Symmetric Pattern Matching Analysis for English Coordinate Structures

Akitoshi Okumura and Kazunori Muraki
NEC Corp. Information Technology Research Laboratories
4-1-1 Miyazaki, Miyamae-ku, Kawasaki, Kanagawa 216, JAPAN
okumura@hum.cl.nec.co.jp

Abstract

The authors propose a model for analyzing English sentences including coordinate conjunctions such as "and","or","but" and the equivalent words. Syntactic analysis of the English coordinate sentences is one of the most difficult problems for machine translation (MT) systems. The problem is selecting, from all possible candidates, the correct syntactic structure formed by an individual coordinate conjunction, i.e. determining which constituents are coordinated by the conjunction. Typically, so many possible structures are produced that MT systems cannot select the correct one, even if the grammars allow to write the rules in the simple notations. This paper presents an English coordinate structure analysis model, which provides top-down scope information of the correct syntactic structure by taking advantage of the symmetric patterns of the parallelism. The model is based on a balance matching operation for two lists of the feature sets, which provides four effects: the reduction of analysis cost, the improvement of word disambiguation, the interpretation of ellipses, and robust analysis. This model was practically implemented and incorporated into the English-Japanese MT system, and provided about 75% accuracy in the practical translation use.

1 Introduction

This paper presents a model for analyzing English sentences including coordinate conjunctions such as "and","or","but" and equivalent words.

Syntactic analysis of English coordinate sentences is one of the most difficult problems for machine translation (MT) systems. The problem is selecting, from all possible candidates, the correct syntactic structure formed by an individual coordinate conjunction, i.e., determining which constituents are conjoined by the conjunction. Although the conjunction appears to have a simple function in the English language, it has been researched as a conjunct scope problem by both theoretical and computational linguists. Theoretically, it is possible to describe the syntactic and semantic constraints that govern the acceptability of a structure in which two constituents are conjoined by the conjunction(Lesmo and Torasso, 1985; Gazdar, 1981; Schachter, 1977). Computationally, it is possible to describe the grammar and heuristic rules for these constraints by ATN networks, logic grammars, HPSG, and categorical grammars(Kosy, 1986; Fong and Berwick, 1985; Huang, 1984; Boguraev et al., 1983; Blackwell, 1981; Niimi et al., 1986). However, it is not easy to apply these techniques to large-scale MT systems, because there exist a variety of conjoined patterns, many word ambiguities, some unknown words and ellipses of the words simultaneously, in real environments. Also, there may be several conjunctions and the equivalent words, such as commas, in a single sentence. Typically, the methods produce so many possible structures that MT systems cannot select the correct one, even if the grammars allow to write the rules in the simple notations.

Often, conjunctions might produce the reading difficulty even for the human readers. However, they also give the readers a kind of symmetry as a reading indication. They exhibit the tendency to conjoin the same kind of syntactic patterns, which has been called "parallelism"(Beaugrande and Dressler, 1981; Shinoda, 1981). In Japanese, the similarity is used for analyzing conjunctive structures and the method is found effective(Kurohashi and Nagao, 1992). While Japanese language has several coordinate conjunctions according to the syntactic levels (a noun phrase and a predicative clause), English coordinate conjunctions are used for any level of the structures. More robust methods are necessary for dealing with English conjunctive structures.

We propose here an English coordinate structure analysis model, which can determine the correct syntactic structure in real environments by taking advantage of the symmetric patterns of the parallelism.

The model is based on a balance matching operation for two lists of the feature sets. The lists represent the left-side and right-side structures of the coordinate conjunctions, including word ambiguities. The operation determines the most symmetric structure by comparing both sides of the conjunction.

First, the problems of the English coordinate sentences are explained. Second, it is mentioned that the parallelism can be effective information for top-down analysis of the sentences. Third, the balance matching analysis model is presented for solving the problems. This model, which was implemented in the PIVOT English-Japanese MT system with the dictionary of 100,000 words, has been working in the analysis module. Finally, the results in the MT system are reported together with the MT system configuration.

