

法律文における並列構造の特徴と それに基づく制限言語モデルについて

岩本秀明[†] 長野馨 永井秀利 中村貞吾 野村広郷

[†]NTT情報通信網研究所
九州工業大学 情報工学部

法律文における並列構造を分類・整理し、その特徴を明らかにする。さらに、これに基づく制限言語モデルを示す。法律文において、並列構造は極めて頻繁に現われる。これらの並列構造は、内部にまた別の並列構造を埋め込むことが多い。しかしながら、法律文の並列構造は明示的な標識を持ち、しかもその標識は並列構造間の埋め込み関係に従って用いられる。本稿では、この特徴を表現した言語モデルを示す。この言語モデルは、法律文の並列構造を正確かつ効率的に解析することができる。

An Analysis of Coordinate Structures and its Application to A Controlled Linguistic Model for Law Sentences

Hideaki IWAMOTO[†], Kaoru NAGANO, Hidetoshi NAGAI, Teigo NAKAMURA, and Hirosato NOMURA

[†]NTT Network Information Systems Laboratories
1-2356 Take Yokosuka 238-03 Japan
iwamoto@nttnlk.ntt.jp

Department of Artificial Intelligence
Kyusyu Institute of Technology
Iizuka 820 Japan
nomura@ai.kyutech.ac.jp

This paper presents a controlled linguistic model for law sentences. The linguistic model is investigated through analyses of actual law sentences with coordinate structures. In the law sentences the coordinate structures appear very frequently and embed other coordinate structures in some cases. However, the coordinate structures in law sentences have explicit markers and the choice of those markers depends on the embedding relations among the coordinate structures. We can parse law sentences correctly and efficiently by applying the linguistic model which represents the usage of these markers.

1 はじめに

現在の自然言語処理システムにおいては、詳細な辞書や文法の確立、およびより高度な言語理解のための文脈情報処理など残された問題は多い。したがって、一般の自然言語を包括的に対象とした実用性の高いシステムを構築することは極めて困難となる。そこで、対象分野を限定することにより、問題を簡潔化し、実用性および実現性の高いシステムの構築を目指すという目標を掲げることができる。本研究では、そのような対象分野として法律文を扱う。

したがって、本研究は、法律を一個のシステムとして計算機上に実現し、それが持つ情報を有効に活用するための自然言語処理技術の確立が目的となる。これにより、法律という価値の高い情報を提供できるだけでなく、自然言語処理技術の高度化をも図ることができる。なぜならば、情報の価値が高いということは、その情報の内容が高度であることを意味し、高度な内容は、高度な表現を必然的に持つ。また、それゆえに、その情報は扱い難く、流通せず、価値が高まる。したがって、法律が持つ情報を表現する法律文の言語処理には、高度な技術が必要とされるからである。

法律文は、一般の自然言語に比べて言語的制約が強く、それらに基づく特徴的な言語表現が多く現われる。これは、法律文が持つ内容の高度さが要求とするものであるが、そのために既存の技術による処理が不可能であったり非効率的であったりすることがある。法律文を正しくかつ効率的に処理するためには、この言語的制約を記述した言語モデルを利用することが必要となる。我々は、そのような言語モデルを法律文の計算機内部表現とともに詳細化を重ねつつ提案してきた。

法律文の言語的制約として、まず第一に補文標識に関する特徴が挙げられる。いくつかの補文標識は、法律要件と法律効果と呼ばれる2つの事象を結び付ける機能を持ち、言語表現と論理構造とを明確に対応付ける。このことからわかる通り、法律文を計算機の内部表現として持つためには、言語情報(格文法)と論理情報(一階述語論理)とを同時に表現する枠組みが必要となる。我々は、このような枠組みとして、素性論理構造表現形式を提案した[3]。

法律文の言語的制約として、第二に文末表現に関する特徴が挙げられる。文末表現は、それが規律する対象を明確にする。文末表現が規律する対象は、基本的に3つしかない。人・物・法律がそれである。文末表現は、まず文末表現自体を助動詞相当語化することにより、無駄な解析を避けることができる。また、文末表現が規律する対象は、すでに文末表現と依存関係にあることが分かるので、これに関する解析は省くことができる。これらのことを考慮して言語モデルを文末表現から単一化可能のように構成した[4]。

