] [Role] [State-change] [State-change] [Action] [Adjective] [Thing] [Agent] [State-change] [Other] [Verb]

**MoPs** [Agent]*[Verb],
[Agent]*[Organization]*[Verb],
[Agent]*[Action][State-change]*[Verb],
[Agent]*[Organization]*[State-change]*[Verb] .

Figure2.15: Example1: An example of sentences which obtained high scores in the experiment with their morphosemantic structure and extracted morphosemantic patterns.

1. Score=1.77

**RJ** *Ippō, senkyo kikan-chū ni '400 man-ri wokoeru Jimintō shiji no shomei o atsume, ōen shita'(tō kanbu) to iwa reru sekiyu, gasu nado enerugī kanren dantai ni taisuru hiaringu de wa, kūki ga ippen.*

**E** On the other hand, saying during the elections that they "collected the signatures of more than 4 million people supporting the Liberal-Democratic Party" (citation after the party's leader), during the hearing for the organizations related to the energy [sources] such as oil and gas, completely changed the atmosphere.

**MS** `[Action]` `[Action]` `[Numeric]` `[Verb]` `[Action]` `[Object]` `[State-change]` `[No-State-change-activity]` `[Citation]` `[Verb]` `[Thing]` `[Thing]` `[Action]` `[Action]` `[Object]` `[State-change]`

**MoPs** `[Action]*[State-change]`,
`[Action]*[State-change]*[Object]*[State-change]`,
`[Action]*[State-change]*[State-change]`,
`[Action]*[Action]*[State-change]*[State-change]`.

Figure2.16: Example2: An example of sentences which obtained high scores in the experiment with their morphosemantic structure and extracted morphosemantic patterns.

# Chapter 3
# The Effectiveness of Future Reference Sentences

## 3.1 Evaluation Experiment of Supporting Future Trend Prediction

### 3.1.1 FRS-based Future Trend Prediction

In the following section I conduct an experiment to confirm whether future reference sentences (for convenience abbreviated later to **FRS**) extracted with my proposed method are effective for future trend prediction support.

When predicting future trends, people synthesize multiple sources of information. This includes their own knowledge, experience, experts' claims regarding the future, past examples, and news from the web, radio, television, and newspapers. Such a large number of information sources potentially gives an access to an incalculably vast amount of knowledge. In practice it is difficult to follow all of them, however, even if knowledge and expertise are in short supply, it is possible to acquire a vast

amount of information through Internet search. Unfortunately, in a keyword search, millions of pages are retrieved and it can be extremely difficult to find the information one needs. Searching through related sites often brings up similar information. This is due to the fact that search engines function according to statistical data storing based on access frequencies.

Professionals such as data scientists are carrying out predictions of future trends according to statistical analysis and processing on numerical data. Applying data-mining techniques it is possible to blend the expert's experience and knowledge found on the Internet to predict future trends. The most important factor in prediction activities is to efficiently and effectively obtain the data actually useful in such trend prediction.

In the following section I conduct an experiment to assess the scale for contribution of the proposed FRS classification method in actual trend predictions. I conduct two kinds of experiments: one experiment for supporting future trend prediction by laypeople, and second for testing a prototype model for fully automatic trend prediction.

### 3.1.2 Experiment Setting

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 Chapter. 2. 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 was performed by a group of thirty laypeople (balanced gender distribution, age groups from university students 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.

The questions were taken from the Future Prediction Competence Test (*Senken-ryoku Kentei*), released by the Language Responsibility Assurance Association (*Gen-ron Sekinin Hoshō Kyōkai*)[1], a nonprofit organization focused on supporting people of increased public responsibility (company leaders, 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 (such as objective facts), while restraining form revealing others (one's fears towards the future or negative thoughts, disturbing public opinion). Thus the association helps preparing and managing one's public speeches and responsibility bound presentations.

The Future Prediction Competence Test 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., to describe economic situation of a country after next two years), and questions that must be answered using numbers (e.g., "What will

[1] http://homepage3.nifty.com/genseki/kentei.html

be the exchange rate of Japanese Yen to US Dollar after two years"), and they are scored after the results of the particular events in question 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 future prediction tests[2], 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 Future Prediction Competence Test 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 also 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.

In the evaluation the the forecasting support system we developed in this study is intended to apply future prediction support sentences (FPSS) related to a given question and provide assistance for humans on which answer to choose. Therefore for evaluation we limited the questions to multiple-choice questions. Questions with two or more (multiple) choices were selected from the 4th Future Prediction Competence Test and applied as questions for the experiment. (Fig. 3.1, Fig. 3.2 (in Japanese)).

[2]http://homepage3.nifty.com/genseki/senken/index4.html

評価実験にご協力お願いいたします。
この実験は，将来イベント予測支援のための評価実験です。
2009年の新聞記事を読み，2010年,2011年にそのイベントがどのようになっているかを予測する設問です。
どうぞよろしくお願いいたします.

<評価を行う前に読んでください。>
1．次の問題1〜6について"予測支援実験データ(1007).xlsx"の各シートの文を読んで、各設問に答えてください。
シート名は問題ごとに記載していますので、随時そのタブをクリックして参照してください。
参考にする文の数についてはシート名の横に記載しています。
2．選択肢を選ぶために参考になった文の番号を必ず答えてください。（複数可）
3．第１〜３候補を答える問題の場合、第１候補のみの回答でも良い。
4．回答は【】の中に答えてください。
5．毎日新聞2009(１年分) の記事を使用しています。
6．提出：10月9日（金）中 yoko@kushiro-ct.ac.jp まで.
予測問題.txt 予測支援実験用データ(1006).xlsx を添付して返信ください.

===ここから設問====

問題1: 以下の法案が、２０１０年６月末時点で可決成立しているか否かを予測せよ。
【シート：foreingers】17文
(1) 永住外国人への地方参政権付与：(a) 成立 (b) 未成立
回答:【 】
参考にした文の行番号：【 】

【シート：dual-surname】19文
(2) 夫婦別姓を認める民法改正：(a) 成立 (b) 未成立
回答:【 】
参考にした文の行番号：【 】

問題2: ２０１０年に行われるアメリカ中間選挙の結果を予測せよ。第三候補まで答えてください。
【シート：america-senkyo53】30文
(a) 上院・下院とも民主党が過半数をとる。
(b) 上院は民主党、下院は共和党が過半数をとる。
(c) 上院は共和党、下院は民主党が過半数をとる。
(d) 上院・下院とも共和党が過半数をとる。

回答:【 第１候補： ／第２候補： ／第３候補 】
参考にした文の行番号：
第１候補：【 】
第２候補：【 】
第３候補：【 】

問題3: ２０１１年６月末時点におけるアフガニスタンでの米軍の駐留状況を予測せよ。
【シート：afganistan】30文（ただし、必要ならば31以降の文も参照可）
(a) ２００９年１０月より増派して駐留。
(b) ２００９年１０月と同程度の規模で駐留。
(c) ２００９年１０月より規模を縮小して駐留。
(d) 米軍は完全撤退している。

回答:【 第１候補： ／第２候補： ／第３候補 】
参考にした文の行番号：
第１候補：【 】
第２候補：【 】
第３候補：【 】

問題4 ２０１１年７月２４日時点における地上デジタル・地上アナログ放送の状況を予測せよ。
【シート：ad-housou】30文（ただし、必要ならば31以降の文も参照可）
(a) 日本全国でデジタル放送の受信が可能となり、日本全国でアナログ放送が一斉に終了する。
(b) 日本全国でデジタル放送の受信が可能となっているが、一部の地域でアナログ放送は継続する。
(c) 日本全国でデジタル放送の受信が可能となっておらず、一部の地域でアナログ放送は継続する。

Figure3.1: Questions for experiment in supporting future trend prediction (page1)

(d) 一部の地域でアナログ放送は終了するが、大部分の地域で継続される。
(e) ２０１１年７月２４日時点では、アナログ放送が終了する地域はない。

回答:【 第１候補： ／第２候補： ／第３候補 】
参考にした文の行番号：
第１候補：【 】
第２候補：【 】
第３候補：【 】

問題5: ２０１１年度における子ども手当の支給状況を予測せよ。
【シート：kodomoteate】30文（ただし、必要ならば31以降の文も参照可）
(a) 公約通り、所得制限なしで全額(26,000円／人）支給される。
(b) 所得制限なしで減額して支給される。
(c) 所得制限ありで全額支給される。
(d) 所得制限ありで減額して支給される。
(e) 子ども手当自体実施されない。

回答:【 第１候補： ／第２候補： ／第３候補 】
参考にした文の行番号：
第１候補：【 】
第２候補：【 】
第３候補：【 】

問題６：２０１０年に行われる参議院選挙の結果を予測せよ。
【シート：sangiin-senkyo】30文（ただし、必要ならば31以降の文も参照可）
(a) 民主党が単独過半数をとる。
(b) 民主党だけでは単独過半数をとれないが、民主党を含む与党で過半数をとる。
(c) 民主党を含む与党で過半数を確保できない。

回答:【 第１候補： 】
参考にした文の行番号：
第１候補：【 】

ご協力ありがとうございました。
最後にアンケートにお答えください。

質問：３０文を読むことは(a)ちょうど良い (b)負担に感じた。(c)足りない
回答【 】

→(b)と答えた方に質問です。
何文程度が適当に感じますか？ 回答【 文程度】

→(c)と答えた方に質問です。
何文程度が適当に感じますか？ 回答【 文程度】

Figure3.2: Questions for experiment in supporting future trend prediction (page2)

### 3.1.3 Data Preparation

Here, I describe the experiment setup. A total of 7 multiple-choice questions were selected from the 4th future prediction test as benchmarks (Figure 3.1, Figure 3.2). Laypeople read the FPSSs presented and responded immediately. The FPSSs for each question presented to the laypeople participants were gathered by keywords from the Mainichi Newspaper's entire 2009 year; MoPs were generated using the fully optimized model (calculated with length_awarded on set130) according to 1–cross–validation, and after the text of each and every question was classified, a score was assigned.

Classified texts with scores over 0.0 were determined to be FRS, and up to 30 of them were presented to subjects as reference FPSSs. FPSSs were arranged and presented in chronological order, from the oldest to the newest. In consideration of the possibility that some respondents would request more FPSSs for prediction, I also included FPSSs of over thirty sentences. In addition, for questions for which less than 30 FPSSs were extracted in general, I presented all of the sentences that were classified into FRS. As an example, FPSSs for Question 3 are presented at the end of this section. The questions were answered directly after reading only the FPSSs. Additionally, the respondents were asked to report the ID number of the FPSSs they referred to in their answer.

Each of the questions 1, 2, 7, were allocated 3 points. Moreover, in questions 2–5 the participants were allowed to make up to three candidate choice answers: primary candidate, secondary candidate and third candidate, allocated 3 points, 2 point and 1 point, respectively. Additionally, for comparison, I made a different point allocation, allowing strictly only one point per question.

The questions were collected with the following keywords. We tried several kinds of keywords on each question, and after careful examination of the sentences, decided on them by hand.

**Q1-1:** (Participation of foreigners with permanent resident status in regional government) — (Participation in regional government permanent resident alien)

**Q1-2:** Husband and wife retaining separate family names

**Q2:** midterm elections — (Republican — Democrat) & (United States — America)

**Q3:** Afghanistan

**Q4:** Analog broadcasting — Digital broadcasting

**Q5:** Child allowance

**Q6:** (Democratic Party — Ruling Party — Liberal Democratic Party) & elections

Examples of future prediction supporting FRS for Question 3. (ordered according to time line [**MonthDay**] in 2009) are represented below.

1. [**0118**] 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.

2. [**0121**] 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.

3. [**0121**] Additionally, it reveals 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.

4. [**0122**] Substantial negotiations toward realizing the campaign pledge to reduce the number of stationed US forces "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.

5. [**0122**] Ahmad Saif (29), an engineer in Baghdad, 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.

6. [**0207**] 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.

7. [**0226**] At the conference, a US-Japan joint investigation into strategies regarding Afghanistan was agreed upon, and a special envoy will be dispatched to the US to settle the details.

8. [**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.

9. [**0402**] Also, while a joint statement from the two countries confirmed their cooperation on the problems surrounding Afghanistan, North Korea, and Iran, they also specified differences on their stance regarding the causes of the conflict in Georgia and the missile defense deployment plan in Eastern Europe.

10. [**0408**]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.