2 Problems of the conjunctions

Coordinate conjunctions for MT systems present three difficulties(Kosy, 1986; Huang, 1983; Niimi et al., 1986; Okumura et al., 1987).

1. Analysis cost: English coordinate conjunctions have a variety of linguistic functions. The conjunctions can syntactically conjoin any parts of speech; nouns, adjectives, verbs, etc., and all sorts of constituents; words, phrases and clauses. They produce so many possible structures that MT systems cannot select the correct one, even if some grammars provide the simple notations of the rules. The complexity of the rules impose a burden on the analysis process.

2. Ambiguities of the words: Most English words are ambiguous in their parts of speech ,such as Noun and Verb, Preposition and Conjunction, Auxiliary Verb and others, and Adjective and Noun. The complex rules governing the conjunctions make the word disambiguation more complicated, which results in the reduction of analysis precision.

3. Ellipses of the words: It is possible for a conjunct to contain "gaps" (elided elements) which are not allowed if the conjunction is removed. These gaps must be filled with elements from other conjunct for a proper interpretation, as in (1) and (1)'.

(1) NO.2 landing gear selector valves should be closed to the full position,and *NO.3 to the half.*

(1)' NO.2 landing gear selector valves should be closed to the full position, and *NO.3 [landing gear selector valves should be closed] to the half [position].*

3 Parallelism of the conjunctions

3.1 Symmetric patterns of parallelism

English coordinate conjunctions have a tendency to conjoin the same kinds of syntactic patterns. We identify three levels of symmetric pattern.

- Phrase(Clause) symmetric patterns: Phrase(Clause)-level symmetric patters such as [Verb-Phase AND Verb-Phase] in the conjunct scope of "as well as " in (2) and a series of commas in (3).

(2) Such coupling is desirable because it enables a development engineer *to move easily within this hierarchy* **as well as** *to exploit the distinctive features of each system.*

(3) The add operators cause POSTSCRIPT to pick up the top two numbers from the stack, remove them, *add them,* and *leave the sum on the stack.*

- Word symmetric patterns: Word-level symmetric patterns such as [Quantifier Preposition Abstract-Noun AND Quantifier Preposition Abstract-Noun] in the conjunct scope of the "and" in (4). Some patterns are represented by the semantic features such as [Instrument AND Instrument] about "and_1" of (5).

(4) The container need not be large; if it is *10cm in diameter* **and** *12cm in depth,* that is enough.

(5) Inspect the cockpit *indicators* and_1 *levers* for cracked glass and missing control knobs.

- Morphological symmetric patterns: Morphological symmetric patterns are recognized by the sorts of characters, uppercase or lowercase letters, as in (6) and (7) as well as an exactly same morphological pattern [CIC ... hatches AND CIC ..., hatches] in (8).

(6) An atomic bomb is a device for producing an explosively rapid neutron chain reaction in *uranium-235* **or** *plutonium-239* which is called a fissile material.

(7) Technical orders described in *AFR 8-2* **and** *PFR 7-2* are registered in the on-line file in the form of inspection workcards.

(8) There are CIC_1 $ditching_2$ $hatches_3$ and CIC_4 $escape_5$ $hatches_6$ in the compartment.

Some symmetric patterns may appear in combined form [Preposition Gerund Nominal-Phrase AND Preposition Gerund Nominal-Phrase] in (9).

(9) Radioisotopes have played an important part in_1 developing effective insecticides in_2 the country **and** in_3 finding the best ways of applying them.

3.2 Analysis by the symmetric patterns

The symmetric patterns can be effective information for top-down analysis of the conjunct scope. For example, though (9) allows another counterpart ("in_2 the country") as the conjoined phrase ("in_3 finding ..."), the symmetric pattern information makes it easy to to select the correct counterpart of the phrase ("in_1 developing ...."). Also, where other examples syntactically allow other counterparts of the conjoined noun phrases, the symmetric pattern information enables easy selection. Often, the scope of each conjunct is explicitly demarcated with commas and morphological patterns, as in (3) and (8). The symmetric patterns can be also effective for word disambiguation. For example, (8) contains verb/noun ambiguities for $escape_5$ and $hatches_6$. The symmetric patterning of $\underline{ditching_2\ hatches_3}$ facilitates their disambiguation.

