Vol. 40 No. 1 Transactions of Information Processing Society of Japan Jan. 1999

Regular Paper

Three-Word Dependency Relations and Their Application to Structural Ambiguity Resolution

EDUARDO DE PAIVA ALVES† and TEIJI FURUGORI†

Resolving syntactic ambiguities is an important issue in natural language processing. By measuring the strengths of association that hold among the words in co-occurrence relations, one may be able to determine the correct syntactic structure for an ambiguous construction. In this paper, we propose a syntactic disambiguation method that uses lexical preferences estimated in three-word dependency relations and, at the same time, conduct a disambiguation experiment in Japanese noun phrases containing the particle *no*. The result shows a better performance than those of other methods, and indicates the applicability of the method to resolving other syntactic ambiguities that appear in coordinated constructions, prepositional phrase attachments, and the like.

1. Introduction

Structural ambiguity is a major obstacle to building sound systems for natural language processing (NLP). It appears in constructions such as nominal compounds, coordinated structures and prepositional phrases:

(1) *hydrogen* ion exchange
(2) *cardiac* and vascular patients
(3) I bought flowers *with Jane.*

Often the correct syntactic structure is determined by the lexical preferences of the words involved. In the examples above, we know that *hydrogen* prefers *ion* to *exchange*, *cardiac* prefers *vascular* to *patient*, and *with Jane* prefers *bought* to *flowers.* In studying syntactic disambiguation, many researchers have used word co-occurrences in large corpora as an indicator that shows lexical preferences [10]. Weischedel, *et al.* [21], Resnik [19], Wu and Furugori [22], for instance, used the co-occurrences between a verb and the head word of prepositional phrase (PP) and between a noun and the head word of the same PP to determine whether the verb or noun was modified by the PP. Resnik [20] and Lauer [17] analyzed the structure of a noun compound by examining the co-occurrence relations among the nouns in it.

The method for resolving ambiguities that measures the preferences, or strengths of association, between two syntactic objects does not work well for some constructions, however. Consider the following examples:

(4) corn and *peanut butter*
(5) *put* the dress *on the rack*
(6) *Rangoon's* north *outskirts*

In (4), for instance, the strength of association for the two words *corn* and *peanut* may be greater than that of *corn* and *butter*, but the correct syntactic structure should be [*corn and* [*peanut butter*]]; in (5) it may be that the PP is attached to the noun *dress* when measuring the strengths of association between *dress* and *on rack* and *put* and *on rack*, but the PP should be attached to the verb *put* in this construction; and *Rangoon's* and *north* may have a stronger two word association than that of *Rangoon's* and *outskirts*, but *Rangoon's* modifies *outskirts* in (6). In all these examples, the strength of associations between two words is an improper measure for determining the correct syntactic structure of a phrase containing three or more words.

Instead of measuring the strength of association between two syntactic objects, in this paper we propose the use of co-occurrences among three syntactic objects involved in ambiguous constructions as an indicator of lexical preference. We then devise a way of measuring the strengths of association among the three words, and finally apply the method to a disambiguation experiment for Japanese noun phrases containing the particle 'の' (*no*).

2. Dependency Relations and Class-based Estimation Method

2.1 Three-Word Dependency Relation

Constituents have dependency relations, or modifier-modifee relations, with other constituents in phrases and sentences. An important task for a parser in NLP is to determine the

† Department of Computer Science, The University of Electro-Communications

correct dependency structures in a sentence.

In this paper we introduce a new kind of syntactic relations, three-word dependency relations (TWDR), as a tool for analysing dependency structures. A three-word dependency relation is a syntactic relation that holds among any three words in a sentence. It may be defined as:

<three-word dependency relation> ::=
[W[<two word dependency relation>]] |
[[<two-word dependency relation>]W] |
[[W] <two word dependency relation>] |
[<two-word dependency relation>[W]]
<two-word dependency relation> ::= [W[W]] |
[[W]W]

Here W stands for any head word in a constituent, and we understand that for any constituent d_i, $[d_1[d_2]]$ means that d_1 modifies d_2 and $[[d_1]d_2]$ means that d_2 modifies d_1.