文末表現の特徴につく言語的制約として並列構造に関する特徴が挙げられるが、並列構造は従来の言語モデルでは扱っておらず、現在の処理技術においても解決困難な課題である。並列構造では、並列される語句同志が互いに互いの並列構造外との係受けの隣接性を妨げるからである。本稿では、法律文の並列構造の分析を述べ、これを基に並列構造の特徴を言語モデルとして提案する。

なお、分析の対象として、「国際動産売買契約に関する国連条約(ウィーン統一売買法)」[1]を用いる。ウィーン統一売買法は、101条320文からなる法律である。

2 並列構造の標識について

並列構造は、語句と語句とを並べることによって生じる。しかも、ただ並べるだけでは必ずしも並列構造とはならない。例えば、「女男」では「女のような男」を意味する可能性がある。さらに、「男女」では「男のような女」の他に「だんじょ」とよんで「男性と女性の対」というある種の全体概念を表す可能性もでてくる。「女と男」もしくは「男と女」というような表現であれば、問題なく並列構造と見做すことができる。このように、並列構造は、接続詞、接続助詞、個条書、点、およびポーズ等により確実に示すことができる。

また、例えば「男と女」のような並列構造で、並列されるそれぞれは、「人」という上位概念を共有するという類似性を持つ。しかし、「美女と野獣」のような例では、類似性よりも差異性がより顕著である。対象分野によっては、このような差異性を考慮しなければならない。しかし、差異性が顕著であれば、接続詞などの並列構造を示す標識は、かえって強調されることになる。

体言句の並列構造をさらに広く捉えると、列挙や網羅のための表現をこれに含めることが可能である。箇条書や句点がこの場合の標識となる。例えば、「着替え、寝具、洗面用具、バスタオル」のように表す。これに加え、「…等の…」という表現を用いて「着替え、寝具、洗面用具、バスタオル等の身の回りの品」と表すと、「着替え、寝具、洗面用具、バスタオル」という列挙項目の上位概念が「身の回りの品」であることが示される。これに対して、「…その他の…」を用いて「着替え、寝具、洗面用具、バスタオルその他の身の回りの品」と表すと、その列挙項目の上位概念が「身の回りの品」であることが示されるだけでなく、列挙項目以外の「身の回りの品」も網羅することを意味する。「…等の…」や「…その他の…」は、それぞれ列挙および網羅の標識となる。また、列挙および網羅の表現では、類似性だけが顕著であり、差異性はほとんどない。

用言の並列構造は、接続詞などの明示的な標識を取らないかぎり、連用中止形がその標識となる。例えば、

犬はよろこび庭かけまわり猫はこたつで丸くなる。

という文では、「よろこぶ」の活用型「よろこび」と「かけまわる」の活用型「かけまわり」とが連用中止である。連用中止型「かけまわり」によって「犬はよろこび庭かけまわり」と「猫はこたつで丸くなる」とは並列されるが、連用中止型「よろこび」によって「よろこぶ」と「庭かけまわり」とが必ずしも並列されるとは限らない。なぜなら、それらの用言句の関係として、継起的なもの、因果的なもの等が考えられうるからである。これらは、純粋な等位関係とは相異なる。その活用型が連用中止である用言を含む節は、従属節になりやすく、並列構造でなければ主節とはならない。これに対して、連用中止型「かけまわり」によって並列される節は、用言(「かけまわり」と「丸くなる」)とそれに対して同様の格関係を持つ格要素同志(「犬」および「猫」ならびに「庭」および「こたつ」)が類似しているだけでなく、それぞれ及びそれらの組み合わせの対比が鮮明である。対比は類似性を基にした差異性に基き、差異性が顕著であるほどその鮮明さはます。この差異性により、連用中止型「かけまわり」が示す並列構造は明確になる。

つまり、連用中止のみを並列構造の標識と見なすことはできない。例文は、句点を使ってより内容を明確にすることができるが、「犬はよろこび、庭かけまわり、猫はこたつで丸くなる」と表現するより、「犬はよろこび庭かけまわり、猫はこたつで丸くなる」と表現するほうがより本来の意味を表わす。この場合、並列構造の標識は、連用中止型+句点「かけまわり、」である。また、「犬はよろこび、庭かけまわり猫はこたつで丸くなる」に関しては非文に近い。句点も用い方を制限すれば、並列構造を効果的に示す標識である。