In the above example sentences, the words immediately following the conjunction play important roles for detecting the structures, because there is usually strong similarity between the starting words of each conjunct scope and the words following the conjunctions. However, the following examples also contain kinds of symmetric patterns, though the words following the conjunctions don't have similarity with the starting words of the conjunct scopes.

(10) In 1985 the government offered offshore registration to the companies, and, <u>in consequence,</u> in 1985 incorporation fees generated about two million dollars.

(11) The damage of the landing gear selector valve caused the leakage of the hydraulic fluid, and <u>completely</u> blockaded the return path.

(12) Close the cockpit ditching hatches, and the cabin pressure will be dumped to relieve the <u>air loads on the hatches.</u>

In both (10) and (11), an adverbial modifier is inserted at the start of the second conjunct; this is a common pattern extension. In (12), there is no real parallelism: the first conjunct clause is an imperative, and the second its result.

4 Balance matching analysis model

The balance matching analysis model determines the correct structure by taking advantage of the symmetric patterns. In this section, first, the representation of symmetric patterns is presented. Then the balance matching operation is presented. Finally, analysis by the balance matching is described.

4.1 The pattern representation

The symmetric patterns are represented by a list of three feature sets; Phrase features, Word features, and Morphological features, based on the symmetric pattern levels.

- Phrase feature $[\phi]$: Values are Predicative, Nominal, Nominal-Premodifier, Nominal-Postmodifier, Predicate-modifier. These values are assigned to all the constituents in the phrase. For example, all the words in "the effective insecticides" have $\phi.(Nominal)$

- Word feature $[\gamma]$: This feature includes 120 values which subclasses of general parts of speech, according to their grammatical and semantic function. For example, some values are {NounInstrument}, {NounHuman}, {NounAction}, {PredicateStatic}, etc.

- Morphological feature $[\delta]$: The values shows the morphological attributes of the words, which are a pair of the word and the morphological type. For example, "uranium-235" is represented by $\delta.(uranium\text{-}235,\ alphabet\_hyphen\_arabic\text{-}numbers)$.

Each word in the sentence is represented by the set of the three features. ϕ and γ can include ambiguous values. The ith word-feature set $\mathcal{W}_i$ and the n-word sentence $\mathcal{S}_1^n$ are respectively represented as follows.

$$\mathcal{W}_i = \begin{pmatrix} \Phi_i \\ \Gamma_i \\ \Delta_i \end{pmatrix} \quad \begin{aligned} \Phi_i &= \{\phi_{i1}, \cdots, \phi_{im}\} \\ \Gamma_i &= \{\gamma_i, \cdots, \gamma_{in}\} \\ \Delta_i &= \{\delta_{i1}, \delta_{i2}\} \end{aligned} \quad (13)$$

$$\mathcal{S}_1^n = (\mathcal{W}_1, \cdots, \mathcal{W}_n) \quad (14)$$

When the conjunction is the mth word of the n-word sentence, the left-side list $\mathcal{S}_1^{m-1}$ and the right-side list $\mathcal{S}_{m+1}^n$ are respectively represented as follows.

$$\mathcal{S}_1^{m-1} = (\mathcal{W}_1, \cdots, \mathcal{W}_{m-1}) \quad (15)$$

$$\mathcal{S}_{m+1}^n = (\mathcal{W}_{m+1}, \cdots, \mathcal{W}_n) \quad (16)$$

The goal of the balance matching is to find the most symmetric pair of $\mathcal{S}_x^{m-1}(1 \le x < m)$ and $\mathcal{S}_{m+1}^y(m < y \le n)$, i.e., to find the values of x and y.

4.2 The balance matching operation

By definition, the most symmetric pair shares the maximum number of the word-feature sets in the lists. The pair is detected by three operations: the intersection operation for two features, the matching operation for two word-feature sets, and the balancing operation for two lists.