It is possible to recursively decompose any complex dependency structure into a set of three-word dependency relations. For example, the possible dependency structures for the sentence *Time flies like an arrow* (there are four head words in it: Time, flies, like, arrow) can be expressed as the following TWDRs:

$[Time[[[flies]][like[arrow]]]] =$
$[Time[[flies]like]] + [[[flies]][[like]arrow]]$
$[[Time[flies]][[[like]arrow]]] =$
$[[Time[flies]][[like]]]] + [[flies][[[like]arrow]]]]$

A word may modify another word to its right or to its left in a construction. In a language like Japanese, however, a word always modifies another word to its right. Thus, we have two possible dependency relations for a three-word dependency construction: $[w_1[w_2[w_3]]]$ and $[[w_1[w_2]][[[w_3]]]]$, where w_i stands for a head word in a constituent. In this paper, we conveniently use the $[w_1[w_2w_3]]$ for $[w_1[w_2[w_3]]]$, and $[[w_1w_2]w_3]$ for $[[w_1[w_2]][[[w_3]]]]$. The two expressions are equivalent to:

2.2 Estimation of the Strength of Lexical Association

Mutual information (MI) [6] is a standard way of estimating the strength of lexical association between any two words appearing in a text. The MI between the two words w_1 and w_2 is defined as:

$$I(w_1, w_2) = \log_2 \frac{N * f(w_1, w_2)}{f(w_1)f(w_2)} \quad (1)$$

where N is the size of the corpus used in the estimation, $f(w_1, w_2)$ is the frequency of co-occurrence, and $f(w_1)$ and $f(w_2)$ are the frequencies of the individual words.

We extend the definition of MI for our purpose, which is to determine the dependency structure for three words. Equations (2) and (3) show the ways of obtaining mutual information between two occurrences of $\langle w_1 \rangle$ and $\langle w_2, w_3 \rangle$ for $[[w_1w_2]w_3]$ and $[w_1[w_2w_3]]$.

$$I([[w_1w_2]w_3]) = \log_2 \left(\frac{N * f([[w_1w_2]w_3])}{f(w_1)f([w_2w_3])} \right), \quad (2)$$

$$I([w_1[w_2w_3]]) = \log_2 \left(\frac{N * f([w_1[w_2w_3]])}{f(w_1)f([w_2w_3])} \right), \quad (3)$$

where f is the frequency of co-occurrence and N the size of the training corpus.

The attachment of w_1 can be determined by comparing mutual information between two occurrences of $\langle w_1 \rangle$ and $\langle w_2, w_3 \rangle$ for $[[w_1w_2]w_3]$ and $[w_1[w_2w_3]]$.

A difficulty encountered in the calculation of the mutual information is the sparse data problem. How do we obtain reliable statistical results when no or only a few word co-occurrences are observed? For many researchers [2),4),13),14),23)] an answer to the question is to use word classes that contain the words in question and calculate the mutual information on the basis of the word class co-occurrences.

To deal with the sparse data problem, Eqs. (2) and (3) may be replaced by Eqs. (4) and (5):

$$I([[w_1w_2]w_3]) \approx \log_2 \left(\frac{N * f([[C_1C_2]C_3])}{f(C_1)f([C_2C_3])} \right), \quad (4)$$

$$I([w_1[w_2w_3]]) \approx \log_2 \left(\frac{N * f([C_1[C_2C_3]])}{f(C_1)f([C_2C_3])} \right). \quad (5)$$

Here C_i stands for a word class that includes the word w_i.

This way of estimating mutual information with word classes risks the problem of overgeneralization, that is, we may use a word class that is too general for the word in question. Another problem is how to choose the best or most appropriate class for the word concerned.

The solution we offer for this problem is to choose word classes from a taxonomy by using t-scores☆ as a measure of reliability[1].

