# Named Entity Chunking Techniques in Supervised Learning for Japanese Named Entity Recognition

**Manabu Sassano**
Fujitsu Laboratories, Ltd.
4-4-1, Kamikodanaka, Nakahara-ku,
Kawasaki 211-8588, Japan
sassano@flab.fujitsu.co.jp

**Takehito Utsuro**
Department of Information
and Computer Sciences,
Toyohashi University of Technology
Tenpaku-cho, Toyohashi 441-8580, Japan
utsuro@ics.tut.ac.jp

## Abstract

This paper focuses on the issue of named entity chunking in Japanese named entity recognition. We apply the supervised decision list learning method to Japanese named entity recognition. We also investigate and incorporate several named-entity noun phrase chunking techniques and experimentally evaluate and compare their performance. In addition, we propose a method for incorporating richer contextual information as well as patterns of constituent morphemes within a named entity, which have not been considered in previous research, and show that the proposed method outperforms these previous approaches.

## 1 Introduction

It is widely agreed that named entity recognition is an important step for various applications of natural language processing such as information retrieval, machine translation, information extraction and natural language understanding. In the English language, the task of named entity recognition is one of the tasks of the Message Understanding Conference (MUC) (e.g., MUC-7 (1998)) and has been studied intensively. In the Japanese language, several recent conferences, such as MET (Multilingual Entity Task, MET-1 (Maiorano, 1996) and MET-2 (MUC, 1998)) and IREX (Information Retrieval and Extraction Exercise) Workshop (IREX Committee, 1999), focused on named entity recognition as one of their contest tasks, thus promoting research on Japanese named entity recognition.

In Japanese named entity recognition, it is quite common to apply morphological analysis as a preprocessing step and to segment the sentence string into a sequence of morphemes. Then, hand-crafted pattern matching rules and/or statistical named entity recognizer are applied to recognize named entities. It is often the case that named entities to be recognized have different segmentation boundaries from those of morphemes obtained by the morphological analysis. For example, in our analysis of the IREX workshop's training corpus of named entities, about half of the named entities have segmentation boundaries that are different from the result of morphological analysis by a Japanese morphological analyzer BREAK-FAST (Sassano et al., 1997) (section 2). Thus, in Japanese named entity recognition, among the most difficult problems is how to recognize such named entities that have segmentation boundary mismatch against the morphemes obtained by morphological analysis. Furthermore, in almost 90% of cases of those segmentation boundary mismatches, named entities to be recognized can be decomposed into several morphemes as their constituents. This means that the problem of recognizing named entities in those cases can be solved by incorporating techniques of base noun phrase chunking (Ramshaw and Marcus, 1995).

In this paper, we focus on the issue of named entity chunking in Japanese named entity recognition. First, we take a supervised learning approach rather than a hand-crafted rule based approach, because the former is more promising than the latter with respect to the amount of human labor if requires, as well as its adaptability to a new domain or a new definition of named entities. In general, creating training data for supervised learning is somewhat easier than creating pattern matching rules by hand. Next, we apply Yarowsky's method for supervised decision list learning[1] (Yarowsky, 1994) to

---

[1] We choose the decision list learning method as the

Table 1: Statistics of NE Types of IREX

| NE Type | frequency (%) | |
| --- | --- | --- |
| | Training | Test |
| ORGANIZATION | 3676 (19.7) | 361 (23.9) |
| PERSON | 3840 (20.6) | 338 (22.4) |
| LOCATION | 5463 (29.2) | 413 (27.4) |
| ARTIFACT | 747 (4.0) | 48 (3.2) |
| DATE | 3567 (19.1) | 260 (17.2) |
| TIME | 502 (2.7) | 54 (3.5) |
| MONEY | 390 (2.1) | 15 (1.0) |
| PERCENT | 492 (2.6) | 21 (1.4) |
| Total | 18677 | 1510 |