4.2.1 The intersection operation

The intersection operation is one of the normal set operations for the features:

$\Phi_{m-1}.\{$uranium-235,alphabet_hyphen_arabic-numbers$\}$
$\cap \ \Phi_{m+1}.\{$plutonium-239, alphabet_hyphen_arabic-numbers$\}$ = $\{$alphabet_hyphen_arabic-numbers$\}$

The mutual dependency information among the ϕ_{ij} and γ_{ik} is managed by bi-directional lists in the background. If all ϕ_{ij} dependent of γ_{ik} are disambiguated by the operation, γ_{ik} are removed.

4.2.2 The matching operation

The matching operation $\cap$ for the the word-feature sets $\mathcal{V}$, $\mathcal{W}$ is defined as follows:

$$\mathcal{V}_i \cap \mathcal{W}_j = \mathcal{V}_i \mathcal{W}_j = \begin{pmatrix} \Phi_i \cap \Phi'_j \\ \Gamma_i \cap \Gamma'_j \\ \Delta_i \cap \Delta'_j \end{pmatrix} \quad (17)$$

The matching word-feature set $\mathcal{V}_i \mathcal{W}_j$ can exist, when $\Phi_i \cap \Phi'_j \neq NULL$ or $\Gamma_i \cap \Gamma'_j \neq NULL$.

Especially, as is mentioned in Section 3.2, in most of the conjoined sentence, the word-feature sets of $\mathcal{W}_{m+1}$, which immediately follows the conjunction, play an important role for detecting the structure, because there is strong similarity between the starting word-feature set of the conjunct scope and $\mathcal{W}_{m+1}$.

4.2.3 The balancing operation

The balancing operation $\otimes$ for the lists $\mathcal{L}_1^n$, $\mathcal{R}_1^m$ is defined as follows:

$$\mathcal{L}_1^n \otimes \mathcal{R}_1^m = \{(\mathcal{V}_i \mathcal{W}_j, \cdots, \mathcal{V}_k \mathcal{W}_l, \cdots) | \mathcal{V} \in \mathcal{L}_1^n, \mathcal{W} \in \mathcal{R}_1^m,$$
$$k > j, \mathcal{L}_1^n = (\mathcal{V}_1, \cdots, \mathcal{V}_n), \mathcal{R}_1^m = (\mathcal{W}_1, \cdots, \mathcal{W}_n)\}$$

Every word-feature set in the list doesn't always match one of the other list. Some word-feature sets in $\mathcal{L}_1^n$ can find matching counterparts in $\mathcal{R}_1^m$ but others cannot. In (18), The_1 and $operations_3$ respectively matches the_1 and $results_2$.

(18) The_1 $arithmetic_2$ $operations_3$ and the_1 $results_2$.

Therefore, the balancing operation creates a set of the lists for exhausting all the possible combinations. The lists consist of the matching word-feature sets ($\mathcal{V}_k \mathcal{W}_l$) which are selected to avoid crossing any existing lines when $\mathcal{V}_i$ and $\mathcal{W}_j$ are connected by a line as in the following:

$$\begin{matrix} (\mathcal{V}_1, & ,,,, & \mathcal{V}_i, & ,,,, & \mathcal{V}_k, & ,,,, & \mathcal{V}_m) \\ & \nearrow & \downarrow & \downarrow & \downarrow & \searrow & \\ (\mathcal{W}_1, & ,,,, & \mathcal{W}_j, & ,,,, & \mathcal{W}_l, & ,,,, & \mathcal{W}_n) \end{matrix} \quad (19)$$

For example, $\mathcal{L}_1^3 \otimes \mathcal{R}_1^2$ is :