11. [**0409**] The Obama administration is advancing their aim of pivoting the terrorism countermeasures to Afghanistan, but the stabilization of Iraq is essential to the realization of their plan.

12. [**0430**] Saying "We're celebrating the progress, but are still not fully satisfied," they showed an intention to work hard to find solutions to important issues such as bringing the economic crisis under control, stabilizing Iraq and Afghanistan, health-care insurance reform, and so on.

13. [**0528**] The technique of armed assault followed by suicide-bombing resembles the attacks that killed over 60 people in the Marriott Hotel of Islamabad last year in September, and it appears that Baitullah Mehsud, commander of militants in the Afghan border region, who had claimed responsibility for the Islamabad attack, was involved this time as well.

14. [**0602**] As part of their comprehensive strategy in the war on terror in Afghanistan, the Obama administration has taken notice of the dialogue, saying that improving ties between the two countries is a required condition for the Pakistani forces to concentrate on domestic counter-insurgency operations.

15. [**0605**] It is undeniable that the President's careless statements have given the strikes against Afghanistan and the war in Iraq an "anti-Islam" impression.

16. [**0816**] 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, saying that "[civilian assistance is] already in place."

17. [**0917**] Furthermore, Okada brought up North Korea, global warming, US military bases in Okinawa, and support to Afghanistan as problems to be solved in 100 days, suggesting these are the most important issues for the US-Japan relations.

18. [**0917**] Regarding alternative policies he said, "This isn't something that is to be implemented by the [Ministry of Defense] alone; I want the administration to discuss it thoroughly," thereby disclosing the government's plan to examine new Afghanistan assistance measures.

19. [**0917**] Officials at the United Nations Assistance Mission in Afghanistan (UNAMA) stated that, "Even if there is a run-off vote, it will turn into a battle of accusations, which will further deepen the chaos," in a sigh of desperation at the deadlocked situation.

20. [**0923**] On the Indian Ocean refueling mission slated to end in January next year, Foreign Minister Katsuya Okada said, that "We're not doing a simple extension," while at the same time strongly emphasizing the importance of continuing support to Afghanistan.

21. [**1010**] When facing down the two major issues, namely, health insurance reform and the problem of a reinforcement surge for the Afghan War, the administration is promoting their "internationally respected president" and making an aggressive turn-around.

22. [**1017**] Furthermore, mentioning that "the United States is emphasizing cooperation in Afghanistan as well," they indicated their plan to prioritize the establishment of the new Afghan assistance measures.

23. [**1019**] It was said that in addition to those, "there is a number of useful policies contributing to the development and cultivation of human resources

and so on," disclosing expectations that Japan will consider their Afghani aid policy anew, and the plans will be implemented.

24. [**1027**] Above all, Afghanistan, while coping with the threat of terrorism, is rebuilding the country and aiming to establish peace and stability.

25. [**1107**] Opposition against sending reinforcements to Afghanistan is increasing in the United States, and a debate is growing over whether to strengthen the national forces to bear the responsibility of maintaining public order, currently held by US troops, in order to facilitate future troop withdrawals even in the event of reinforcements.

26. [**1114**] According to the AP and others, the defendant Hasan planned to be deployed to Afghanistan this month, however he did not want to participate in the mission.

27. [**1119**] It appears that Pakistani men (39) were arrested by the Korean National Police Agency after repeatedly smuggling acetic anhydride into Korea from Japan and elsewhere, and smuggling it out via Iran to Kandahar, Afghanistan.

28. [**1211**] Moreover, as Commander in Chief of the US military through two wars in Iraq and Afghanistan, he spoke frankly about the problem of "war and peace," asking that we understand this as a "necessary war."

29. [**1218**] The Obama administration promoted "assistance for the Pakistani civil government" as a part of their new strategy in Afghanistan, in addition to expanding missile strikes with drones.

30. [**1231**] It is hard to say whether the economy is on the road to recovery, while national crises like the troop surge in Afghanistan and foiled terror attempts continue.

### 3.1.4 Experiment Results

Table 3.1 shows the results of 30 people's responses. The scoring was performed in accordance with future prediction test scoring procedure, wherein each question is worth 3 points with a total of 21 possible points. For questions with up to 3 choices possible, I awarded 3 points when the first choice was correct, 2 points when the second was correct, and 1 point when the third choice was correct. Apart from the experiment with 30 respondents, I analyzed the responses of the actual participants in the 4th Future Prediction Competence Test. The total possible score of was equal to 120 points. The test was taken by 11 people. Of those points, prediction questions account for 90 points, while essay questions account for 30 points. The comparison was based on prediction questions with a maximum score of 90 points.

In the performed experiment, the average score of our participant was 35.71%. In comparison, the average score of the test participants was 33.4%. These results were similar, which shows that even though the events for prediciton in our experiment were in fact from the past, the experiment participants performed similarly to original test participants. Therefore it can be said that the participants did not use (or did not have) the knowledge about the predicted events and that they based their judgments on the provided FPSS. Furthermore, in comparison with original test results, an improvement of approximately 2% was noticed. This can be considered as the contribution of our system.

Additionally, the highest score in our experiment was 61.9%, while the lowest was 14.29%. In comparison with the 4th Future Prediction Competence Test these results indicate an improvement of 0.8 percentage points for the highest score range and 7.62 percentage points for the lowest range. The accuracy of the results (when any of the options 1–3 were correct) is shown in Table 3.7. The accuracy of the method described in this study was approximately 43%.

Moreover, the Future Prediction Competence Test has an established ranking system based on the number of points received. On the 4th future prediction test,

a score over 60 (out of 120) earns a 1st class ranking; over 50, 2nd class; and over 40, 3rd class. This refers to the level of competence of a participant in future event prediction. Table 3.3 shows the results calculated for each grade as applied to the experimental results of this study. On the 4th Future Prediction Competence Test, 2 people earned 1st class, none earned 2nd, and 2 people earned 3rd class. In comparison, experiment participants performing predictions with the use of FPSS produced significantly more accurate results, if their scores were calculated at the time of test submission: 6 people earned 1st class, 6 earned 2nd, and 4 earned 3rd. Also, the aggregate of both the statements referred to by accurate respondents as well as those referred to by incorrect respondents are shown from Figure 3.3 to Figure 3.9.

Here, the horizontal axis is the FPSS statement number, and the vertical axis is the number of times the statement was considered a useful reference for prediction.

Table3.1: Result of correct accuracy with weighted scoring (including original results of 4th Future Prediction Competence Test for comparison).

| | Average | Highset | Lowest |
|---|---|---|---|
| experiment results | 0.3571 | 0.6190 | 0.1428 |
| future prediction test results | 0.3344 | 0.6111 | 0.0666 |

Table3.2: Result of correct accuracy rate with no weighted scoring (added original results of 4th future prediction test for comparison).

| | Average | Highset | Lowest |
|---|---|---|---|
| experiment results | 0.4285 | 0.8571 | 0.1428 |
| future prediction test results | 0.3344 | 0.6111 | 0.0666 |

Table3.3: Comparison of number of participants who passed (or would have passed) the 1-3 grades of the future prediction test between original examinees and participants in the future prediction support experiment.

| | 1st grade | 2nd grade | 3rd grade | number of examinees |
|---|---|---|---|---|
| Original examinees | 2 | 0 | 2 | 11 |
| Experiment participants | 6 | 6 | 4 | 30 |

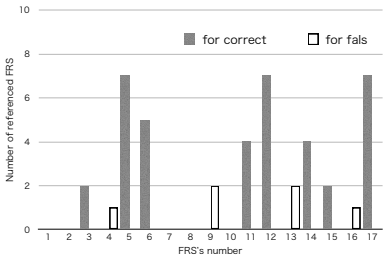


Figure3.3: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question1-1

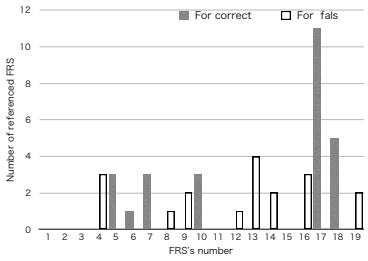


Figure3.4: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question1-2

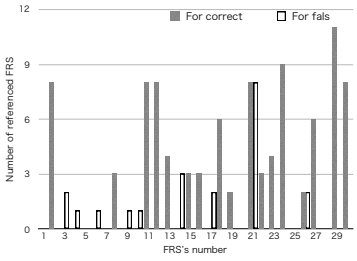


Figure3.5: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question2

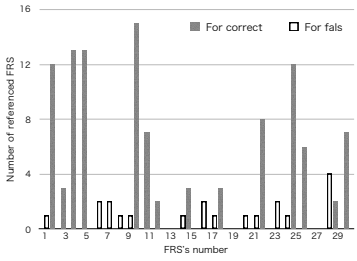


Figure3.6: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question3

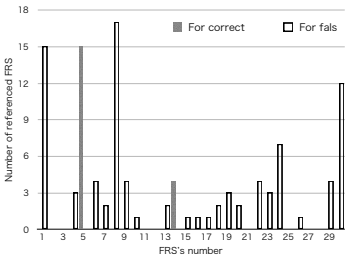


Figure3.7: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question4

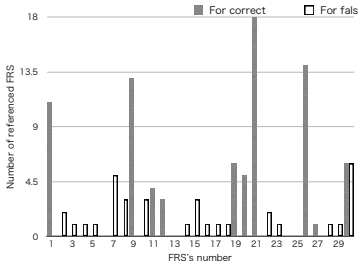


Figure3.8: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question5

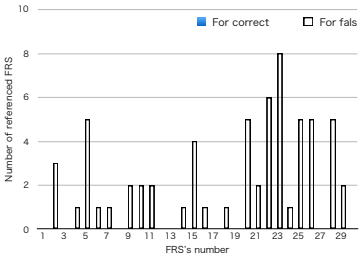


Figure3.9: Number of times FPSS statements were used as references by participants who got correct and incorrect answers for Question6

### 3.1.5 Discussion

In this section, I discuss the effectiveness of FRS for future trend prediction while comparing the results of the experiment with those of the future prediction test.

A comparison of the average scores of the future prediction test (33.44%) and the results of the experiment (35.71%) revealed a small difference of 2 percentage points in favor of the proposed method. Although the 2 percentage-point difference is not large, it can be considered meaningful. This is because the results of experiment participants were based only on 30 automatically extracted FPSS statements and the questions were answered immediately after reading through the FPSS, whereas in the actual future prediction test, the respondents had an entire year to answer the questions, and could consult any document they like, and were free to seek the opinions of others, including experts.

As shown in Table 3.4, if we look at the accuracy of the 4th Future Prediction Competence Test, the average was 31%, demonstrating that when people have every means at their disposal, they still only accurately predict the future on around one third of the time. Kakeya et al. [34] analyzed currents of the answer results on the 1st Future Prediction Competence Test and verified whether the idea of collective intelligence is true or not. The accuracy rate at that time was 33.17 %. Moreover, Kakeya et al. concluded that the collective intelligence is not possible when it comes to future prediction. It means it is not easy for people to predict future trends.

The accuracy of my proposed method was 43%, an improvement of 13 percentage points over that average. Furthermore, a consideration of the certification breakdown from 1st class to 3rd shows the Future Prediction Competence Test participants at 36.36% and FPSS at 53.33%; an improvement of roughly 17 percentage points. Thus, it is evident that when predicting future trends, FRS can dramatically reduce time and effort spent gathering information and achieve above-average predictive accuracy. Therefore, using FRS to support future trends forecasting is effective.

Next, I analyze the FPSS referred to by experiment participants as most useful. Figure 3.3, Figure 3.4, Figure 3.5, Figure 3.6, and Figure 3.8 show for each question graphs with highly accurate responses, while Figure 3.7, and Figure 3.9 show questions with low accuracy responses. Gray bars indicate the number of statements referred to by successful respondents, while white bars indicate the number of statements referred to by respondents who failed the task of prediction.

The contribution of these statements 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 statements were referred to. Taking Question 3 as an example, I analyze both the content of statements that were only referred to by incorrect answers as well as those that contributed to correct responses.

In the experiment, 83.33% of responses to Question 3 were accurate. Examples (a)–(c) show FPSSs contributing to correct answers, whereas examples (d)–(f) show FPSSs that contributed to inaccurate responses. The values on the horizontal axis of Figure 3.6 correspond to FRS numbers. FPSS (a)–(f) were all selected using the keyword "Afghanistan". However, (a)–(c) contain references to troops in Afghanistan, whereas (d)–(f) contain the word "Afghanistan" but do not refer to troops. Therefore, in order to improve the prediction accuracy, it is necessary to devise a better keyword setting for selecting FPSS from newspaper corpora.