Table 2: Statistics of Boundary Match vs. Mismatch of Morphemes (M) and Named Entities (NE)

| Match/Mismatch | | freq. of NE Tags (%) | |
| --- | --- | --- | --- |
| 1 M to 1 NE | | 10480 (56.1) | |
| $n(\geq 2)$ Ms to 1 NE | $n = 2$ | 4557 (24.4) | |
| | $n = 3$ | 1658 (8.9) | 7175 (38.4) |
| | $n \geq 4$ | 960 (5.1) | |
| other boundary mismatch | | 1022 (5.5) | |
| Total | | 18677 | |

Japanese named entity recognition, into which we incorporate several noun phrase chunking techniques (sections 3 and 4) and experimentally evaluate their performance on the IREX workshop's training and test data (section 5). As one of those noun phrase chunking techniques, we propose a method for incorporating richer contextual information as well as patterns of constituent morphemes within a named entity, compared with those considered in the previous research (Sekine et al., 1998; Borthwick, 1999), and show that the proposed method outperforms these approaches.

## 2 Japanese Named Entity Recognition

### 2.1 Task of the IREX Workshop

The task of named entity recognition of the IREX workshop is to recognize eight named entity types in Table 1 (IREX Committee, 1999). The organizer of the IREX workshop provided 1,174 newspaper articles which include 18,677 named entities as the training data. In the formal run (general domain) of the workshop, the participating systems were requested to recognize 1,510 named entities included in the held-out 71 newspaper articles.

### 2.2 Segmentation Boundaries of Morphemes and Named Entities

In the work presented here, we compare the segmentation boundaries of named entities in the IREX workshop's training corpus with those of

morphemes which were obtained through morphological analysis by a Japanese morphological analyzer BREAKFAST (Sassano et al., 1997).[2] Detailed statistics of the comparison are provided in Table 2. Nearly half of the named entities have boundary mismatches against the morphemes and also almost 90% of the named entities with boundary mismatches can be decomposed into more than one morpheme. Figure 1 shows some examples of such cases.[3]

## 3 Chunking and Tagging Named Entities

In this section, we formalize the problem of named entity chunking in Japanese named entity recognition. We describe a novel technique as well as those proposed in the previous works on named entity recognition. The novel technique incorporates richer contextual information as well as patterns of constituent morphemes within a named entity, compared with the techniques proposed in previous research on named entity recognition and base noun phrase chunking.

### 3.1 Task Definition

First, we will provide our definition of the task of Japanese named entity chunking. Suppose

---

supervised learning technique mainly because it is easy to implement and quite straightforward to extend a supervised learning version to a minimally supervised version (Collins and Singer, 1999; Cucerzan and Yarowsky, 1999). We also reported in (Utsuro and Sassano, 2000) the experimental results of a minimally supervised version of Japanese named entity recognition.

[2]The set of part-of-speech tags of BREAKFAST consists of about 300 tags. BREAKFAST achieves 99.6% part-of-speech accuracy against newspaper articles.

[3]In most cases of the "other boundary mismatch" in Table 2, one or more named entities have to be recognized as a part of a correctly analyzed morpheme and those cases are not caused by errors of morphological analysis. One frequent example of this type is a Japanese verbal noun "hou-bei (*visiting United States*)" which consists of two characters "hou (*visiting*)" and "bei (*United States*)", where "bei (*United States*)" has to be recognized as <LOCATION>. We believe that boundary mismatches of this type can be easily solved by employing a supervised learning technique such as the decision list learning method.

Table 3: Encoding Schemes of Named Entity Chunking States

| Named Entity Tag | | | <ORG> | | <LOC> | | | <LOC> | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Morpheme Sequence | ··· | $M$ | $M$ | $M$ | $M$ | $M$ | $M$ | $M$ | $M$ | ··· |
| Inside/Outside Encoding | | O | ORG_I | O | LOC_I | LOC_I | LOC_I | LOC_B | O | |
| Start/End Encoding | | O | ORG_U | O | LOC_S | LOC_C | LOC_E | LOC_U | O | |

**2 Morphemes to 1 Named Entity**

<ORGANIZATION>
··· | Roshia (*Russian*) | gun (*army*) | ···

<PERSON>
··· | Murayama (last name) | Tomiichi (first name) | shushou ( $\frac{prime}{minister}$ ) ···