For a class co-occurrence $\langle C_1; C_2C_3\rangle$, the t-scores for $[[C_1C_2]C_3]$ and $[C_1[C_2C_3]]$ may be approximated by:

$$t([[C_1C_2]C_3]) \approx \frac{f([[C_1C_2]C_3]) - \frac{1}{N}f(C_1)f(C_2,C_3)}{\sqrt{f([[C_1C_2]C_3])}} \tag{6}$$

$$t([C_1[C_2C_3]]) \approx \frac{f([C_1[C_2C_3]]) - \frac{1}{N}f(C_1)f(C_2,C_3)}{\sqrt{f([C_1[C_2C_3]])}} \tag{7}$$

t becomes very small or negative when $f([[C_1C_2]C_3])$ or $f([C_1[C_2C_3]])$ is zero or low, and becomes bigger as the frequency becomes higher. t also becomes low when the classes in the co-occurrence contain too many words irrelevant to the estimation.

Class-based estimation of mutual information using t-scores can be carried out in the following way:

a. Set a threshold for t.
b. Search the lowest class co-occurrence in the taxonomy for which t is above the threshold.
c. Choose the most probable dependency structure using Eqs. (4) and (5).

The process is expressed in algorithmic form as follows:

Algorithm: Structural ambiguity resolution using class-based estimation

step 1: set $C_1 = w_1$, $C_2 = w_2$, $C_3 = w_3$
step 2: while $t([[C_1, C_2]C_3]) < threshold$
  else if $f(C_1) < f(C_2, C_3)$
    replace C_1 with its upper class
  else if $f(C_2) < (C_3)$
    replace C_2 with its upper class
  else if $C_3 \neq \mathbf{C}$
    replace C_3 with its upper class
  else $I_{left} = na$ (not available)
    goto step 4
step 3: $I_{left} = I([[C_1, C_2]C_3])$
step 4: set $C_1 = w_1$, $C_2 = w_2$,$C_3 = w_3$
step 5: while $t([C_1[C_2, C_3]]) < threshold$
  else if $f(C_1) < f(C_2, C_3)$
    replace C_1 with its upper class
  else if $f(C_2) < (C_3)$
    replace C_2 with its upper class
  else if $C_3 \neq \mathbf{C}$
    replace C_3 with its upper class
  else $I_{right} = na$ (not available)
    goto step 7
step 6: $I_{right} = I([C_1[C_2C_3]])$
step 7: if $I_{left} > I_{right}$
    attach w_1 to w_2
  else attach w_1 to w_3

Here, $\mathbf{C}$ is the topmost word class in the taxonomy.

3. Application to Japanese Noun Phrases

The method of estimating the strengths of word association may be effectively used for resolving syntactic ambiguities. We would like to show an example from a disambiguation study that deals with the construction of Japanese noun phrases.

3.1 Japanese Phrases with the Particle *no*

The function of the particle *no* (の) in Japanese is in a sense similar to that of the preposition *of* in English. It builds up a noun phrase: *Tokyo no kita* or 東京の北 (North of Tokyo). By combining various parts of speech, an arbitrary number of instances of *no* is used to make compound noun phrases: *watashi no ie no niwa* or 私の家の庭 (garden of my house).

A noun phrase can be built up recursively, using any number of instances of *no*:

< *no_construction* > ::= N no N
| N no < *no_construction* >

Here, N stands for a nominal phrase.

In $[w_1$ *no* $w_2]$, w_1 modifies w_2. But when three or more than three N's are involved, a noun phrase containing instances of *no* becomes structurally ambiguous. In the example *Rangoon no kita no hazure* or ラングーンの北のはずれ, (Rangoon's north outskirts), two syntactic structures are possible:

*[[ラングーンの北] のはずれ]
([[Rangoon's north] outskirts])
[ラングーンの [北のはずれ]]
([Rangoon's [north outskirts]])

Theoretically a noun phrase containing instances of *no* can be infinite in its length. But the maximum number of *no* in a noun phrase

☆ The t-score[5] is a standard measure of the likelihood that an occurrence can be attributed to chance.

Table 1 Frequencies and t-score.