**[Question3]**

**Predict whether there will be more or less US troops stationed in Afghanistan at the end of June, 2011.**

**(a)**

**ID number of FRS:** 10 (10th sentence)

**number of references:** 15 (sentence was marked as useful in prediction 15 times)

**FRS score=** 1.8329

**Japanese sentence:** その一方でアフガニスタンへの増派計画を発表しており、イラクからの撤退によって対テロ戦争の軸足をアフガンへ移す姿勢を示している。

**English:** On the other hand, there has been announced a reinforcement plan to Afghanistan, and it shows the attitude to move the pivoting foot of the war on terror more to Afghanistan by withdrawing troops from Iraq.

**MS:** [Adnominal] [Action] [Place] [Object] [Role] [State change] [Time-Point] [State change] [Action] [Object] [Noun][State change] [Object] [State change]

**(b)**

**ID number of FRS:** 4

**number of references:** 13

**score=** 1.969

**Japanese sentence:** 公約である「就任後１６カ月以内」の駐留米軍撤退に向けた実質的な協議に着手し、対テロ戦争の主戦場と位置付けるアフガニスタンへの米軍増派を含む総合計画の早期策定を目指す。

**English:** [They] embarked on substantial consultations regarding the pledge of withdrawal of US troops “within 16 months after taking office”, aimed at the early development of a comprehensive plan, including the US military surge to locate the main battlefield of the war on terror in Afghanistan.

**MS:** [Things] [Action] [Time] [Action] [State change] [Auxiliary verb] [Object] [Action] [State change] [Action] [Citation] [Verb] [Place] [Object] [State change] [Action] [Object] [No-state change(Action)]

**(c)**

**ID number of FRS:** 5

**number of references:** 13

**score=** 1.9039

**Japanese sentence:** バグダッドのエンジニア、アフマド・サイフさん（２９）は、オバマ大統領が対テロ戦争の軸足をアフガニスタンに移す考えを改めて強調したことで、米軍の早期イラク撤退が現実味を増したと喜んだ。

**English:** An engineer from Baghdad, Ahmad Saif (29) expressed his joy regarding the fact that early withdrawal of US troops from Iraq have become more realistic after hearing that President Obama have re-emphasized the idea to move the pivot foot of the war on terrorism to Afghanistan.

**MS:** [Place] [Person] [Number] [Agent] [Action] [Object] [Goal] [State change] [Object(Role)]] [Adverb][State change] [Other] [organization] [Agent] [Object] [State change] [State change]

**(d)**

**ID number of FRS:** 7

**number of references:** 2

**score=** 2.2268

**Japanese sentence:** 会談では、アフガニスタン戦略に関し日米共同で検討することが決まり、今後、特使を米国に派遣して内容を深めていくことになった。

**English:** At the meeting, it was decided that there will be a joint US-Japan discussion regarding dispatching a special envoy [of Japanese Self Defense Forces] to the US, and both sides decided to further work on the details.

**MS:** [No-state change(Action)] [Place] [Action] [No-state change(Action)] [Object] [State change] [Time] [Object] [Goal] [State change] [Object] [State change] [Other] [Verb]

**(e)**

**ID number of FRS:** 8

**number of references:** 1

**score=** 1.7977

**Japanese sentence:** ロシア外務省は6日に声明を発表し、両国にはアフガニスタン情勢やテロとの戦いなど立場が近接または一致するテーマがあり、外相会談の結果を「控えめに楽観している」と述べた。

**English:** Russian Ministry of Foreign Affairs announced the statement on 6th, and there was proximity or a theme to accord, and a situation including the situation and war on terrorism in Afghanistan spoke the result of the Minister of Foreign Affairs talk in the two countries, "I am optimistic modestly".

**MS(Japanese):** 動作主 時間 対象 状態変化あり モノ 場所 その他 動作 対象 生成物 動作 状態変化あり モノ 動詞 動作 対象 動作 状態変化あり 状態変化あり

**MS:** [Agent] [Time [Object] [State change] [Things] [Place] [Other] [Action] [Object(product) [Action] [State change] [Things] [Verb][Action] [Object] [Action] [State change] [State change]

**(f)**

**number of FRS:** 9

**number of refernece:** 1

**score=** 1.7549

**Japanese sentence:** また両国は共同声明で、アフガニスタン、北朝鮮、イラン問題での協調を確認する一方、ＭＤの東欧配備計画やグルジア紛争の原因と評価については立場の相違があることを明記。

**English:** In addition, the two countries continue confirming Afghanistan, North Korea, the cooperation by the problem in Iran by joint communique and specify that there is the difference in situation about Eastern Europe deployment plan of MD and a cause and the evaluation of the Georgian dispute.

**MS(Japanese):** 身体部分 動作主 動作 場所 場所 場所 対象 状態変化あり 動作 モノ 動作 動作 その他 状態変化あり その他 その他 動詞 対象 対象

**MS:** [Body part] [Agent] [Action] [Place] [Place] [Place] [Object] [State change] [Action] [Things][Action] [Action] [Other] [State change] [Other] [Other] [Verb] [Object] [Object]

We propose a method to improve prediction accuracy that references more detailed information when employing FPSS. This takes FPSS as a key statement, and references all articles containing it. Consider the example articles below, which contain both the key statement and FRS.

Example 1

As you can see from the topic, this is not an article about stationing troops in Afghanistan, but does contain content referring to Afghanistan. As in this case, there are times when important information can be gleaned even from articles on a different topic.

Example 2

This article contains direct information on troop increases in Afghanistan. Referring to an article containing an FRS that contains the keyword “Afghanistan,”

the sentence prior - "In February the President announced his plan to reduce the number of US troops stationed in Iraq from 140,000 to 40,000 by August of 2010, and stage a complete withdrawal by 2011." - is an even stronger expression of the plan to increase the number of staff in Afghanistan. In this way, acquiring articles that contain the key statements allows the acquisition of clearer information and makes decision-making easier.

Example 3

This article contained statements with a score = 1.4311 and rank = 166. The score's BEP was over 0.98; however at around 1.7077, it was ranked 30th and thus the recall was insufficient, making this an example of a necessary statement that was overlooked. A survey of the number of statements necessary for respondents to answer one question (i.e. without getting tired) is 22.8. This time I used the 30 statements with the highest scores, but I think using statements with a more balanced score around BEP = 0.98 could be one way to reform the tendency to overlook important information.

Example 4

This statement was selected with the keyword "Afghanistan," and was extracted as an FRS. While it does contain content referring to the future, it does not contain information related to stationing of troops in Afghanistan. If I read the article that contains this statement, it references the contents of the conversation at the Japan-Russia Foreign Ministers' meeting, but it does not constitute a reference for the question. In a case like this, when there is an FRS but the contents refer to something unrelated, predictive results depend on the user.

In summary, I have seen that more detailed information can be acquired by refer-

ring to the content of articles found through searches using FRS as key statements. In addition, necessary information can be found in articles that could not be collected on the basis of their headline alone. Thus, in this way it is effective to use FRS as key phrases for browsing article contents.

In today's society, there is simply too much information, and it is extremely difficult to acquire the information you want. Furthermore, the functionality of Web searches is such that frequently searched pages come to the top of the results, but it is certainly the case that there is important information that is not often searched for. Using FRS as an indicator and the information contained therein as a reference eliminates the need to search randomly amongst the enormous mass of available information, making for more efficient information retrieval. We believe that this method is useful for efficiently gathering information because it saves time and effort to read all articles from beginning to end.

Table3.4: An accuracy rate for each future prediction test.

| | accuracy rate | highest point | lowest point |
|---|---|---|---|
| the 3rd | 30.5 | 57 | 0 |
| the 4th | 33.4 | 55 | 6 |
| the 5th | 30.1.5 | 39 | 11 |
| the 6th | 30.1 | 54 | 15 |

Table3.5: An accuracy rate of each questions.

| Question No. | accuracy rate | number of FPSS |
|---|---|---|
| Q1-1 | 63.33 | 16 |
| Q1-2 | 20.69 | 19 |
| Q2 | 63.33 | 29 |
| Q3 | 83.33 | 30–172 |
| Q4 | 13.33 | 30–85 |
| Q5 | 43.33 | 30–267 |
| Q6 | 3.45 | 30–430 |

**Example1:**

**No.:** 29

**ID:** 91218062

**Topic:** Pakistan: An opposition party, the President impeachment are great blows to "corruption innocence" destruction, rice of the support, too

**Sentence:** The Obama U.S. government advocates "civil administration support of Pakistan" in an Afghanistani new strategy, too and extends the missile strike using the unmanned aircraft.

**Article:** The Pakistani Supreme Court canceled "a nation conciliation agreement" to perform an acquittal of corruption punishments of the politician on 16th, and the trial resumption of seven corruption punishment of a total of 1,500 million dollars said to that colander President Dali participated was enabled. I received this, and the opposition force executive including largest opposition party "federation of Muslim Nawaz Sharif group" suggested the start of the impeachment procedure for 17 days saying "the President should resign". The resignation demand begins to smolder in the governing coalition, and the United States which strengthened an anti-terrorism war in Pakistan by supporting colander Dali seems to be pressed for a strategic large review.

About Supreme Court order, the Executive Office of the President denies resignation for 17 days saying "there is no problem". But the prime minister prefecture showed a posture for the trial resumption saying "I respect an order" and highlighted the difference in viewpoint. The Secretary of Thiaw Dally Supreme Court which acted as Judge chief of the order of 16th is opposition intensely in the Musharraf ex-government and the past when I threw it on cooperation to U.S. I was engaged in the

location elucidation of several hundred people who became missing in terrorism mop-up operation. I collected the big support of the strong nation of the anti-American sentiment, and such a posture led to a political crisis of Musharraf.

This was because it enfeoffed his force that it gave an order for the effect stop of the national reconciliation agreement Musharraf declared a state of emergency in November, 07, and to have dismissed tea Uda Lee by "power". But the political reflux of Musharraf does not stop, and the government ruling party suffers a crushing defeat in a general election of February, 08. "The Populist party" where assassinated Bhutto former First Husband, colander Dali took office as the top recaptured the government.

It was colander Dali who restored civil rights by a conciliation agreement, and took office as the President, but I continued refusing reinstatement of tea Uda Lee who repelled the agreement, and the repulsion increased from some opposition force and ruling parties.

Colander Dali maintains the President authority that Musharraf extended by the reason of anti-terrorism war accomplishment to make up for the weakness of the domestic administration base and strengthens a cooperation route to U.S. still more. The Obama U.S. government advocates "civil administration support of Pakistan" in an Afghanistani new strategy, too and extends the missile strike using the unmanned aircraft. The Kiyani Chiefs of Staff Pakistan military authorities resist this with "sovereignty disregard".

Blast terrorism occurs successively in the country as a result that the Obama U.S. government demanded military operation reinforcement from Pakistan. The voice for the switch of the cooperation to U.S. is strengthened with criticism to colander Dali who is cooperative with the United

States among the nations.

If trial is reopened, it is certain that colander Dali loses political centripetal force, and the new strategy of the Obama Administration considering, "cooperation of Pakistan is indispensable for stability of Afghan" will be greatly out of order.

**Example2:**

**No.:** 10

**ID:** 90408013

**Sentence:** On the other hand, I announce the reinforcement plan to Afghanistan and show a posture to move a pivot leg of the war on terrorism to Afghan by withdrawal from Iraq.

**Topic:** U.S. President Obama: To the visit prime ministers and a talk lightning in Iraq

**Article:** The U.S. President Obama who visited Europe, Turkey visited (the Japan time night of the same day), capital Baghdad, Iraq as a surprise on the afternoon of 7th. After the assumption of office, this is the first visit of the President to Iraq. (to six a related story)

The President talked together with the Odieruno commander of the stationing United States Armed Forces in US base camping fish basket Tolly of Baghdad on the same day. I am going to talk together with the Iraqi Mariki prime ministers afterwards. According to Associated Press, the President stated, "time when the Iraqi nation achieved the responsibility to one's country came". It is said that I decided a visit to Iraq to show

the will of the thanks to the stationing corps on the way to the United States from Istanbul.