**3 Morphemes to 1 Named Entity**

<TIME>
··· | gozen (*AM*) | ku (*nine*) | ji (*o'clock*) | ···

<ARTIFACT>
··· | ( $\frac{North}{America}$ ) hokubei | jiyuu-boueki (*free trade*) | kyoutei (*treaty*) | ···

Figure 1: Examples of Boundary Mismatch of Morphemes and Named Entities

that a sequence of morphemes is given as below:

$$( \frac{\text{Left}}{\text{Context}} ) \quad (\text{Named Entity}) \quad ( \frac{\text{Right}}{\text{Context}} )$$

$$\cdots M_{-k}^{L} \cdots M_{-1}^{L} \quad M_1^{NE} \cdots M_i^{NE} \cdots M_m^{NE} \quad M_1^{R} \cdots M_l^{R} \cdots$$
$$\uparrow$$
$$(\text{Current Position})$$

Then, given that the current position is at the morpheme $M_i^{NE}$, the task of named entity chunking is to assign a chunking state (to be described in Section 3.2) as well as a named entity type to the morpheme $M_i^{NE}$ at the current position, considering the patterns of surrounding morphemes. Note that in the supervised learning phase we can use the chunking information on which morphemes constitute a named entity, and which morphemes are in the left/right contexts of the named entity.

## 3.2 Encoding Schemes of Named Entity Chunking States

In this paper, we evaluate the following two schemes of encoding chunking states of named entities. Examples of these encoding schemes are shown in Table 3.

### 3.2.1 Inside/Outside Encoding

The Inside/Outside scheme of encoding chunking states of base noun phrases was studied in Ramshaw and Marcus (1995). This scheme distinguishes the following three states: O – the word at the current position is outside any base noun phrase. I – the word at the current position is inside some base noun phrase. B – the word at the current position marks the beginning of a base noun phrase that immediately follows another base noun phrase. We extend this scheme to named entity chunking by further distinguishing each of the states I and B into eight named entity types.[4] Thus, this scheme distinguishes $2 \times 8 + 1 = 17$ states.

### 3.2.2 Start/End Encoding

The Start/End scheme of encoding chunking states of named entities was employed in Sekine et al. (1998) and Borthwick (1999). This scheme distinguishes the following four states for each named entity type: S – the morpheme at the current position marks the beginning of a named entity consisting of more than one morpheme. C – the morpheme at the current position marks the middle of a named entity consisting of more than one morpheme. E – the morpheme at the current position marks the ending of a named entity consisting of more than one morpheme. U – the morpheme at the current position is a named entity consisting of only one morpheme. The scheme also considers one additional state for the position outside any named entity: O – the morpheme at the current position is outside any named entity. Thus, in our setting, this scheme distinguishes $4 \times 8 + 1 = 33$ states.

## 3.3 Preceding/Subsequent Morphemes as Contextual Clues

In this paper, we evaluate the following two models of considering preceding/subsequent

---

[4] We allow the state $x$_B for a named entity type $x$ only when the morpheme at the current position marks the beginning of a named entity of the type $x$ that immediately follows a named entity of the same type $x$.

morphemes as contextual clues to named entity chunking/tagging. Here we provide a basic outline of these models, and the details of how to incorporate them into the decision list learning framework will be described in Section 4.2.2.

### 3.3.1 3-gram Model

In this paper, we refer to the model used in Sekine et al. (1998) and Borthwick (1999) as a 3-gram model. Suppose that the current position is at the morpheme $M_0$, as illustrated below. Then, when assigning a chunking state as well as a named entity type to the morpheme $M_0$, the 3-gram model considers the preceding single morpheme $M_{-1}$ as well as the subsequent single morpheme $M_1$ as the contextual clue.

$$\left(\begin{array}{c} \text{Left} \\ \text{Context} \end{array}\right) \quad \left(\begin{array}{c} \text{Current} \\ \text{Position} \end{array}\right) \quad \left(\begin{array}{c} \text{Right} \\ \text{Context} \end{array}\right)$$
$$\cdots \quad M_{-1} \qquad M_0 \qquad M_1 \quad \cdots \qquad (1)$$

The major disadvantage of the 3-gram model is that in the training phase it does not take into account whether or not the preceding/subsequent morphemes constitute one named entity together with the morpheme at the current position.