Co-occurrence			$f(C_1)$	$f(C_2)$	$f(C_3)$	$f(C_2C_3)$	$f([C_1[C_2C_3]])$	t
region	*direction*	*location*	181	42	513	10	1	.84
urban area	*cardinal points*	*location*	28	33	513	10	0	0
urban area	*cardinal points*	*extreme*	28	33	10	10	0	0
urban area	north	*extreme*	28	9	10	6	0	0
urban area	north	outskirt	28	9	2	0	0	0
Rangoon	north	outskirt	0	9	2	0	0	0

Table 2 *no*-constructions in the EDR corpus.

no's	Noun Phrase	Frequency
1	共通の部分 (shared part)	207,802
2	米国の制裁の動き (America's sanction movement)	21,415
3	私のかつてのアメリカの生活 (my life once in America)	1,321
4	ラングーンの北のはずれの緑の森 (the green forest in Rangoon's north outskirts)	76

with *no*'s is 4 in the EDR Japanese Corpus [12)☆]. **Table 1** shows the numbers of *no*'s used in noun phrases.

3.2 An Example of Ambiguity Resolution

Let us take an example and try to find the correct structure for three word *no*-constructions that can be either $[[w_1\ no\ w_2]\ no\ w_3]$ or $[w_1\ no\ [w_2\ no\ w_3]]$. Consider:

ラングーンの北のはずれ
(Rangoon's north outskirts)

We first calculate the strength of association for [ラングーンの [北のはずれ]]. To do so, we replace the words in co-occurrence with their upper classes according to the algorithm described in Section 2☆☆ and obtain a co-occurrence with a t-score just above the threshold (we set the threshold value empirically at 0.70).

Table 2 shows the conceptual hierarchy of *Rangoon*, *north*, and *outskirt*, and various frequencies of the concepts involved. Here, the numbers in columns $f(C_1)$, $f(C_2)$, and $f(C_3)$ are the occurrences in the EDR corpus for the classes in the conceptual hierarchy of the words *Rangoon*, *north* and *outskirts*, respectively.

In the case for [ラングーンの [北のはずれ]], we use the classes (**region** *no* **direction** *no* **location**) to estimate the mutual information. We repeat the same process for [[ラングーンの北] のはずれ] ([[Rangoon's north] outskirts]). The mutual information we obtain through this process is 2.58 for [ラングーンの [北のはずれ]] and 2.19 for [[ラングーンの北] のはずれ]. We thus select the structure [ラングーンの [北のはずれ]] as the most probable one, and this in fact accords with the syntactic annotation in the corpus.

Table 3 Experimental results.

Method	Success
(1) Closest attachment	73.2% (314/429)
(2) Two-word dependency	72.7% (312/429)
(3) **Three-word dependency**	77.6% (333/429)
(4) Human	80.0% (120/150)

4. Results and Evaluation

We tested 429 *no*-constructions in the EDR Japanese Corpus, each of which contains two instances of *no*. This number is somewhat arbitrary, but all of the constructions have semantic annotations. The training data, separate from the test data, are semantically annotated for 11,224 *no*-constructions obtained from the 21,415 *no*-constructions in the EDR Japanese Corpus.

Table 3 shows the disambiguation results (success rate) for the test data. It contains the results from various other methods with the same training data and test data. Here, the first result (1) is the one obtained by attaching the modifier to the nearest modifee (head) [15)]. The second one (2) shows the performance by estimating the strength of association in the two-word dependency relations

☆ The EDR Japanese Corpus is provided by the Japan Electronic Dictionary Research Institute. It contains 220,000 parsed sentences with syntactic and semantic annotations.

☆☆ For this purpose, we use the Conceptual Dictionary provided by the EDR Institute. It contains a hierarchical structure of all the words in the EDR corpus. The number of concepts in Conceptual Dictionary is about 400,000.

Table 4 Some examples from test constructions.

Construction	$I([w_1 w_2])$	$I([w_1 w_3])$	$I([[w_1 w_2] w_3])$	$I([w_1 [w_2 w_3]])$
[ラングーンのはずれ] の森 Rangoon - outskirt - forest	0.44	0.85	**1.43**	na
[ニュートンの住宅地] の自宅 Newton - residencial area - home	0.91	3.91	**0.21**	na
はずれの [緑の森] outskirts - green - forest	3.15	0.08	0.35	**0.51**
[被乗数のビット数分] の部分積 multiplicand - n-bit - partial product	1.80	5.32	**3.52**	2.38
ラングーンの [北のはずれ] Rangoon - north - outskirts	5.35	2.38	2.19	**2.58**

$[w_1 w_2]$ and $[w_1 w_3]$ [2),16),17)]☆. The third one (3) shows the performance of our method. The last one (4) is the performance of two native Japanese speakers to whom we presented the *no*-constructions without surrounding contexts (in this case, we tested 150 of the 429 constructions because it was too laborious to go through all of them).