The President withdraws about 100,000 among a little over 140,000 Iraqi stationing United States Armed Forces by August, 2010 in February and expresses a policy to withdraw entirely by 11 years. On the other hand, I announce the reinforcement plan to Afghanistan and show a posture to move a pivot leg of the war on terrorism to Afghan by withdrawal from Iraq.

**Example3:**

**No.:** 166

**ID:** 91210225

**Sentence:** On the other hand, criticism is given for the awarding that announced the United States Armed Forces reinforcement of 30,000 people to Afghanistan just after that at home and abroad.

**Topic:** The Nobel Prize: U.S. President starts on an award ceremony

**Article:** The U.S. President Obama left the air force base of the suburbs of Washington to attend at the award ceremony of Nobel Peace prize practiced in Oslo on 10th on 9th.

President Obama expresses the leadership of America about the action to "the nuclear-free world" which became one of the awarding reasons by the speech of the award ceremony some other time.

On the other hand, criticism is given for the awarding that announced the United States Armed Forces reinforcement of 30,000 people to Afghanistan just after that at home and abroad. Therefore, the President is going to

appeal for being the part that reinforcement desires Afghan and world peace.

**Example4:**

**No.:** 8

**ID:** 090307040

**Sentence:** Russian Ministry of Foreign Affairs announced the statement on 6th, and there was proximity or a theme to accord, and a situation including the situation and war on terrorism in Afghanistan spoke the result of the Minister of Foreign Affairs talk in the two countries, "I am optimistic modestly".

**Topic:** U.S.-Russia Minister of Foreign Affairs talk: Axis including milestone START1 and MD of the U.S.-Russia improvement

**Article:** Foreign Minister of Rablove Russia talks together with U.S. Secretary of the State Clinton in (the Japan time early morning of 7th), Geneva on the afternoon of 6th. After Obama U.S. government start, the both sides want to do it at a Foreign Minister of first United States and Russia talk with a milestone for the improvement of the relations that they made cooling under ex-government of Bush.

Prior to a talk, Secretary Clinton points it out based on North Atlantic Treaty Organization (NATO) having selected the reopening of cabinet minister level talks with Russia as 6th, "many in terrorism, the field that can cooperate including the nonproliferation of weapons of mass destruction". I stated, "the United States and Russia cooperated and researched and developed it, and there will be an opportunity to carry

out even joint deployment in the future" about an MD deployment plan and promoted Russian participation.

Russian Ministry of Foreign Affairs announced the statement on 6th, and there was proximity or a theme to accord, and a situation including the situation and war on terrorism in Afghanistan spoke the result of the Minister of Foreign Affairs talk in the two countries, "I am optimistic modestly". For the first talk of both Obama rice that this Minister of Foreign Affairs talk is carried out again in London on April 2, Medvedev dew Presidents when was stepped, placed it.

Russia repelled that ex-government of Bush pushed forward the MD Eastern Europe deployment that could spoil power of the strategic nuclear missile of the country until now. But "MD is not necessary without a menace of Iran", and the Obama Administration is more likely to use cooperation and the MD deployment by the nuclear issue in Iran for the business materials by negotiations to Russia for a posture with (Defense Secretary Gates). It is considered that the progress of the issue of MD gives positive effect for the nuclear disarmament negotiation over the START succession treaty at the same time.

But Russia is consistency in a dispute in Georgia in a situation "not to recognize independence" by the problem that approved South Ossetia and independence of Abkhazia in the United States in the Georgia territory. There is no sign that U.S.-Russia both sides compromise, and it seems to be difficult to expect progress.

## 3.2 Evaluation Experiment of Prototype Method

### 3.2.1 Automatic Prediction Based on FRS

In this section, I conduct an experiment to evaluate the automatic prediction performance of the prototype method on answer choices using MoPs as FPSS.

Just as when humans read the prediction support statements and selected answers, this implementation answered the same questions used in the experiment for supporting future trend prediction described in the previous section. For each question, I conducted training on SPEC using the semantic role label (SRL) assigned to the FPSS. Just as in the experiment in Section 3.1.2 where I categorized FRS, this time I used a calculation method that considers length_awarded, and in order to use all the patterns, weighted MoPs were generated using 1-cross-validation. In other words, learning was conducted with the same experimental setup as the experiment for supporting future trend prediction.

The following two sentence patterns were used for FPSS learning. The number of statements differed for each question (Table 3.5).

1. FRS 17–30, which were read by subjects in the prediction support system
2. As indicated in the previous section, FRS above BEP = 0.98.

Next, the text of the choices for each question was taken as test data, and classified using the generated MoPs. Among those results, the choices with the highest score were used as the prototype method's answers. Both weighted scoring and non-weighted scoring methods were employed to score the results. The prototype method experiment results are shown in Table 3.8, and the scoring results are in Table 3.6.

Table3.6: Results with weighted scoring (comparison prototype method for future prediction and experiment in supporting future trend prediction.

| number of sentences for learning | Q1-1 | Q1-2 | Q-2 | Q-3 | Q-4 | Q-5 | Q-6 | score(/21.00) | percentage of full score |
|---|---|---|---|---|---|---|---|---|---|
| score of system prediction | | | | | | | | | |
| 1. up to 30 sentences | 3.00 | 3.00 | 3.00 | 1.00 | 0.00 | 2.00 | 0.00 | 12.00 | 57.14% |
| 2. more than BEP=0.98 | 3.00 | 3.00 | 3.00 | 1.00 | 0.00 | 2.00 | 0.00 | 12.00 | 57.14% |
| average score of laypeople | | | | | | | | | |
| | 1.90 | 0.60 | 1.40 | 2.27 | 0.17 | 1.067 | 0.10 | 7.50 | 35.70% |

Table3.7: Result of correct accuracy rate (experiment results compared with future prediction test results).

| | Average | Highest | Lowest |
|---|---|---|---|
| experiment results | 0.4285 | 0.8571 | 0.1428 |
| future prediction test results | 0.3344 | 0.6111 | 0.0666 |

When scoring the choices, in cases where there were under 30 FPSS statements and when all FRS with scores over BEP = 0.98 were used, the scoring rate was 57.14%, an improvement of 21.19% over results obtained by human predictions. In comparison with the scoring rate on the 4th Test of Insight, these results represent a 28.5% improvement (Table 3.7). As for accuracy, 6 out of 7 total questions were answered correctly for a score of 85.71%. The results of this experiment were of higher accuracy than either the experiment for supporting future trend prediction or 4th future prediction test.

Table3.8: Results of prototype method for future prediction. Numbers in brackets represent number of FRS used in training for each question. Correct answer is marked with an asterisk("*"). For each question highest score achieved by prototype method is marked in bold type font.

| | correct (*) | alternative/ SRL | score $\geq$ 0.00 | score $\geq$ 0.98 |
|---|---|---|---|---|
| Q1-1 | | | (16) | (10) |
| | | (a) The right to participate in regional politics will be established for foreign permanent residents by the end of June, 2010. | | |
| | | [Time] [Noun] [Reason] [Action] [Creation] [State-change] | 1.82 | 1.82 |
| | * | (b) The right to participate in regional politics will not be established for foreign permanent residents by the end of June, 2010. | | |
| | | [Time] [Noun] [Reason] [Action] [Action] [Action] | **2.00** | **2.00** |
| Q1-2 | | | (18) | (7) |
| | | (a) By the end of June, 2010 the civil code will be amended to recognize married couples who retain separate family names. | | |
| | | [Time] [Noun] [Reason] [Object] [State-change] [Creation] [State-change] | 2.05 | 1.75 |
| | * | (b) By the end of June, 2010 the civil code will not have been amended to recognize married couples who retain separate family names. | | |
| | | [Time] [Noun] [Reason] [Object] [State-change] [Action] [Action] | **2.28** | **2.18** |
| Q2 | | | (30) | (26) |
| | | (a) The Democratic Party will take the majority in both the House and the Senate. | | |
| | | [Organization] [Organization] [Other] [Verb] | 1.67 | 1.67 |
| | * | (b) The Democratic Party will take a majority in the Senate, and the Republican Party will take a majority in the House. | | |
| | | [Organization] [Organization] [Organization] [Organization] [Other] [Verb] | **1.70** | **1.70** |
| | | (c) The Republican Party will take a majority in the Senate, and the Democratic Party will take a majority in the House. | | |
| | | [Organization] [Organization] [Organization] [Organization] [Other] [Verb] | **1.70** | **1.70** |
| | | (d) The Republican Party will take the majority in both the House and the Senate. | | |
| | | [Organization] [Organization] [Other] [Verb] | 1.67 | 1.67 |
| Q3 | | | (30) | (172) |
| | * | (a) More reinforcements will stationed than in October, 2009. | | |
| | | [Time] [Noun] [Action] [Action] | 1.71 | 2.17 |
| | | (b) The same amount of troops as in October, 2009 will be stationed. | | |
| | | [Time] [Noun] [Noun] [Other] [Action] | 1.82 | 2.22 |
| | | (c) The number of troops stationed will be reduced than in October, 2009. | | |
| | | [Time] [Noun] [Object-role] [State-change] [Action] | **2.04** | **2.41** |
| | | (d) The US military will be completely withdrawn. | | |
| | | [Agent] [Action] | 1.50 | 1.91 |
| Q4 | | | (30) | (86) |
| | | (a) The reception of digital broadcasting will be enabled throughout Japan, while analog broadcasting will be simultaneously terminated. | | |
| | | [Place] [Action] [State-change] [Auxiliary verb] [Place] [Agent(experiencer)] [Time] [State-change] | **2.85** | **2.79** |
| | | (b) The reception of digital broadcasting will be enabled throughout Japan, while analog broadcasting will continue in certain regions. | | |
| | | [Place] [Action] [State-change] [Other] [Verb] [Things] [Place] [Agent(creation)] [No state change(activity)] | 2.80 | 2.70 |
| | | (c) The reception of digital broadcasting will not be enabled throughout all of Japan, and analog broadcasting will continue in certain regions. | | |
| | | [Place] [Action] [State-change] [Other] [Verb] [Things] [Place] [Agent(creation)] [No state change(activity)] | 2.801 | 2.70 |
| | * | (d) Analog broadcasting will be terminated in certain areas, but it will continue in most regions. | | |
| | | [Thing] [Place] [Agent(creation)] [State-change] [Things] [Place] [No state change(activity)] | 2.56 | 2.32 |
| | | (e) No regions will terminate analog broadcasting by July 24th, 2011. | | |
| | | [Time] [Noun] [Time] [Agent(creation)] [State-change] [Place] [adjective] | 2.70 | 2.60 |
| Q5 | | | (30) | (267) |
| | | (a) As promised, the full amount will be rationed out (26,000 yen per person) regardless of income levels. | | |
| | | [Action] [Action] [Action] [Action] | **2.73** | **2.95** |
| | * | (b) A smaller amount will be rationed out, regardless of income levels. | | |
| | | [Action] [Action] [State-change] | 2.11 | 2.23 |
| | | (c) The full amount will be rationed out, with limitations depending on income. | | |
| | | [Action] [Action] | 1.67 | 2.00 |
| | | (d) A smaller amount will be rationed out, with limitations depending on income. | | |
| | | [Action] [Action] [State-change] | 2.11 | 2.23 |
| | | (e) A child allowance will not be implemented. | | |
| | | [Action] | 1 | 1 |
| Q6 | | | (30) | (392) |
| | | (a) The Democratic Party will take the sole majority. | | |
| | | [Organization] [Noun] [Verb] | 1.40 | 2.00 |
| | | (b) A coalition including but not limited to the Democratic Party will take the majority. | | |
| | | [Organization] [Noun] [Verb] [Object] [State-change] [Organization] [Other] [Verb] | **2.62** | **3.24** |
| | * | (c) A coalition including the Democratic Party will not be able to maintain the majority. | | |
| | | [Object] [State-change] [Organization] [Other] [Action] | 2.18 | 2.69 |

### 3.2.2 Discussion

In this experiment, I systematized the thought processes a human goes through when they read FPSS and answer questions to predict the future, and conducted experiments evaluating the accuracy of the system. We found that my prediction results were 21.43% more accurate than human predictions, and 28.74% better than the 4th future prediction test. In addition, 5 out of 7 questions were answered correctly, for an accuracy of 71.43%. In comparison with an average accuracy of 30.3% in the prediction support experiment (high: 85.71%, low: 14.29%), this represents a 2-fold increase. Consideration of the accuracy on each question shows that the questions answered correctly and incorrectly are quite similar to the results of the prediction experiment. This fact is extremely interesting. I think there is a possibility that increasing the problem samples and the number of times the experiment is run in further studies may shed light on the relationship between this method of predicting the future and the human thought process.