$\mathcal{L}_1^3 = (\mathcal{V}_1, \mathcal{V}_2, \mathcal{V}_3) \quad \mathcal{R}_1^2 = (\mathcal{W}_1, \mathcal{W}_2)$
$\mathcal{L}_1^3 \otimes \mathcal{R}_1^2 = \{(\mathcal{V}_1 \mathcal{W}_1), (\mathcal{V}_1 \mathcal{W}_2), (\mathcal{V}_2 \mathcal{W}_1), (\mathcal{V}_2 \mathcal{W}_2),$
$(\mathcal{V}_3 \mathcal{W}_1), (\mathcal{V}_3 \mathcal{W}_2), (\mathcal{V}_1 \mathcal{W}_1, \mathcal{V}_2 \mathcal{W}_2), (\mathcal{V}_1 \mathcal{W}_1, \mathcal{V}_3 \mathcal{W}_2),$
$(\mathcal{V}_2 \mathcal{W}_1, \mathcal{V}_3 \mathcal{W}_2)\}$

$$(20)$$

The balancing degree θ for a list $\mathcal{I}$ is defined as follows:

$$\theta_{\mathcal{I}} = \sum_{k \in \{\Phi, \Gamma, \Delta\}} w_k \mathcal{N}(k) \qquad \mathcal{I} = (\mathcal{V}_i \mathcal{W}_j, \cdots, \mathcal{V}_l \mathcal{W}_m)$$

$\mathcal{N}(\Phi)$ and $\mathcal{N}(\Gamma)$ respectively represent the total number of the each feature in the list $\mathcal{I}$. $\mathcal{N}(\Delta)$ is the total number of the values of Δ in the list $\mathcal{I}$. Herein,

w_k is defined as a binary value (1.0 or 0.0) for simplifying the model. Through the analysis of 10,000 English conjunctive sentences in technical manuals, the structures could be divided into about 300 coordinate patterns, which are represented in the form of the word-feature sets. We manually assigned the weights to the features Δ, Γ and Φ according to the patterns, in order to select the correct structure.

4.3 Analysis by balance matching

The n-word sentence including the conjunction as the mth word is analyzed according to the following steps.

I. Collect the word-feature set $G(\mathcal{W})$: $G(\mathcal{W}) = \{\mathcal{W}_x | \mathcal{W}_x \cap \mathcal{W}_{m+1} \neq NULL\}$
This step collects a set of the words similar to $\mathcal{W}_{m+1}$. The collection considers some definite concord markers and boundary markers, such as "and", "both", "either", commas, periods, and colons. In order to deal with the cases of (10),(11) and (12), the starting words of the clause $\mathcal{W}_k$ are added to $G(\mathcal{W})$ when a comma is preceded by the conjunction.

II. Create the list set $H(\mathcal{L})$:
$H(\mathcal{L}) = \{\mathcal{W}_j | i \leq j \leq m - 1, \mathcal{W}_i \in G(\mathcal{W})\}$

III. Create the list $\mathcal{R}_1^{n-m}$:
$\mathcal{R}_1^{n-m} = \{\mathcal{W}_{m+1}, \mathcal{W}_{m+2}, \cdots, \mathcal{W}_n\}$

IV. To create the list set $F(\mathcal{LR})$ by the balancing operation.
$F(\mathcal{LR}) = \{\mathcal{L}_i \otimes \mathcal{R}_1^{n-m} | \mathcal{L}_i \in H(\mathcal{L})\}$

V. Select the list $F(\mathcal{LR})_{max}$ which has the highest balancing degree from all the possible lists in $F(\mathcal{LR})$.

For example, (9) are analyzed by the following steps. Here, for the easy understanding, the word-feature set and the list are represented by the simplified expression.

(9) Radioisotopes have played an important part in_1 developing effective insecticides in_2 the country **and** in_3 finding the best ways of applying them.

I. $G(\mathcal{W}) = \{in_1, in_2\}$

II. $H(\mathcal{L}) = \{(in_1, developing, effective,$
$insecticides, in, the, country), (in_2, the, country)\}$

III. $\mathcal{R}_1^8 =$
$(in_3, finding, the, best, ways, of, applying, them)$

IV. $F(\mathcal{LR}) = \{\cdots, (Preposition, Nominal),$
$(Preposition, Gerund, Nominal)\}$

V.

$F(\mathcal{LR})_{max} = (Preposition, Gerund, Nominal)$
$\mathcal{L} = (in_1, developing, effective, \cdots, in, the, country)$
$\mathcal{R} = (in_3, finding, the, best, ways, of, applying, them)$
The and_5 in (21) can be analyzed as the same way.