The lower and upper bounds for the performance of our method seem to be 73.2% for the simple heuristic of closest attachment (1) and 80.0% for human beings (4). In the first case, a word is always attached to its successor in a phrase. Nevertheless, this method performed better than that of using two-word dependency relation (2). This result may be language-specific because Japanese has a strong tendency for word to modify the adjacent word to its right. However, the performance of our method (3) is much better than those of (1) and (2). We attribute the better performance to the estimation procedure, which seems to capture the syntactic and semantic constraints better than other methods, as we have taken a "wider context" in the disambiguation process. Using three-word dependency relations, we can capture more syntactic information that is not included in two-word dependency relations. For example, the distance between two words or constituents has a strong influence on the dependency relations but it is not considered in the work based on two word dependency relations. Our method, on the other hand, can capture this information and then differentiate the most probable structure from the candidates.

Table 4 shows a few examples in which attachment by two-word dependency relations failed and our method succeeded☆☆. In the first example, *Rangoon* is taken to modify the third constituent, *forest*, when we use the two-word dependency relation, since $I([w_1 w_3])$ is bigger than $I([w_1 w_2])$. In our method, however *Rangoon* modifies *outskirt*, since $I([[w_1 w_2] w_3])$ is bigger than $I([w_1 [w_2 w_3]])$. In Table 4, the mutual information in bold font indicates the properly selected syntactic structure for the construction in our method.

Our method takes its toll in some cases, however. Sometimes, the annotations in the corpus are doubtful. In *Konnichi no kankoku no fuhai* or 今日の韓国の腐敗 (today's Korean corruptions), for instance, either [today's [Korean corruption]] or [[today's Korean] corruption] seems semantically acceptable. This fact may explain why the result obtained by human beings is worse than we expected.

Other incorrect choices were due mainly to the data sparseness. In this case, our method finds no class co-occurrence with the t-score above the threshold. Estimates are then made for all co-occurrences, and the algorithm chooses the most probable structure, which in effect becomes that of right attachment.

Although it is not a methodological deficiency, our method apparently suffers from the size of the training data (11,224): we found that about 37% of the test data suffer from data sparseness. When we eliminated these data and tested our method, we obtained a success rate

☆ We use mutual information between the classes of the two words to measure the strength of association between them,

$$I([w_1 w_2]) = \log_2 \left(\frac{N * f(C_1, C_2)}{f(C_1) f(C_2)} \right), \qquad (8)$$

where C_1 and C_2 are word classes that correspond to the words w_1 and w_2, respectively. The classes were selected from the EDR Concept Dictionary according to their t-score, using Algorithm.

☆☆ In Table 4, *na* means *not available.*

of 79.3%. This tells us that using a larger set of training data helps improve the overall success rate.

5. Discussion

5.1 Applications to Other Kinds of Structural Disambiguation Tasks

The method presented in this paper can be used for resolving various kinds of structural ambiguities, such as complicated noun phrases, coordinate structures, and prepositional phrase attachments.

Let us take an example and show how to disambiguate the sentence *Put the dress on the rack.* It has two possible dependency structures [Put [dress][on rack]] and [Put [dress [on rack]]]☆ in which the prepositional phrase *on (the) rack* is attached to the verb *Put* and the noun *dress*, respectively. We decompose these structures into TWDRs:

[Put [dress][on rack]] =
[[put dress] on] + [put [on rack]]
[Put [dress [on rack]]] =
[put [dress on]] + [dress [on rack]]

We calculate the class-based probability for each TWDR by using the classes selected by the Algorithm. We then estimate the probability for each structure by multiplying the probabilities of its TWDRs [7),8)] and select the structure that has the highest probability. In the example, since both [put [on rack]] and [dress [on rack]] are possible, the probabilities will reflect the fact that [[put dress] on] is more likely than [put [dress on]], since the verb *put* requires both a direct object and a locational object. The comparison leads us to identify the first structure as the most probable interpretation.