In support statement training, a comparison of predictions made using MoPs obtained as pattern 1 and pattern 2 produced identical results ( Figure 3.6 ). This suggests a weak relationship between predictive accuracy and the number of statements to be learned. When learning with an increased number of future prediction supporting sentences, a larger number of statements produced a higher score, but had no effect on ranking. Further, even when future prediction supporting sentences and the length of the SRL of the question options differed, because MoPs consist of 1 – 6 elements, this method is effective even for short sentences like the question options.

On the other hand, focusing on the length of the SRL elements within the question options reveals that the SRL lengths of Q1–2, Q–1–2, and Q3 are approximately equal. In this case, I think that means my method is working effectively. For

example, the SRLs for the options of Q6 are configured as follows:

(a) organization noun verb

(b) organization noun verb object state-change organization other verb

(c) object state-change organization other verb

Even when there is a difference in the length of the morphosemantic structure, as with (a) and (b), since MoPs are assigned a score calculated by the weight of the elements, if MoPs with longer elements have contributed, the score may increase accordingly. We suspect that using an alternate version (total of 14 versions) that does not calculate weights may provide a clue toward improving the PM's accuracy in selecting answers from the question options. And though it was mentioned earlier, it was proven again in this experiment that stable results were obtained by training SPEC on far less data than would be typically required for machine learning. I see increasing accuracy as a challenge for the future.

# Chapter 4
# Conclusions and Future Work

## 4.1 Conclusion

We have pursued this research with the hypothesis that when they predict future trends, humans find sentences referring to the future helpful.

In Chapter 2, I conducted a survey of sentences referencing the future, centered primarily on newspaper corpora. We found that these expressions varied greatly, expressed things not limited to the words and phrases, and further, that temporal expressions alone are not comprehensive.

We proposed the morphosemantic method, a combination of morphology and semantics, as a way to find sentences implicitly referencing the future. The morphosemantic method extracts MoPs by training SPEC on the morphosemantic structure, which combines compound-clause processing with SRL based on verb argument structure and assigns SRLs to FRS. The morphosemantic method extracted MoPs by training SPEC on the morphosemantic structure, which synthesizes compound-clause processing with SRL based on predicate argument structure and assigns SRLs to FRS.

The highest results achieved were F=0.71 with P=0.56 and R=0.98 for the version of the classifier which used pattern list with zero-patterns deleted and length awarded.

Next I compared in detail the results between the two datasets, set50 and set130. For set50, the F-score reached plateau at around 0.67-0.71 for patterns and 0.67-0.70 for n-grams.

Also, the performance for the algorithm as a whole is similar for set50 and set130. Larger dataset usually contains more ambiguities, thus the results would be expected to degrade. With the proposed approach the differences are negligible and statistically not significant.

Experiments using small amounts of data are possible with SPEC, as is processing without vectorizing the morphosemantic structure; furthermore, I were able to confirm which of the MoPs thereby obtained was contributing.

For verification, I ran an experiment to extract FRS from 170 statements in the Mainichi Newspaper using MoPs (5–15) with a high length_awarded score (set130). Our results obtained accuracy of F-score = 0.43 on a mere 5–15 MoPs at most. These results are a 0.33 improvement over the method of Jatowt et al. [11], verifying effectiveness of my FRS extraction method.

In Chapter 3, I conducted a validation experiment to determine whether FRS is effective in supporting future trend prediction. We drew questions from the future prediction test and, using keywords from those questions, gathered newspaper articles from the entire year of 2009, extracted FPSS from them, and had 30 laypeople read these and respond. The results yielded only a 0.023 improvement over the results of the future prediction test. However, if one considers that the future prediction test allows respondents to spend a lot of time on it, use any available information source, and seek the opinions of others, while my experiment had subjects reply immediately after reading a mere 30 FPSSs, then the significance of these results for prediction support has surely been demonstrated.

Furthermore, I built a future prediction system and conducted an experiment to automatically read questions and predict answers. With a score of 57.14%, I achieved results nearly twice as good as human answers.

Only one FRS was extracted for support, but my discussion clarified that using this sentence as a key phrase to search for articles yielded very detailed information. We think it is easier to refine the information desired with this method than with a keyword search. Furthermore, I believe that a combination of FRS and statistical data in future trend predictions increases the potential for obtaining information that cannot be reduced to numbers.

Looking forward, I are planning to use this method with other corpora to conduct experiments on real-world problems (company management support, economic trend prediction, and so on). Also, carrying out a timeline analysis of FRS and the addition of polarity could lead to the discovery of new knowledge.

# References


[1] K. Radinsky, S. Davidovich, and S. Markovitch, "Learning causality for news events prediction," Proceedings of the 21st International Conference on World Wide Web, pp.909–918, ACM, 2012.

[2] H. SAKAJI and S. MASUYAMA, "A method for extracting sentences including causal relations from newspaper articles," The IEICE transactions on information and systems (Japanese edetion), vol.94, no.8, pp.1496–1506, aug 2011.

[3] H. SAKAJI and S. MASUYAMA, "An extraction method of causal knowledge from newspaper corpus," IEICE technical report. Natural language understanding and models of communication, vol.111, no.119, pp.7–10, jun 2011.

[4] R. Baeza-Yates, "Searching the future," SIGIR Workshop on MF/IR, 2005.

[5] N. Kanhabua, R. Blanco, and M. Matthews, "Ranking related news predictions," Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval, pp.755–764, 2011.

[6] K. Kanazawa, A. Jatowt, S. Oyama, and K. Tanaka., "Exracting explicit and implicit future-related information from the web(o) (in japanese)," DEIM Forum 2010, pp.paper ID: A9–1, 2010.

[7] O. Alonso, J. Strotgen, R. Baeza-Yates, and M. Gertz, "Temporal information retrieval: Challenges and opportunities," TWAW, vol.11, pp.1–8, 2011.

[8] K. Kanazawa, A. Jatowt, and K. Tanaka, "Improving retrieval of future-related information in text collections," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), pp.278–283, 2011.

[9] A. Jatowt, H. Kawai, K. Kanazawa, K. Tanaka, K. Kunieda, and K. Yamada, "Multi–lingual, longitudinal analysis of future-related information on the web," Proceedings of the 4th International conference on Culture and Computing 2013, pp.27–32, IEEE Press, 2013.

[10] O. Popescu and C. Strapparava, "Behind the times: Detecting epoch changes using large corpora," International Joint Conference on Natural Language Processing, pp.347–355, 2013.

[11] A. Jatowt and C. man Au Yeung, "Extracting collective expectations about the future from large text collections," Proceedings of the 20th ACM international conference on Information and knowledge management, pp.1259–1264, 2011.

[12] A. Jatowt, K. Kanazawa, S. Oyama, and K. Tanaka, "Supporting analysis of future-related information in news archives and the web," Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries, pp.115–124, 2009.

[13] E. Aramaki, S. Maskawa, and M. Morita, "Twitter catches the flu: Detecting influenza epidemics using twitter," Proceedings of the Conference on Empirical Methods in Natural Language Processing, pp.1568–1576, 2011.

[14] H. Sakaji, S. Sekine, and S. Masuyama, "Extracting causal knowledge using clue phrases and syntactic patterns," in Practical Aspects of Knowledge Management, pp.111–122, Springer, 2008.

[15] S. TANIGUCHI, H. SAKAJI, H. SAKAI, and S. MASUYAMA, "A polarity assignment method to basis expressions of economic trends extracted from economic newspaper articles," The IEICE transactions on information and systems (Japanese edetion), vol.94, no.6, pp.1039–1043, jun 2011.

[16] O. Y. and M. P., Chance Discovery, Springer, 2003.

[17] K. Takeuchi, K. Inui, N. Takeuchi, and A. Fujita, "A thesaurus of predicate-argument structure for japanese verbs to deal with granularity of verb meanings," The 8th Workshop on Asian Language Resources, pp.1–8, 2010.

[18] A.Z. Broder, S.C. Glassman, M.S. Manasse, and G. Zweig, "Syntactic clustering of the web," Computer Networks and ISDN Systems, vol.29, no.8, pp.1157–1166, 1997.

[19] J. Cohen *et al.*, "A coefficient of agreement for nominal scales," Educational and psychological measurement, vol.20, no.1, pp.37–46, 1960.

[20] J.L. Fleiss, "Measuring nominal scale agreement among many raters.," Psychological bulletin, vol.76, no.5, p.378, 1971.

[21] K. Takeuchi, S. Tsuchiyama, M. Moriya, and Y. Moriyasu, "Construction of argument structure analyzer toward searching same situations and actions," IEICE Technical Report, vol.109, no.390, pp.1–6, 2010.

[22] J. Bresnan, Lexical-functional syntax, Wiley-Blackwell, 2001.

[23] B. Levin and M.R. Hovav, "Morphology and lexical semantics," in The handbook of morphology, ed. eds, pp.248–271, Oxford: Blackwell, 1998.

[24] P. Kroeger, "Morphosyntactic vs. morphosemantic functions of indonesian -*kan*," in Architectures, Rules, and Preferences: Variations on Themes of Joan

Bresnan, ed. A. Zaenen, J. Simpson, T.H. King, J. Grimshaw, J. Maling, and C. Manning, pp.229–251, Stanford, CA: CSLI Publications, 2007.

[25] C. Fellbaum, A. Osherson, and P.E. Clark, "Putting semantics into wordnet's morphosemantic links," Human Language Technology. Challenges of the Information Society, vol.5603, pp.350–358, 2009.

[26] I. Raffaelli, "The model of morphosemantic patterns in the description of lexical architecture," Lingue e linguaggio, vol.12, no.1, pp.47–72, 2013.

[27] T. Koichi, T. Suguru, M. Masato, M. Yuuki, and S. Koichi, "Verb sense disambiguation based on thesaurus of predicate-argument structure-an evaluation of thesaurus of predicate-argument structure for japanese verbs.," KEOD, pp.208–213, 2011.

[28] Y. Kobayashi, T. Tokunaga, and H. Tanaka, "Meishi-kan no imi-teki kyōki jōhō o mochiita fukugō meishi no kaiseki. (analysis of compound nouns using semantic co-occurrence information between nouns) [in japanese]," Shizen Gengo Shori (Natural Language Processing), vol.3, no.1, pp.29–43, 1996.

[29] Y. Matsumoto, "Nihongo no goi-teki fukugō dōshi ni okeru dōshi no kumiawase.(combinations of verbs in lexical compound verbs in japanese) [in japanese]," Gengo Kenkyū (Language Research), vol.1998, no.114, pp.37–83, 1998.

[30] M. Ptaszynski, R. Rzepka, K. Araki, and Y. Momouchi, "Language combinatorics: A sentence pattern extraction architecture based on combinatorial explosion," International Journal of Computational Linguistics (IJCL), vol.2, no.1, pp.24–36, 2011.

[31] M. Ptaszynski, F. Masui, R. Rzepka, and K. Araki, "Emotive or non-emotive: That is the question," the 5th Workshop on Computational Approaches to

Subjectivity, Sentiment & Social Media Analysis (WASSA 2014), pp.59–65, ACL, 2014.

[32] M. Ptaszynski, D. Hasegawa, and F. Masui, "Women like backchannel, but men finish earlier: Pattern based language modeling of conversations reveals gender and social distance differences," Samsung HLT Young Researchers Symposium colocated with 9th International Conference on Natural Language Processing (PolTAL 2014), p.paper ID:02, 2014.

[33] M. Ptaszynski, F. Masui, Y. Kimura, R. Rzepka, and K. Araki, "Extracting patterns of harmful expressions for cyberbullying detection," 7th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics (LTC'15), 2015.

[34] T. Kurokawa and H. Kakeya, "Analysis of a tendency to answer in foresight official approval (in japanese)," The fifth media information inspection arts and sciences meeting, pp.50–55, 2009.