44

(21) Use extreme care when using cleaning solvents $like_1$ acetone and methyl ethyl ketone which are highly $flammable_2$ and shall be $used_3$ in areas $with_4$ adequate fire extinguishing devices and_5 $free_6$ of ignition sources.

$G(\mathcal{W}) = \{like_1, flammable_2, used_3, with_4, free_6\}$
$F(\mathcal{LR})_{max} = (NominalPostmodifier)$
$\mathcal{L} = (with_4, adequate, fire, extinguishing, devices)$
$\mathcal{R} = (free_6, of, ignition, sources)$

5 Empirical results on the MT system

5.1 MT configuration

The model was incorporated as a balance matching module into the PIVOT English-Japanese MT system as shown in Figure.1(Muraki, 1986; Okumura et al., 1987; Okumura et al., 1991). The system works for a practical use together with the dictionary of 100,000 words.

The input sentences, represented by the list of the feature sets based on the results of the morphological analysis, are transferred to the balance matching module. The module produces the conjunct scope information as well as the results of the balance matching. The syntactic-semantic analysis module analyzes the sentences according to this top-down information. After the analysis, the conceptual analysis module creates an interlingua representation. From the interlingua, output sentences are generated(Muraki, 1986; Okumura et al., 1991).

5.2 Effects of the model

When 15,000 conjunctive sentences in the technical reports and manuals were translated apart from the analyzed 10,000 sentences, the following effects are confirmed about the model.

1. Reduction of the analysis cost: This model provides the correct top-down information about the conjunct scopes for about 75% of the sentences, which result in the accurate and effective syntactic analysis. Most of the sentences had required backtracking for the analysis without the model. The backtracking is almost all suppressed by the model.

2. Improvement of the word disambiguation: The results of the balance matching improve word disambiguation and the inferences of the unknown words, because the ambiguities of each word are intersected by the counterparts of the symmetric list. The results provide top-down information for the analysis of the ambiguous words and unknown words, as in (8). Most of the sentences contained some word ambiguities and one sixth contained unknown words. By using the top-down information, the accuracy was twice better improved.

3. Interpretation of the ellipses: The model makes it easy to interpret the elided elements, because the balance matching results can suggest the missing elements. In sentences such as (1), $\mathcal{R}$ is completely included $\mathcal{L}$. The differences of $\mathcal{L}$ and $\mathcal{R}$ complement the missing elements.

4. Robust analysis: The model helps make the system robust because the balance matching operation is based on the three different kinds of features. In the MT domain of technical reports and manuals, there are some unknown words as well as some ambiguous words as in (6),(7) and (8). Robustness is achieved because the morphological features are considered as well as the other features.

5.3 Discussions

To increase the accuracy, the model is improved from three points:

- Lexical disambiguation: The model is based on lexical information. Therefore, when many words provide too many feature ambiguities, the model cannot always determine a correct structure. In order to solve this problem, filtering rules are applied to the sentence before the balance matching operation. The rules are local constraint rules, which checks the two or three words before a focused word to remove some ambiguities of the focused word. The filtering rules improve the model.

- Weight optimization: The weights for each feature set are manually assigned based on 300 patterns. They should be more appropriately assigned as real values instead of binary values according to the domain and text styles. We have developed a learning method for the feature structures(Okumura et al., 1992). The method is applicable for determining the weights according to the input patterns.

- Semantic calculation: Some conjunctive structures should be analyzed by the more subdivided semantic features and semantic similarity calculation. We are introducing some semantic taxonomy and the semantic distance measurement algorithm(Knight, 1993; Okumura and Hovy, 1994; Resnik, 1993).

6 Conclusion

The authors propose an English coordinate structure analysis model, which provides top-down scope information of the correct syntactic structure by taking advantage of the symmetric patterns of the parallelism.