Let us take another example in disambiguating a Japanese noun phrase with an adjective in it, 日本の美しい川の流れ. It contains three possible interpretations: [日本の [美しい [川の流れ]]] (*the beautiful flow of river(s) of Japan*), [日本の [[美しい川] の流れ]] (*the flow of beautiful river(s) of Japan*), and [[日本の [美しい川]] の流れ] (*the flow of Japanese beautiful river(s)*)☆☆. To deal with this problem, we decompose the structures into TWDRs of [日本の [美しい川]], [日本の [川の流れ]], [[日本の川] の流れ], [[美しい川] の流れ], [美しい [川の流れ]] using the method in Section 2. We compute the probability for each dependency structure and select the one with the highest probability as the most probable interpretation of 日本の美しい川の流れ.

We have no difficulty in extending our method to noun phrases with more than two instances of *no* (e.g., 私の父の昔の友人の写真 (the photograph of my father's former friend)). What we need is to decompose such a construction into a series of TWDRs. However, it is noted here that we may need to incorporate a distance measure into our method, since such a measure is effective for determining the structure of compound nouns. Kobayashi *et al.* [16)] proposed a method for analyzing structure of Japanese compound nouns composed of kanji (Chinese) characters, using collocational information from a corpus and semantic information from a thesaurus, and strengthened their method by introducing a distance measure to improve the performance.

5.2 Conclusion

We have presented a structural disambiguation method using lexical preferences estimated from three-word dependency relations. We proved experimentally that the method works well for determining the correct syntactic structure in ambiguous constructions.

There have been some proposals using three-word co-occurrences in the literature [18),19)]. However, none of them have used dependency relations in three-word co-occurrences. We are the first to use three-word dependency relations in estimating the strengths of association and to apply it to structural disambiguation tasks.

The *t*-score seems to be effective for selecting reliable classes from a taxonomical hierarchy. To our knowledge, the *t*-score, a standard measure of the likelihood that an occurrence can be attributed to chance, has not been used before as a measure of reliability in selecting class co-occurrences.

☆ Here, we ignore the determiner *the* in the sentence and simplify the bracketing notations.

☆☆ In Japanese, the attachments or dependency relations should not be crossed.

References

1) Alves, E., Wu, H. and Furugori, T.: A Method for Estimating Strength of Association and Its Application to Structural Disambiguation, *Journal of Natural Language Processing*, Vol.5, No.3 (1998).

2) Alves, E.: The Selection of the Most Probable Dependency Structure in Japanese Using Mutual Information, *Proc. 34th Annual Meeting of the Association for Computational Linguistics*,

pp.372–374 (1996).

3) Basili, R., Pazienza, M. and Velardi, P.: Combining NLP and Statistical Techniques for Lexical Acquisition, *AAAI Fall Symposium on Probabilistic Approaches to Natural Language*, pp.1–9 (1991).

4) Brown, P., Della Pietra, V. and Mercer, R.: Word Sense Disambiguation Using Statistical Methods, *Proc. 30th Annual Meeting of the Association for Computational Linguistics*, pp.264–270 (1992).

5) Church, K. and Mercer, R.: Introduction to the Special Issue on Computational Linguistics Using Large Corpora, *Computational Linguistics*, Vol.19, No.1, pp.1–24 (1993).

6) Church, K. and Hanks, P.: Word Association Norms, Mutual Information and Lexicography, *Computational Linguistics*, Vol.16, No.1, pp.22–29 (1990).

7) Collins, M.: A New Statistical Parser Based on Bigram Lexical Dependencies, *Proc. 34th Annual Meeting of the Association for Computational Linguistics*, pp.184–191 (1996).

8) Eisner, J.: Three New Probabilistic Models for Dependency Parsing: An Exploration, *Proc. 16th International Conference on Computational Linguistics*, pp.340–345 (1996).