### 3.3.2 Variable Length Model

In order to overcome this disadvantage of the 3-gram model, we propose a novel model, namely the "Variable Length Model", which incorporates richer contextual information as well as patterns of constituent morphemes within a named entity. In principle, as part of the training phase this model considers which of the preceding/subsequent morphemes constitute one named entity together with the morpheme at the current position. It also considers several morphemes in the left/right contexts of the named entity. Here we restrict this model to explicitly considering the cases of named entities of the length up to three morphemes and only implicitly considering those longer than three morphemes. We also restrict it to considering two morphemes in both left and right contexts of the named entity.

$$\left(\begin{array}{c} \text{Left} \\ \text{Context} \end{array}\right) \qquad (\text{Named Entity}) \qquad \left(\begin{array}{c} \text{Right} \\ \text{Context} \end{array}\right)$$
$$\cdots \quad M_{-2}^L M_{-1}^L \quad M_1^{NE} \cdots M_i^{NE} \cdots M_{m(\leq 3)}^{NE} \quad M_1^R M_2^R \quad \cdots$$
$$\uparrow \qquad\qquad\qquad (2)$$
$$(\text{Current Position})$$

## 4 Supervised Learning for Japanese Named Entity Recognition

This section describes how to apply the decision list learning method to chunking/tagging named entities.

### 4.1 Decision List Learning

A decision list (Rivest, 1987; Yarowsky, 1994) is a sorted list of decision rules, each of which decides the value of a *decision* $D$ given some *evidence* $E$. Each decision rule in a decision list is sorted in descending order with respect to some preference value, and rules with higher preference values are applied first when applying the decision list to some new test data.

First, the random variable $D$ representing a decision varies over several possible values, and the random variable $E$ representing some evidence varies over '1' and '0' (where '1' denotes the presence of the corresponding piece of evidence, '0' its absence). Then, given some training data in which the correct value of the decision $D$ is annotated to each instance, the conditional probabilities $P(D = x \mid E = 1)$ of observing the decision $D = x$ under the condition of the presence of the evidence $E$ ($E = 1$) are calculated and the decision list is constructed by the following procedure.

1. For each piece of evidence, we calculate the *log of likelihood ratio* of the largest conditional probability of the decision $D = x_1$ (given the presence of that piece of evidence) to the second largest conditional probability of the decision $D = x_2$:

$$\log_2 \frac{P(D = x_1 \mid E = 1)}{P(D = x_2 \mid E = 1)}$$

Then, a decision list is constructed with pieces of evidence sorted in descending order with respect to their log of likelihood ratios, where the decision of the rule at each line is $D = x_1$ with the largest conditional probability.[5]

---

[5]Yarowsky (1994) discusses several techniques for avoiding the problems which arise when an observed count is 0. From among those techniques, we employ the simplest one, i.e., adding a small constant $\alpha$ ($0.1 \leq \alpha \leq 0.25$) to the numerator and denominator. With this modification, more frequent evidence is preferred when several evidence candidates exist with the same

2. The final line of a decision list is defined as 'a default', where the log of likelihood ratio is calculated from the ratio of the largest marginal probability of the decision $D = x_1$ to the second largest marginal probability of the decision $D = x_2$:

$$\log_2 \frac{P(D = x_1)}{P(D = x_2)}$$

The 'default' decision of this final line is $D = x_1$ with the largest marginal probability.

## 4.2 Decision List Learning for Chunking/Tagging Named Entities

### 4.2.1 Decision

For each of the two schemes of encoding chunking states of named entities described in Section 3.2, as the possible values of the decision $D$, we consider exactly the same categories of chunking states as those described in Section 3.2.

### 4.2.2 Evidence

The evidence $E$ used in the decision list learning is a combination of the features of preceding/subsequent morphemes as well as the morpheme at the current position. The following describes how to form the evidence $E$ for both the 3-gram model and variable length model.