# An Appendix

**A:** Classification Results (in Section 2.5.3, Sections 2.5.5)

**A-1:** All results averaged for all folds in 10-fold cross validation for set50.

**A-2:** All results averaged for all folds in 10-fold cross validation for set130.

**B:** Evaluation Experiment of Future Trend Prediction (in Section 3.1)

**B-1:** Future prediction supporting sentence (FPSS) for question3 in Japanese.

**A-1** Results from all experiments. All results averaged for all folds in 10-fold cross validation. Experiments performed for all versions of the algorithm, with all pattern list modifications, all ways of weight calculation, summarized separately for patterns and n-grams for **set50**.

unmodified pattern list

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.50 | 0.50 | 0.61 | 0.60 | 0.55 | 0.56 | 0.54 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.16 | 0.34 | 0.56 | 0.76 | 0.84 | 0.94 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.24 | 0.40 | 0.58 | 0.67 | 0.66 | 0.71 | 0.70 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.48 | 0.53 | 0.56 | 0.60 | 0.62 | 0.57 | 0.60 | 0.57 | 0.52 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 0.96 | 0.90 | 0.78 | 0.64 | 0.48 | 0.30 | 0.26 | 0.16 | 0.04 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | -0.07 | 0.07 | 0.10 | 0.21 | 0.26 | 0.16 | 0.25 | 0.20 | 0.07 | 0.03 | 0.03 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.43 | 0.57 | 0.61 | 0.58 | 0.55 | 0.56 | 0.54 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.16 | 0.38 | 0.56 | 0.78 | 0.84 | 0.94 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.23 | 0.45 | 0.58 | 0.67 | 0.66 | 0.71 | 0.70 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.47 | 0.51 | 0.58 | 0.60 | 0.61 | 0.57 | 0.60 | 0.57 | 0.52 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 0.94 | 0.86 | 0.78 | 0.64 | 0.44 | 0.30 | 0.26 | 0.16 | 0.04 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | -0.10 | 0.01 | 0.16 | 0.21 | 0.24 | 0.16 | 0.25 | 0.20 | 0.07 | 0.03 | 0.03 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.10 | 0.22 | 0.37 | 0.50 | 0.50 | 0.53 | 0.56 | 0.54 | 0.55 | 0.54 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 |
| Recall | 0.06 | 0.12 | 0.28 | 0.38 | 0.52 | 0.66 | 0.82 | 0.86 | 0.92 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.07 | 0.15 | 0.32 | 0.43 | 0.51 | 0.59 | 0.67 | 0.66 | 0.69 | 0.69 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.48 | 0.48 | 0.54 | 0.53 | 0.54 | 0.54 | 0.59 | 0.56 | 0.58 | 0.56 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 |
| Specificity | 0.90 | 0.84 | 0.80 | 0.68 | 0.56 | 0.42 | 0.36 | 0.26 | 0.24 | 0.16 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | -0.07 | -0.04 | 0.08 | 0.07 | 0.07 | 0.08 | 0.20 | 0.14 | 0.20 | 0.15 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.00 | 0.00 | 0.00 |

length awarded

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.45 | 0.43 | 0.41 | 0.52 | 0.61 | 0.57 | 0.58 | 0.56 | 0.56 | 0.54 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 |
| Recall | 0.52 | 0.54 | 0.56 | 0.64 | 0.74 | 0.80 | 0.90 | 0.92 | 0.96 | 0.98 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.49 | 0.48 | 0.47 | 0.57 | 0.67 | 0.67 | 0.71 | 0.69 | 0.71 | 0.69 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.62 | 0.60 | 0.57 | 0.60 | 0.62 | 0.60 | 0.62 | 0.59 | 0.60 | 0.56 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 |
| Specificity | 0.72 | 0.66 | 0.58 | 0.56 | 0.50 | 0.40 | 0.34 | 0.26 | 0.24 | 0.14 | 0.04 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 |
| phi-coefficient | 0.25 | 0.21 | 0.14 | 0.22 | 0.26 | 0.21 | 0.28 | 0.23 | 0.27 | 0.17 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.00 | 0.00 |

length awarded zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.63 | 0.58 | 0.56 | 0.57 | 0.56 | 0.55 | 0.56 | 0.55 | 0.56 | 0.53 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 |
| Recall | 0.76 | 0.78 | 0.82 | 0.88 | 0.88 | 0.90 | 0.92 | 0.94 | 0.98 | 0.98 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.69 | 0.67 | 0.67 | 0.69 | 0.69 | 0.69 | 0.70 | 0.69 | 0.71 | 0.69 | 0.67 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.66 | 0.61 | 0.59 | 0.60 | 0.59 | 0.58 | 0.59 | 0.58 | 0.60 | 0.55 | 0.51 | 0.52 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 |
| Specificity | 0.56 | 0.44 | 0.36 | 0.32 | 0.30 | 0.26 | 0.26 | 0.22 | 0.22 | 0.12 | 0.04 | 0.04 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.00 | 0.00 |
| phi-coefficient | 0.34 | 0.25 | 0.20 | 0.23 | 0.21 | 0.18 | 0.22 | 0.20 | 0.27 | 0.15 | 0.03 | 0.07 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.00 | 0.00 |

length awarded ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.56 | 0.56 | 0.55 | 0.54 | 0.54 | 0.54 | 0.54 | 0.55 | 0.55 | 0.53 | 0.51 | 0.50 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 |
| Recall | 0.84 | 0.88 | 0.90 | 0.90 | 0.90 | 0.92 | 0.94 | 0.96 | 0.98 | 0.98 | 0.98 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.67 | 0.68 | 0.69 | 0.68 | 0.67 | 0.68 | 0.69 | 0.70 | 0.70 | 0.69 | 0.67 | 0.66 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.58 | 0.58 | 0.58 | 0.57 | 0.56 | 0.57 | 0.57 | 0.58 | 0.58 | 0.55 | 0.52 | 0.50 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 |
| Specificity | 0.32 | 0.28 | 0.26 | 0.24 | 0.22 | 0.22 | 0.20 | 0.20 | 0.18 | 0.12 | 0.06 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 | 0.02 |
| phi-coefficient | 0.17 | 0.18 | 0.18 | 0.17 | 0.13 | 0.17 | 0.17 | 0.20 | 0.22 | 0.13 | 0.07 | 0.00 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 | 0.03 |

ngrams

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.00 | 0.10 | 0.30 | 0.46 | 0.63 | 0.65 | 0.56 | 0.55 | 0.53 | 0.52 | 0.52 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.08 | 0.18 | 0.48 | 0.62 | 0.76 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.00 | 0.03 | 0.13 | 0.26 | 0.55 | 0.63 | 0.64 | 0.70 | 0.69 | 0.69 | 0.69 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.49 | 0.50 | 0.50 | 0.50 | 0.58 | 0.61 | 0.58 | 0.58 | 0.55 | 0.54 | 0.54 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 0.98 | 0.98 | 0.92 | 0.82 | 0.68 | 0.60 | 0.40 | 0.20 | 0.10 | 0.08 | 0.08 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | -0.03 | 0.00 | 0.00 | 0.02 | 0.18 | 0.24 | 0.18 | 0.20 | 0.17 | 0.13 | 0.13 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ngrams zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.05 | 0.15 | 0.28 | 0.41 | 0.63 | 0.60 | 0.56 | 0.55 | 0.53 | 0.52 | 0.52 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.02 | 0.04 | 0.08 | 0.20 | 0.48 | 0.66 | 0.78 | 0.96 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.03 | 0.06 | 0.12 | 0.27 | 0.54 | 0.63 | 0.66 | 0.70 | 0.69 | 0.69 | 0.69 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.49 | 0.50 | 0.49 | 0.50 | 0.57 | 0.59 | 0.59 | 0.58 | 0.55 | 0.54 | 0.54 | 0.52 | 0.52 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 0.96 | 0.96 | 0.90 | 0.80 | 0.66 | 0.52 | 0.40 | 0.20 | 0.10 | 0.08 | 0.08 | 0.04 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | -0.03 | 0.00 | -0.02 | 0.00 | 0.16 | 0.21 | 0.20 | 0.20 | 0.17 | 0.13 | 0.13 | 0.07 | 0.07 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ngrams ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.56 | 0.56 | 0.57 | 0.59 | 0.61 | 0.59 | 0.60 | 0.60 | 0.57 | 0.57 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 |
| Recall | 0.32 | 0.32 | 0.34 | 0.38 | 0.52 | 0.60 | 0.64 | 0.74 | 0.76 | 0.76 | 0.94 | 0.94 | 0.94 | 0.94 | 0.94 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 1.00 |
| F-score | 0.41 | 0.41 | 0.43 | 0.46 | 0.56 | 0.59 | 0.62 | 0.66 | 0.65 | 0.65 | 0.67 | 0.67 | 0.67 | 0.67 | 0.66 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.51 | 0.52 | 0.56 | 0.57 | 0.59 | 0.60 | 0.56 | 0.56 | 0.54 | 0.54 | 0.54 | 0.54 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.50 |
| Specificity | 0.68 | 0.68 | 0.68 | 0.66 | 0.60 | 0.54 | 0.54 | 0.46 | 0.36 | 0.36 | 0.14 | 0.14 | 0.14 | 0.14 | 0.10 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.00 |
| phi-coefficient | 0.02 | 0.02 | 0.04 | 0.06 | 0.14 | 0.15 | 0.20 | 0.23 | 0.15 | 0.15 | 0.10 | 0.10 | 0.10 | 0.10 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.07 | 0.00 |

ngrams length awarded

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.10 | 0.40 | 0.52 | 0.62 | 0.62 | 0.55 | 0.61 | 0.66 | 0.62 | 0.56 | 0.54 | 0.55 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.50 | 0.50 |
| Recall | 0.02 | 0.10 | 0.16 | 0.22 | 0.32 | 0.40 | 0.50 | 0.66 | 0.72 | 0.76 | 0.86 | 0.98 | 0.98 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.03 | 0.16 | 0.24 | 0.33 | 0.42 | 0.46 | 0.55 | 0.66 | 0.66 | 0.64 | 0.66 | 0.71 | 0.68 | 0.68 | 0.69 | 0.69 | 0.69 | 0.68 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.54 | 0.53 | 0.53 | 0.54 | 0.54 | 0.58 | 0.63 | 0.62 | 0.58 | 0.55 | 0.58 | 0.54 | 0.54 | 0.54 | 0.54 | 0.54 | 0.53 | 0.51 | 0.50 | 0.50 |
| Specificity | 0.98 | 0.98 | 0.90 | 0.84 | 0.76 | 0.68 | 0.66 | 0.60 | 0.52 | 0.40 | 0.24 | 0.18 | 0.10 | 0.10 | 0.08 | 0.08 | 0.08 | 0.06 | 0.02 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.12 | 0.11 | 0.11 | 0.11 | 0.09 | 0.18 | 0.29 | 0.26 | 0.18 | 0.11 | 0.22 | 0.13 | 0.13 | 0.13 | 0.13 | 0.13 | 0.10 | 0.03 | 0.00 | 0.00 |

ngrams length awarded zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.25 | 0.35 | 0.46 | 0.61 | 0.56 | 0.55 | 0.62 | 0.66 | 0.61 | 0.55 | 0.54 | 0.55 | 0.53 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.50 |
| Recall | 0.06 | 0.10 | 0.16 | 0.22 | 0.32 | 0.40 | 0.52 | 0.68 | 0.72 | 0.76 | 0.86 | 0.98 | 0.98 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.10 | 0.16 | 0.24 | 0.32 | 0.41 | 0.46 | 0.56 | 0.67 | 0.66 | 0.64 | 0.66 | 0.71 | 0.69 | 0.68 | 0.69 | 0.69 | 0.69 | 0.68 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.51 | 0.53 | 0.51 | 0.53 | 0.52 | 0.53 | 0.57 | 0.63 | 0.61 | 0.57 | 0.55 | 0.58 | 0.55 | 0.54 | 0.54 | 0.54 | 0.54 | 0.53 | 0.51 | 0.51 | 0.50 |
| Specificity | 0.96 | 0.96 | 0.86 | 0.84 | 0.72 | 0.66 | 0.62 | 0.58 | 0.50 | 0.38 | 0.24 | 0.18 | 0.12 | 0.10 | 0.08 | 0.08 | 0.08 | 0.06 | 0.02 | 0.02 | 0.00 |
| phi-coefficient | 0.03 | 0.08 | 0.05 | 0.10 | 0.05 | 0.07 | 0.16 | 0.29 | 0.25 | 0.15 | 0.11 | 0.22 | 0.17 | 0.13 | 0.13 | 0.13 | 0.13 | 0.10 | 0.03 | 0.03 | 0.00 |

ngrams length awarded ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.64 | 0.64 | 0.64 | 0.61 | 0.60 | 0.61 | 0.60 | 0.57 | 0.57 | 0.58 | 0.53 | 0.53 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 |
| Recall | 0.62 | 0.62 | 0.66 | 0.68 | 0.74 | 0.76 | 0.78 | 0.78 | 0.82 | 0.84 | 0.92 | 0.92 | 0.94 | 0.94 | 0.94 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 | 0.96 |
| F-score | 0.63 | 0.63 | 0.65 | 0.65 | 0.66 | 0.67 | 0.68 | 0.66 | 0.68 | 0.69 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.61 | 0.61 | 0.62 | 0.61 | 0.61 | 0.62 | 0.61 | 0.57 | 0.58 | 0.59 | 0.54 | 0.54 | 0.53 | 0.53 | 0.53 | 0.54 | 0.54 | 0.53 | 0.52 | 0.52 | 0.52 |
| Specificity | 0.60 | 0.60 | 0.58 | 0.54 | 0.48 | 0.48 | 0.44 | 0.36 | 0.34 | 0.34 | 0.16 | 0.16 | 0.12 | 0.12 | 0.12 | 0.12 | 0.12 | 0.10 | 0.08 | 0.08 | 0.08 |
| phi-coefficient | 0.24 | 0.24 | 0.26 | 0.24 | 0.24 | 0.27 | 0.24 | 0.17 | 0.20 | 0.21 | 0.11 | 0.11 | 0.09 | 0.09 | 0.09 | 0.12 | 0.12 | 0.10 | 0.07 | 0.07 | 0.07 |