This model was practically implemented and incorporated into the PIVOT English-Japanese MT system with the dictionary of 100,000 words. The model provided about 75% accuracy, which lead to
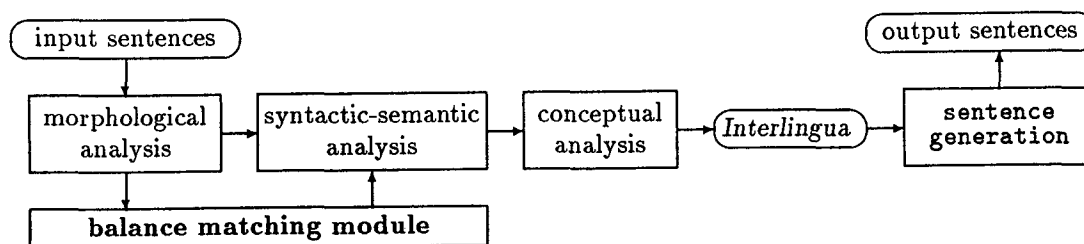
Figure 1: A MT system configuration including the balance matching module

robust analysis for the MT system, as well as helping make rule-based analysis effective. In the future, we plan to improve the model according to the above discussion, and also to extend the analysis to other conjunctions and prepositions.

7 Acknowledgments

The authors thank Eduard Hovy and Kevin Knight of USC/ISI for their precious comments.

References

R. Beaugrande and W. Dressler. 1981. *INTRO-DUCTION TO TEXT LINGUISTICS*. Longman Group Limited, Harlow, Essex.

S.A. Blackwell. 1981. Processing conjunctions in an atn parser. Technical report, University of Cambridge. Unpublished M.Phil. Dissertation.

B.K Boguraev, K. Sparck-Jones, and Y.(eds) Wilks, editors. 1983. *Recognizing Conjunctions within the ATN Framework*. Ellis Horwood.

S. Fong and R.C. Berwick. 1985. New approaches to parsing conjunctions using prolog. In *Proceedings of the 23rd ACL Conference*, Chicago.

G. Gazdar. 1981. *Unbounded Dependencies and Coordinate Structure*. .

X. Huang. 1983. Dealing with conjunctions in a machine translation environment. In *Proceedings of European ACL Conference*.

X. Huang. 1984. Dealing with conjunctions in a machine translation environment. In *Proceeding of COLING84*.

Kevin Knight. 1993. Building a large ontology for machine translation. In *Proceedings of the ARPA Human Language Technology Workshop*, Princeton, New Jersey. ARPA.

D. Kosy. 1986. Parsing conjunctions deterministically. In *Proceedings of the 24th ACL Conference*.

S. Kurohashi and M. Nagao. 1992. Dynamic programming method for analyzing conjunctive structures in japanese. In *Proceedings of COLING-92*, August.

L. Lesmo and P Torasso. 1985. Analysis of conjunctions in a rule-based parser. In *Proceedings of the 23rd ACL Conference*, Chicago.

K. Muraki. 1986. Venus: Two-phase machine translation system. *Future Generations Computer Systems 2*.

S. Niimi, Y. Suwa, J Tsujii., and .M Nagao. 1986. Coordinate sentences by mu-project english analysis grammar. In *Proceedings of Information Processing Society of Japan*. in Japanese.

A. Okumura and E. Hovy. 1994. Building a japanese dictionary associated with an ontology for machine translation. In *ARPA Human Language Technology Workshop*.

A. Okumura, S. Kamei, and K. Muraki. 1987. English coordinate structure analysis. In *Proceedings of Information Processing Society of Japan*, September. in Japanese.

A. Okumura, K. Muraki, and S. Akamine. 1991. Multi-lingual sentence generation from the pivot interlingua. In *Proceedings of MT SUMMIT III*, July.

A. Okumura, K. Muraki, and K. Yamabana. 1992. A pattern-learning based, hybrid model for the syntactic analysis of structural relationships among japanese clauses. In *Proceeding of Fourth International Conference on the Theoretical and Methodological Issues in Machine Translation*, June.

P. Resnik. 1993. Semantic classes and syntactic ambiguity. In *Proceedings of ARPA HLT Workshop*.

P. Schachter. 1977. *Constraints on Coordination*. .

Y. Shinoda. 1981. *TECHNICAL ENGLISH*. Nagumodou, Tokyo Japan. in Japanese.