### 3-gram Model

The evidence $E$ represents a tuple $(F_{-1}, F_0, F_1)$, where $F_{-1}$ and $F_1$ denote the features of immediately preceding/subsequent morphemes $M_{-1}$ and $M_1$, respectively, $F_0$ the feature of the morpheme $M_0$ at the current position (see Formula (1) in Section 3.3.1). The definition of the possible values of those features $F_{-1}$, $F_0$, and $F_1$ are given below, where $M_i$ denotes the morpheme itself (i.e., including its lexical form as well as part-of-speech), $C_i$ the character type (i.e., Japanese (hiragana or katakana), Chinese (kanji), numbers, English alphabets, symbols, and all possible combinations of these) of $M_i$, $T_i$ the part-of-speech of $M_i$:

$$F_{-1} \quad ::= \quad M_{-1} \mid (C_{-1}, T_{-1}) \mid T_{-1} \mid \text{null}$$

---

unsmoothed conditional probability $P(D = x \mid E = 1)$. Yarowsky's training algorithm also differs somewhat in his use of the ratio $\frac{P(D = x_i \mid E = j)}{P(\neg D = x_i \mid E = j)}$, which is equivalent in the case of binary classifications, and also by the interpolation between the global probabilities (used here) and the residual probabilities further conditional on higher-ranked patterns failing to match in the list.

$$F_1 \quad ::= \quad M_1 \mid (C_1, T_1) \mid T_1 \mid \text{null}$$
$$F_0 \quad ::= \quad M_0 \mid (C_0, T_0) \mid T_0$$

As the evidence $E$, we consider each possible combination of the values of those three features.

### Variable Length Model

The evidence $E$ represents a tuple $(F_L, F_{NE}, F_R)$, where $F_L$ and $F_R$ denote the features of the morphemes $M_{-2}^L M_{-1}^L$ and $M_1^R M_2^R$ in the left/right contexts of the current named entity, respectively, $F_{NE}$ the features of the morphemes $M_1^{NE} \cdots M_i^{NE} \cdots M_{m(\leq 3)}^{NE}$ constituting the current named entity (see Formula (2) in Section 3.3.2). The definition of the possible values of those features $F_L$, $F_{NE}$, and $F_R$ are given below, where $F_j^{NE}$ denotes the feature of the $j$-th constituent morpheme $M_j^{NE}$ within the current named entity, and $M_i^{NE}$ is the morpheme at the current position:

$$
\begin{aligned}
F_L \quad &::= \quad M_{-2}^L M_{-1}^L \mid M_{-1}^L \mid \text{null} \\
F_R \quad &::= \quad M_1^R M_2^R \mid M_1^R \mid \text{null} \\
F_{NE} \quad &::= \quad F_i^{NE} F_{i+1}^{NE} F_{i+2}^{NE} \mid F_{i-1}^{NE} F_i^{NE} F_{i+1}^{NE} \\
&\quad \mid F_{i-2}^{NE} F_{i-1}^{NE} F_i^{NE} \mid F_i^{NE} F_{i+1}^{NE} \\
&\quad \mid F_{i-1}^{NE} F_i^{NE} \mid F_i^{NE} \quad\quad (3) \\
F_x^{NE} \quad &::= \quad M_x^{NE} \mid (C_x^{NE}, T_x^{NE}) \mid T_x^{NE}
\end{aligned}
$$

As the evidence $E$, we consider each possible combination of the values of those three features, except that the following three restrictions are applied.

1. In the cases where the current named entity consists of up to three morphemes, as the possible values of the feature $F_{NE}$ in the definition (3), we consider only those which are consistent with the requirement that each morpheme $M_j^{NE}$ is a constituent of the current named entity. For example, suppose that the current named entity consists of three morphemes, where the current position is at the middle of those constituent morphemes as below:

$$
\begin{array}{ccc}
\left(\begin{array}{c}\text{Left}\\\text{Context}\end{array}\right) & (\text{Named Entity}) & \left(\begin{array}{c}\text{Right}\\\text{Context}\end{array}\right) \\
\cdots \ M_{-2}^L M_{-1}^L & M_1^{NE} M_2^{NE} M_3^{NE} & M_1^R M_2^R \ \cdots \\
& \uparrow & (4) \\
& (\text{Current Position}) &
\end{array}
$$