**A-2** Results from all experiments. All results averaged for all folds in 1-fold cross validation. Experiments performed for all versions of the algorithm, with all pattern list modifications, all ways of weight calculation, summarized separately for patterns and n-grams for **set130**.

unmodified pattern list

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.00 | 0.55 | 0.73 | 0.69 | 0.61 | 0.59 | 0.58 | 0.54 | 0.53 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.00 | 0.08 | 0.21 | 0.44 | 0.55 | 0.77 | 0.90 | 0.95 | 0.97 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.00 | 0.13 | 0.32 | 0.53 | 0.57 | 0.67 | 0.70 | 0.69 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.50 | 0.52 | 0.57 | 0.62 | 0.59 | 0.62 | 0.62 | 0.57 | 0.55 | 0.53 | 0.52 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 0.99 | 0.96 | 0.92 | 0.79 | 0.64 | 0.46 | 0.34 | 0.20 | 0.13 | 0.07 | 0.03 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | -0.02 | 0.07 | 0.18 | 0.25 | 0.19 | 0.25 | 0.31 | 0.24 | 0.20 | 0.12 | 0.08 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.20 | 0.55 | 0.72 | 0.67 | 0.61 | 0.59 | 0.57 | 0.54 | 0.53 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.02 | 0.08 | 0.23 | 0.44 | 0.58 | 0.77 | 0.90 | 0.95 | 0.97 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.03 | 0.13 | 0.35 | 0.53 | 0.60 | 0.67 | 0.70 | 0.69 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.50 | 0.52 | 0.57 | 0.61 | 0.61 | 0.61 | 0.62 | 0.57 | 0.55 | 0.52 | 0.52 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 0.99 | 0.96 | 0.90 | 0.78 | 0.63 | 0.45 | 0.33 | 0.20 | 0.13 | 0.07 | 0.03 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | 0.02 | 0.07 | 0.18 | 0.23 | 0.22 | 0.24 | 0.30 | 0.24 | 0.20 | 0.10 | 0.08 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.23 | 0.24 | 0.26 | 0.26 | 0.56 | 0.71 | 0.59 | 0.57 | 0.57 | 0.56 | 0.54 | 0.53 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.03 | 0.04 | 0.07 | 0.16 | 0.28 | 0.41 | 0.52 | 0.66 | 0.84 | 0.91 | 0.95 | 0.97 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.05 | 0.07 | 0.11 | 0.20 | 0.37 | 0.52 | 0.55 | 0.61 | 0.68 | 0.69 | 0.69 | 0.68 | 0.67 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.51 | 0.53 | 0.56 | 0.59 | 0.57 | 0.58 | 0.60 | 0.59 | 0.57 | 0.55 | 0.53 | 0.52 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 0.98 | 0.97 | 0.95 | 0.90 | 0.84 | 0.78 | 0.62 | 0.49 | 0.36 | 0.28 | 0.19 | 0.13 | 0.08 | 0.05 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.02 | 0.02 | 0.04 | 0.06 | 0.12 | 0.23 | 0.15 | 0.18 | 0.24 | 0.26 | 0.24 | 0.20 | 0.12 | 0.12 | 0.04 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

length awarded

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.53 | 0.53 | 0.52 | 0.49 | 0.63 | 0.64 | 0.59 | 0.57 | 0.58 | 0.56 | 0.54 | 0.53 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 |
| Recall | 0.42 | 0.48 | 0.51 | 0.54 | 0.60 | 0.64 | 0.76 | 0.82 | 0.89 | 0.92 | 0.94 | 0.94 | 0.96 | 0.98 | 0.98 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.47 | 0.51 | 0.52 | 0.51 | 0.61 | 0.64 | 0.67 | 0.67 | 0.70 | 0.69 | 0.68 | 0.67 | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.60 | 0.61 | 0.61 | 0.60 | 0.60 | 0.60 | 0.60 | 0.60 | 0.61 | 0.59 | 0.57 | 0.55 | 0.54 | 0.54 | 0.53 | 0.53 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 |
| Specificity | 0.77 | 0.74 | 0.72 | 0.65 | 0.59 | 0.55 | 0.45 | 0.38 | 0.33 | 0.26 | 0.19 | 0.15 | 0.12 | 0.10 | 0.08 | 0.07 | 0.05 | 0.03 | 0.02 | 0.02 | 0.02 |
| phi-coefficient | 0.20 | 0.23 | 0.23 | 0.20 | 0.21 | 0.22 | 0.24 | 0.24 | 0.29 | 0.25 | 0.22 | 0.17 | 0.17 | 0.18 | 0.15 | 0.13 | 0.11 | 0.08 | 0.04 | 0.04 | 0.04 |

length awarded zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.67 | 0.66 | 0.62 | 0.61 | 0.60 | 0.59 | 0.58 | 0.57 | 0.56 | 0.55 | 0.54 | 0.53 | 0.53 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 |
| Recall | 0.55 | 0.59 | 0.62 | 0.68 | 0.72 | 0.77 | 0.84 | 0.87 | 0.92 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.97 | 0.98 | 0.98 | 0.99 | 1.00 | 1.00 | 1.00 |
| F-score | 0.61 | 0.63 | 0.62 | 0.64 | 0.66 | 0.67 | 0.68 | 0.69 | 0.69 | 0.69 | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 | 0.67 | 0.68 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.63 | 0.64 | 0.62 | 0.62 | 0.62 | 0.61 | 0.61 | 0.61 | 0.60 | 0.58 | 0.57 | 0.55 | 0.55 | 0.54 | 0.53 | 0.53 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 |
| Specificity | 0.72 | 0.68 | 0.61 | 0.55 | 0.52 | 0.45 | 0.38 | 0.35 | 0.28 | 0.24 | 0.20 | 0.17 | 0.15 | 0.12 | 0.10 | 0.08 | 0.06 | 0.05 | 0.02 | 0.02 | 0.02 |
| phi-coefficient | 0.28 | 0.28 | 0.23 | 0.24 | 0.25 | 0.24 | 0.26 | 0.27 | 0.27 | 0.24 | 0.22 | 0.19 | 0.19 | 0.17 | 0.14 | 0.12 | 0.10 | 0.11 | 0.06 | 0.04 | 0.04 |

length awarded ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.59 | 0.58 | 0.57 | 0.57 | 0.57 | 0.55 | 0.56 | 0.56 | 0.54 | 0.54 | 0.53 | 0.53 | 0.53 | 0.53 | 0.52 | 0.52 | 0.52 | 0.53 | 0.52 | 0.51 | 0.50 |
| Recall | 0.69 | 0.75 | 0.78 | 0.82 | 0.85 | 0.86 | 0.90 | 0.92 | 0.92 | 0.92 | 0.93 | 0.94 | 0.95 | 0.96 | 0.96 | 0.97 | 0.97 | 0.99 | 0.99 | 1.00 | 1.00 |
| F-score | 0.64 | 0.66 | 0.66 | 0.67 | 0.68 | 0.67 | 0.69 | 0.70 | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 | 0.68 | 0.67 | 0.68 | 0.68 | 0.69 | 0.68 | 0.68 | 0.67 |
| Accuracy | 0.60 | 0.60 | 0.59 | 0.60 | 0.60 | 0.58 | 0.59 | 0.59 | 0.57 | 0.57 | 0.56 | 0.55 | 0.55 | 0.55 | 0.53 | 0.53 | 0.53 | 0.55 | 0.53 | 0.52 | 0.51 |
| Specificity | 0.51 | 0.45 | 0.40 | 0.38 | 0.35 | 0.31 | 0.28 | 0.27 | 0.23 | 0.22 | 0.18 | 0.17 | 0.15 | 0.13 | 0.11 | 0.10 | 0.10 | 0.10 | 0.08 | 0.04 | 0.02 |
| phi-coefficient | 0.22 | 0.23 | 0.22 | 0.24 | 0.25 | 0.23 | 0.25 | 0.26 | 0.22 | 0.22 | 0.19 | 0.18 | 0.17 | 0.17 | 0.14 | 0.14 | 0.14 | 0.19 | 0.13 | 0.10 | 0.04 |

ngrams

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.56 | 0.65 | 0.61 | 0.57 | 0.54 | 0.53 | 0.51 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.15 | 0.37 | 0.56 | 0.78 | 0.91 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.23 | 0.47 | 0.58 | 0.66 | 0.68 | 0.69 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.49 | 0.49 | 0.52 | 0.58 | 0.60 | 0.60 | 0.57 | 0.55 | 0.53 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.96 | 0.89 | 0.79 | 0.64 | 0.41 | 0.23 | 0.12 | 0.05 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.04 | -0.03 | 0.07 | 0.18 | 0.20 | 0.21 | 0.19 | 0.18 | 0.10 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ngrams zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.25 | 0.55 | 0.65 | 0.60 | 0.57 | 0.54 | 0.53 | 0.52 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.02 | 0.16 | 0.40 | 0.58 | 0.78 | 0.90 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.04 | 0.25 | 0.49 | 0.59 | 0.66 | 0.68 | 0.69 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.49 | 0.49 | 0.52 | 0.59 | 0.60 | 0.60 | 0.57 | 0.56 | 0.53 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 | 0.50 |
| Specificity | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.98 | 0.95 | 0.88 | 0.78 | 0.62 | 0.41 | 0.25 | 0.14 | 0.06 | 0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |
| phi-coefficient | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | -0.04 | -0.04 | 0.06 | 0.19 | 0.20 | 0.21 | 0.20 | 0.20 | 0.11 | 0.02 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

ngrams ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.51 | 0.51 | 0.53 | 0.54 | 0.55 | 0.56 | 0.55 | 0.54 | 0.54 | 0.55 | 0.52 | 0.51 | 0.52 | 0.53 | 0.52 | 0.51 | 0.51 | 0.51 | 0.51 | 0.51 | 0.50 |
| Recall | 0.38 | 0.38 | 0.40 | 0.43 | 0.52 | 0.62 | 0.63 | 0.74 | 0.78 | 0.82 | 0.89 | 0.89 | 0.95 | 0.95 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 1.00 |
| F-score | 0.43 | 0.43 | 0.45 | 0.48 | 0.53 | 0.58 | 0.59 | 0.62 | 0.64 | 0.66 | 0.65 | 0.65 | 0.67 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.51 | 0.51 | 0.52 | 0.54 | 0.54 | 0.56 | 0.56 | 0.56 | 0.57 | 0.58 | 0.53 | 0.52 | 0.54 | 0.54 | 0.53 | 0.52 | 0.52 | 0.52 | 0.52 | 0.52 | 0.50 |
| Specificity | 0.65 | 0.65 | 0.65 | 0.65 | 0.56 | 0.51 | 0.48 | 0.38 | 0.35 | 0.33 | 0.16 | 0.15 | 0.13 | 0.13 | 0.08 | 0.05 | 0.05 | 0.05 | 0.05 | 0.05 | 0.00 |
| phi-coefficient | 0.02 | 0.02 | 0.05 | 0.08 | 0.09 | 0.13 | 0.12 | 0.14 | 0.17 | 0.19 | 0.07 | 0.06 | 0.10 | 0.12 | 0.12 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.00 |