9) Hindle, D.: Noun Classification from Predicate-Argument Structures, *Proc. 28th Annual Meeting of the Association for Computational Linguistics*, pp.268–275 (1990).

10) Hindle, D. and Rooth, M.: Structural Ambiguity and Lexical Relations, *Computational Linguistics*, Vol.19, No.1, pp.103–120 (1993).

11) Grishman, R. and Sterling, J.: Acquisition of Selectional Patterns, *Proc. 14th International Conference on Computational Linguistics*, pp.658–664 (1992).

12) Japan Electronic Dictionary Research Institute: *EDR Electronic Dictionary Specifications Guide* (1993) (in Japanese).

13) Jelinek, F. and Mercer, R.: Probability Distribution Estimation from Sparse Data, *IBM Technical Disclosure Bulletin*, Vol.28, pp.2591–2594 (1985).

14) Katz, S.: Estimation of Probabilities from Sparse Data for Language Model Component of a Speech Recognizer, *IEEE Trans. Acoustics, Speech and Signal Processing*, Vol.ASSP-35, No.3, pp.400–401 (1987).

15) Kimball, J.: Seven Principles of Surface Structure Parsing in Natural Language, *Cognition*, Vol.2, pp.15–47 (1973).

16) Kobayasi, Y., Tokunaga, T. and Tanaka, H.: Analysis of Japanese Compound Noun Using Collocational Information, *Journal of Natural Language Processing*, Vol.3, No.1, pp.29–44 (1996) (in Japanese).

17) Lauer, M.: Conceptual Association for Compound Noun Analysis, *Proc. Annual Meeting of the Association for Computational Linguistics*, pp.47–54 (1993).

18) Li, H. and Abe, N.: Generalizing Case Frames Using a Thesaurus and the MDL Principle, *Proc. Recent Advances in Natural Language Processing*, pp.239–248 (1995).

19) Resnik, P.: Wordnet and Distributional Analysis: A Class-Based Approach to Lexical Discovery, *AAAI Workshop on Statistically-Based Natural Language Processing Techniques*, pp.56–64 (1992).

20) Resnik, P.: Selection and Information: A Class-Based Approach to Lexical Relationships, Ph.D. Thesis, University of Pennsylvania (1993).

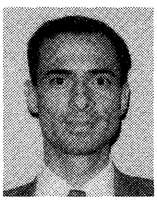
21) Weischedel, R., Meteer, M., Schwartz, R., Ramshaw, L. and Palmucci, J.: Coping with Ambiguity and Unknown Words through Probabilistic Models, *Computational Linguistics*, Vol.19, No.2, pp.359–382 (1993).

22) Wu, H. and Furugori, T.: A Hybrid Disambiguation Model for Prepositional Phrase Attachment, *Literary and Linguistic Computing*, Vol.11, No.4, pp.187–192 (1997).

23) Yarowsky, D.: Word Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora, *Proc. 14th International Conference on Computational Linguistics*, pp.454–460 (1992).

24) Zhang, Y. and Ozeki K.: Dependency Analysis of Japanese Sentences Using the Statistical Property of Dependency Distance between Phrases, *Journal of Natural Language Processing*, Vol.4, No.2, pp.3–19 (1997) (in Japanese).

(Received August 13, 1997)
(Accepted October 21, 1998)



Eduardo de Paiva Alves received a B.S. degree in Engineering from Rio de Janeiro Federal University, a B.A. degree in English Literature from Rio de Janeiro State University, and a M.S. degree in Computer Science also from Rio de Janeiro Federal University, Brazil. He is currently a Ph.D. candidate in Computer Science at the University of Electro-Communications, Tokyo, Japan. He is a member of the Association for Computational Linguistics.

Teiji Furugori is a professor in the Department of Computer Science at the University of Electro-Communications, Tokyo, Japan. He has a Ph.D. in Computer Science from SUNY at Buffalo. His research interests include Computational Linguistics, Natural Language Processing, Computer Aided Instruction, and Cognitive Science. He is a member of ACM, the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers of Japan, the Association for Natural Language Processing of Japan, and the Mathematical Linguistic Society of Japan.