Then, as the possible values of the feature $F_{NE}$, we consider only the following four:

$$F_{NE} \quad ::= \quad F_1^{NE} F_2^{NE} F_3^{NE} \mid F_2^{NE} F_3^{NE}$$

$$| F^{NE}_1 F^{NE}_2 | F^{NE}_2$$

2. In the cases where the current named entity consists of more than three morphemes, only the three constituent morphemes are regarded as within the current named entity and the rest are treated as if they were outside the named entity. For example, suppose that the current named entity consists of four morphemes as below:

$$\left( \begin{array}{c} \text{Left} \\ \text{Context} \end{array} \right) \qquad \text{(Named Entity)} \qquad \left( \begin{array}{c} \text{Right} \\ \text{Context} \end{array} \right)$$

$$\cdots \quad M^L_{-2} M^L_{-1} \quad M^{NE}_1 M^{NE}_2 M^{NE}_3 M^{NE}_4 \quad M^R_1 M^R_2 \quad \cdots$$
$$\uparrow$$
$$\text{(Current Position)}$$

In this case, the fourth constituent morpheme $M^{NE}_4$ is treated as if it were in the right context of the current named entity as below:

$$\left( \begin{array}{c} \text{Left} \\ \text{Context} \end{array} \right) \qquad \text{(Named Entity)} \qquad \left( \begin{array}{c} \text{Right} \\ \text{Context} \end{array} \right)$$

$$\cdots \quad M^L_{-2} M^L_{-1} \quad M^{NE}_1 M^{NE}_2 M^{NE}_3 \quad M^{NE}_4 M^R_1 \quad \cdots$$
$$\uparrow$$
$$\text{(Current Position)}$$

3. As the evidence $E$, among the possible combination of the values of three features $F_L$, $F_{NE}$, and $F_R$, we only accept those in which the positions of the morphemes are continuous, and reject those discontinuous combinations. For example, in the case of Formula (4) above, as the evidence $E$, we accept the combination $(M^L_{-1}, M^{NE}_1 M^{NE}_2, \text{null})$, while we reject $(M^L_{-1}, M^{NE}_2 M^{NE}_3, \text{null})$.

## 4.3 Procedures for Training and Testing

Next we will briefly describe the entire processes of learning the decision list for chunking/tagging named entities as well as applying it to chunking/tagging unseen named entities.

### 4.3.1 Training

In the training phase, at the positions where the corresponding morpheme is a constituent of a named entity, as described in Section 4.2, each *allowable* combination of features is considered as the evidence $E$. On the other hand, at the positions where the corresponding morpheme is outside any named entity, the way the combination of features is considered is different in the variable length model, in that the exception 1 in the previous section is no longer applied. Therefore, all the possible values of the feature $F_{NE}$ in Definition (3) are accepted. Finally, the frequency of each decision $D$ and evidence $E$ is counted and the decision list is learned as described in Section 4.1.

### 4.3.2 Testing

When applying the decision list to chunking/tagging unseen named entities, first, at each morpheme position, the combination of features is considered as in the case of the non-entity position in the training phase. Then, the decision list is consulted and all the decisions of the rules with a log of likelihood ratio above a certain threshold are recorded. Finally, as in the case of previous research (Sekine et al., 1998; Borthwick, 1999), the most appropriate sequence of the decisions that are consistent throughout the whole sequence is searched for. By consistency of the decisions, we mean requirements such as that the decision representing the beginning of some named entity type has to be followed by that representing the middle of the same entity type (in the case of Start/End encoding). Also, in our case, the appropriateness of the sequence of the decisions is measured by the sum of the log of likelihood ratios of the corresponding decision rules.

## 5 Experimental Evaluation

We experimentally evaluate the performance of the supervised learning for Japanese named entity recognition on the IREX workshop's training and test data. We compare the results of the combinations of the two encoding schemes of named entity chunking states (the Inside/Outside and the Start/End encoding schemes) and the two approaches to contextual feature design (the 3-gram and the Variable Length models). For each of those combinations, we search for an optimal threshold of the log of likelihood ratio in the decision list. The performance of each combination measured by F-measure ($\beta = 1$) is given in Table 4.