ngrams length awarded

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.00 | 0.20 | 0.42 | 0.57 | 0.52 | 0.59 | 0.59 | 0.59 | 0.55 | 0.58 | 0.55 | 0.54 | 0.53 | 0.52 | 0.51 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.00 | 0.02 | 0.07 | 0.15 | 0.22 | 0.33 | 0.45 | 0.54 | 0.62 | 0.77 | 0.83 | 0.88 | 0.92 | 0.98 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.00 | 0.03 | 0.12 | 0.24 | 0.30 | 0.42 | 0.51 | 0.56 | 0.58 | 0.66 | 0.66 | 0.67 | 0.67 | 0.68 | 0.68 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.49 | 0.50 | 0.52 | 0.52 | 0.55 | 0.57 | 0.58 | 0.56 | 0.60 | 0.57 | 0.57 | 0.55 | 0.54 | 0.53 | 0.52 | 0.52 | 0.50 | 0.50 | 0.50 |
| Specificity | 0.99 | 0.99 | 0.96 | 0.93 | 0.89 | 0.82 | 0.76 | 0.70 | 0.62 | 0.49 | 0.43 | 0.32 | 0.25 | 0.17 | 0.11 | 0.06 | 0.05 | 0.04 | 0.01 | 0.01 | 0.01 |
| phi-coefficient | -0.02 | -0.02 | -0.05 | -0.01 | 0.06 | 0.04 | 0.11 | 0.15 | 0.16 | 0.12 | 0.22 | 0.18 | 0.19 | 0.15 | 0.15 | 0.11 | 0.10 | 0.09 | 0.02 | 0.02 | 0.02 |

ngrams length awarded zero deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.00 | 0.10 | 0.25 | 0.54 | 0.57 | 0.58 | 0.59 | 0.58 | 0.60 | 0.56 | 0.58 | 0.55 | 0.54 | 0.53 | 0.52 | 0.52 | 0.51 | 0.51 | 0.50 | 0.50 | 0.50 |
| Recall | 0.00 | 0.01 | 0.02 | 0.10 | 0.17 | 0.25 | 0.34 | 0.44 | 0.55 | 0.65 | 0.78 | 0.84 | 0.89 | 0.92 | 0.96 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| F-score | 0.00 | 0.01 | 0.04 | 0.17 | 0.26 | 0.35 | 0.43 | 0.50 | 0.57 | 0.60 | 0.66 | 0.66 | 0.68 | 0.67 | 0.68 | 0.68 | 0.68 | 0.68 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.50 | 0.50 | 0.49 | 0.51 | 0.52 | 0.53 | 0.55 | 0.57 | 0.59 | 0.56 | 0.60 | 0.57 | 0.57 | 0.55 | 0.54 | 0.53 | 0.52 | 0.52 | 0.50 | 0.50 | 0.50 |
| Specificity | 0.99 | 0.99 | 0.96 | 0.92 | 0.88 | 0.81 | 0.76 | 0.69 | 0.63 | 0.48 | 0.43 | 0.31 | 0.25 | 0.18 | 0.12 | 0.07 | 0.05 | 0.04 | 0.01 | 0.01 | 0.01 |
| phi-coefficient | -0.02 | 0.00 | -0.03 | 0.05 | 0.06 | 0.08 | 0.11 | 0.13 | 0.18 | 0.13 | 0.23 | 0.18 | 0.20 | 0.15 | 0.14 | 0.12 | 0.10 | 0.09 | 0.02 | 0.02 | 0.02 |

ngrams length awarded ambiguous deleted

| | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Threshold | 1.00 | 0.90 | 0.80 | 0.70 | 0.60 | 0.50 | 0.40 | 0.30 | 0.20 | 0.10 | 0.00 | -0.10 | -0.20 | -0.30 | -0.40 | -0.50 | -0.60 | -0.70 | -0.80 | -0.90 | -1.00 |
| Precision | 0.53 | 0.53 | 0.53 | 0.53 | 0.54 | 0.56 | 0.56 | 0.56 | 0.55 | 0.54 | 0.52 | 0.52 | 0.52 | 0.52 | 0.51 | 0.51 | 0.51 | 0.51 | 0.52 | 0.52 | 0.51 |
| Recall | 0.55 | 0.55 | 0.58 | 0.59 | 0.68 | 0.77 | 0.80 | 0.80 | 0.82 | 0.83 | 0.88 | 0.88 | 0.88 | 0.89 | 0.90 | 0.92 | 0.92 | 0.93 | 0.95 | 0.96 | 0.98 |
| F-score | 0.54 | 0.54 | 0.56 | 0.56 | 0.61 | 0.65 | 0.66 | 0.66 | 0.66 | 0.66 | 0.65 | 0.65 | 0.65 | 0.66 | 0.65 | 0.66 | 0.66 | 0.66 | 0.67 | 0.67 | 0.67 |
| Accuracy | 0.53 | 0.53 | 0.53 | 0.53 | 0.56 | 0.58 | 0.59 | 0.58 | 0.58 | 0.57 | 0.53 | 0.53 | 0.53 | 0.53 | 0.52 | 0.52 | 0.52 | 0.52 | 0.53 | 0.53 | 0.52 |
| Specificity | 0.52 | 0.51 | 0.48 | 0.47 | 0.43 | 0.40 | 0.38 | 0.36 | 0.33 | 0.31 | 0.18 | 0.18 | 0.18 | 0.17 | 0.13 | 0.12 | 0.12 | 0.11 | 0.11 | 0.11 | 0.05 |
| phi-coefficient | 0.07 | 0.06 | 0.07 | 0.06 | 0.13 | 0.20 | 0.21 | 0.19 | 0.19 | 0.18 | 0.08 | 0.08 | 0.09 | 0.08 | 0.04 | 0.05 | 0.06 | 0.06 | 0.09 | 0.11 | 0.08 |

**B-1**: Examples of future prediction supporting FRS for question3 in Japanese. (ordered time line)

| No. | Future Reference sentence |
|---|---|
| 1 | 毎日が取り上げた北朝鮮の核問題での日米韓３国連携、アフガニスタン復興支援での日韓協力、首脳同士が訪問し合う「シャトル外交」の定着という三つの論点は、他紙も共通している。 |
| 2 | また、イラク駐留米軍戦闘部隊の段階的撤退によってイラク戦争を終結させ、アフガニスタンでの対テロ戦争に全力を挙げる意向を示した。 |
| 3 | また、イラク駐留米軍戦闘部隊の段階的撤退によってイラク戦争を終結させ、アフガニスタンでの対テロ戦争に全力を挙げる意向を示す。 |
| 4 | 公約である「就任後１６カ月以内」の駐留米軍撤退に向けた実質的な協議に着手し、対テロ戦争の主戦場と位置付けるアフガニスタンへの米軍増派を含む総合計画の早期策定を目指す。 |
| 5 | バグダッドのエンジニア、アフマド・サイフさん（２９）は、オバマ大統領が対テロ戦争の軸足をアフガニスタンに移す考えを改めて強調したことで、米軍の早期イラク撤退が現実味を増したと喜んだ |
| 6 | 各国の外相や財務相が参加する閣僚級会合で、アフガニスタンの旧支配勢力タリバンによる治安悪化への対策の他、干ばつ対策のための水資源開発などを議題とし、支援策について協議する。 |
| 7 | 会談では、アフガニスタン戦略に関し日米共同で検討することが決まり、今後、特使を米国に派遣して内容を深めていくことになった。 |
| 8 | ロシア外務省は６日に声明を発表し、両国にはアフガニスタン情勢やテロとの戦いなど立場が近接または一致するテーマがあり、外相会談の結果を「控えめに楽観している」と述べた。 |
| 9 | また両国は共同声明で、アフガニスタン、北朝鮮、イラン問題での協調を確認する一方、ＭＤの東欧配備計画やグルジア紛争の原因と評価については立場の相違があることを明記。 |
| 10 | その一方でアフガニスタンへの増派計画を発表しており、イラクからの撤退によって対テロ戦争の軸足をアフガンへ移す姿勢を示している。 |
| 11 | 対テロ対策の軸足をアフガニスタンに移す方針を打ち出したオバマ政権だが、その実現にはイラク安定化が不可欠。 |
| 12 | 進展を喜ぶが、満足はしていない」と述べ、経済危機の克服やイラク、アフガニスタン安定化、医療保険改革など重要課題解決に向けさらに努力する意向を示した。 |
| 13 | 銃撃後の自爆という手口は昨年９月、イスラマバードのマリオットホテルで起きた６０人以上死亡の自爆テロと酷似しており、当時、犯行声明を出したアフガニスタン国境の武装勢力のベイトラ・メスード司令官が今回も関与したとみられている。 |
| 14 | 両国の関係修復は、パキスタン軍が国内の武装勢力掃討に専念する必須条件といえ、対テロ戦争のアフガニスタン包括戦略を定めたオバマ米政権も対話の行方を注目しているとみられる。 |
| 15 | 大統領の不用意な一言が、その後のアフガニスタン攻撃やイラク戦争に「反イスラム」的な印象を与えてしまったことは否定できまい。 |
| 16 | 鳩山氏がインド洋での海自の給油活動の代替策としてアフガニスタンへの民生支援の充実を挙げたことに対しても「(民生支援は）既に実施している。 |
| 17 | さらに岡田氏は「１００日間で解決しなければいけない問題」として、北朝鮮や地球温暖化問題とならんで沖縄の在日米軍基地移転問題とアフガニスタン支援を挙げ、対米関係を最重視する考えをにじませた。 |
| 18 | 代替策は「(防衛省で）独自にやる話ではなく、政府でしっかり協議をしたい」と述べ、政府として新たなアフガニスタン支援策などを検討する方針を明らかにした。 |
| 19 | 国連アフガニスタン支援ミッション（ＵＮＡＭＡ）幹部は、「仮に決選投票をしても、今度は告発合戦となり、混乱はさらに深まるだろう」と打開策の見えない状況にため息をついた。 |
| 20 | 岡田克也外相は来年１月で期限が切れるインド洋での給油支援について「単純延長はしない」としているが、一方でアフガニスタン支援を継続する重要性も強調している。 |
| 21 | 医療保険改革、アフガニスタン戦争への増派問題という２大課題が大詰めを迎えている時期だけに、政権側は、「国際的に評価された大統領」を前面に押し出し、反転攻勢に出るとみられる。 |
| 22 | さらに「アフガニスタンにおける協力のほうを米国も重要視している」と述べ、新たなアフガン支援策の策定を優先させる考えを示した。 |
| 23 | そのうえで、「開発や人材育成などで有益な貢献策はいくらでもある」と述べ、日本がアフガニスタン支援策を新たに検討し、実行することへの期待感を明らかにした。 |
| 24 | とりわけ、アフガニスタンは今、テロの脅威に対処しつつ、国家を再建し、社会の平和と安定を目指しています。 |
| 25 | 米国ではアフガニスタン増派を巡って対立が激化しており、増派する場合でも将来の撤兵の見通しをつけるため、現在米軍が担っている治安維持機能を国軍が肩代わりできるよう強化すべきだとする議論が強まっている。 |
| 26 | ＡＰ通信などによると、ハサン被告は今月末にもアフガニスタンに派遣される予定だったが、従軍を希望していなかった。 |
| 27 | 韓国警察庁が逮捕したパキスタン国籍の男（３９）らは、日本などから無水酢酸を韓国に密輸入した後、イラン経由でアフガニスタンのカンダハルに密輸出を繰り返していたとみられる。 |
| 28 | その上で、イラクとアフガニスタンで二つの戦争を進める米軍の最高司令官として「戦争と平和」の問題について率直に語り、「必要な戦争」への理解を求めた。 |
| 29 | オバマ米政権も「パキスタンの民政支援」をアフガニスタン新戦略に掲げるとともに、無人機を使ったミサイル攻撃を拡大。 |
| 30 | 経済はとても回復軌道に乗ったとは言い難いし、アフガニスタン増派、テロ未遂とまだまだ国難が続く。 |