In this evaluation, we exclude the named entities with "other boundary mismatch" in Table 2. We also classify the system output according to the number of constituent morphemes of each named entity and evaluate the performance for each subset of the system output. For each sub-

Table 4: Evaluation Results Measured by F-measure ($\beta = 1$)

| | | $n$ Morphemes to 1 Named Entity | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | $n \geq 1$ | $n = 1$ | $n = 2$ | $n = 3$ | $n \geq 2$ | $n \geq 3$ | $n \geq 4$ |
| 3-gram | Inside/Outside | 72.9 | 75.9 | 79.7 | 51.4 | 69.4 | 42.5 | 29.2 |
| | Start/End | 72.7 | 76.6 | 79.6 | 43.7 | 68.1 | 37.8 | 29.6 |
| Variable Length | Inside/Outside | **74.3** | **77.6** | **80.0** | **55.5** | **70.9** | **49.9** | 41.0 |
| | Start/End | 72.1 | 77.0 | 75.6 | 51.5 | 67.2 | 48.6 | **43.6** |

set, we compare the performance of the four combinations of {3-gram, Variable Length} × {Inside/Outside, Start/End} and show the highest performance with bold-faced font.

Several remarkable points of these results of performance comparison can be stated as below:

- Among the four combinations, the Variable Length Model with Inside/Outside Encoding performs best in total ($n \geq 1$) as well as in the recognition of named entities consisting of more than one morpheme ($n = 2, 3, n \geq 2, 3$).
- In the recognition of named entities consisting of more than two morphemes ($n = 3, n \geq 3, 4$), the Variable Length Model performs significantly better than the 3-gram model. This result clearly supports the claim that our modeling of the Variable Length Model has an advantage in the recognition of long named entities.
- In general, the Inside/Outside encoding scheme performs slightly better than the Start/End encoding scheme, even though the former distinguishes considerably fewer states than the latter.

## 6 Conclusion

In this paper, we applied the supervised decision list learning method to Japanese named entity recognition, into which we incorporated several noun phrase chunking techniques and experimentally evaluated their performance. We showed that a novel technique that we proposed outperformed those using previously considered contextual features.

## 7 Acknowledgments

## References

A. Borthwick. 1999. A Japanese named entity recognizer constructed by a non-speaker of Japanese. In *Proc. of the IREX Workshop*, pages 187–193.

M. Collins and Y. Singer. 1999. Unsupervised models of named entity classification. In *Proc. of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 100–110.

S. Cucerzan and D. Yarowsky. 1999. Language independent named entity recognition combining morphological and contextual evidence. In *Proc. of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 90–99.

IREX Committee, editor. 1999. *Proceedings of the IREX Workshop*. (in Japanese).

S. Maiorano. 1996. The multilingual entity task (MET): Japanese results. In *Proc. of TIPSTER PROGRAM PHASE II*, pages 449–451.

MUC. 1998. *Proceedings of the 7th Message Understanding Conference (MUC-7)*.

L. Ramshaw and M. Marcus. 1995. Text chunking using transformation-based learning. In *Proc. of the 3rd Workshop on Very Large Corpora*, pages 83–94.

R. L. Rivest. 1987. Learning decision lists. *Machine Learning*, 2:229–246.

M. Sassano, Y. Saito, and K. Matsui. 1997. Japanese morphological analyzer for NLP applications. In *Proc. of the 3rd Annual Meeting of the Association for Natural Language Processing*, pages 441–444. (in Japanese).

S. Sekine, R. Grishman, and H. Shinnou. 1998. A decision tree method for finding and classifying names in Japanese texts. In *Proc. of the 6th Workshop on Very Large Corpora*, pages 148–152.

T. Utsuro and M. Sassano. 2000. Minimally supervised Japanese named entity recognition: Resources and evaluation. In *Proc. of the 2nd International Conference on Language Resources and Evaluation*, pages 1229–1236.

D. Yarowsky. 1994. Decision lists for lexical ambiguity resolution: Application to accent restoration in Spanish and French. In *Proc. of the 32nd Annual Meeting of ACL*, pages 88–95.