法律文は、並列構造の標識に言語的制約を持つ対象分野である。次節では、その標識を糸口にして、法律文の並列構造を分析する。

3 法律文における並列構造

法律文の並列構造に用いられる標識は、接続詞、接続助詞、その他の標識がある。接続詞は、頻度順に、「又は」、「及び」、「若しくは」、「かつ」、「また」、「並びに」、「あるいは」がある。同様に、接続助詞には、「と」、「や」、「か」がある。接続詞および接続助詞以外の標識として、「その他の」、「その他」、さらに明示的でない標識がある。

法律文全320文中166文までが、少なくとも一つの接続詞による並列構造を持つ。つまり、全文中半分強は、接続詞による並列構造を持つ。接続助詞は、接続詞に比べて頻度が低い。接続詞による並列構造の個数が293であるのに対し、接続助詞による個数は40であり、接続詞の1/7弱である。さらに、接続詞及び接続助詞以外の標識の個数は、54であり、接続助詞による並列構造より多い。法律文における並列構造の総数は、387である。その分類を次ページの表に示す。

法律文の並列構造は、ほとんど明示的な標識を持つ。例外は、用言句および個条書を並列する場合に現われる。文末表現とそれが規律する対象は、並列されることが少なく、文末表現を含む用言句が並列されるときには、連用中止のみによって並列される。通常は、明示的な標識とともに現われる。しかし、個条書きの並列は、個条書き自体が並列構造を表す標識であって、本来接続詞等を用いないほうが多い。

	体言句	用言句	格要素	個条書	合計
又は	130	51	4	2	187
及び	49	0	0	1	50
若しくは	19	4	1	0	24
かつ	0	17	4	0	21
また	1	3	3	0	7
並びに	2	0	0	0	2
あるいは	0	2	0	0	2
と	19	1	0	0	20
や	15	0	0	0	15
か	1	4	0	0	5
その他の	21	0	0	3	24
その他	2	0	0	0	2
∅	0	10	0	18	28
合計	259	92	12	24	387

並列構造の分類

接続詞及び接続助詞以外の標識である「その他の」および「その他」は、第2節で述べたように網羅的な並列構造の標識である。例えば、

当該権利又は請求が、買主の提供した技術設計、デザイン、混合法その他の仕様に売主が従った結果生じた場合

という文では、網羅の標識「その他の」によって、「仕様」が、「技術設計」、「デザイン」、「混合法」の上位概念であることを示し、かつ、それらの3つの列挙項目と「その他の仕様」が並列される。このように、網羅的な並列構造は、接続詞による並列「権利又は請求」などに比べて、かなり複雑な構造を持つ。

接続詞の中で、「また」は、「権利又は請求」のような単純な構造をとることはない。「また」によって並列される語句は、法律要件を埋め込む。例えば、

この条約により規律される事項で、条約中に解決方法が明示されていない問題については、この条約の基礎にある一般原則に従い、そのような原則がない場合には、国際私法の規則により適用される法に従って解決すべきものとする。

という文における「そのような原則がない場合には」がその法律要件である。これは、接続詞を伴わない連用中止が文末表現を並列することと同様、詳細な検討が必要である。さらに、法律文には、「ただし、～する場合に限る」という特徴的な表現がある。これは、接続詞と文末表現の組であり、直前の文の法律効果に自らが持つ要件を付加する。したがって、直前の文にその効果を導く別の要件があれば、「ただし、～する場合に限る」に含まれる要件は、直前の文にある要件と並立する。これらは、並列構造、文末表現、および法律要件と法律効果に関する言語的制約が重なる部分であるから、まず純粋な並列構造に関する制約を明確にしておく必要がある。しかも、これらの現象は法律文全体から見れば、それほど比率を占めず、問題解決によっても、全体として高い効果を期待できない。

以下より問題を簡潔化するために、「また」を除いた接続詞による並列構造に議論を絞る。これらの頻度は286であり、並列構造の総数387に対して73.901...%を占める。第3.1節および第3.2節では、それぞれ体言句および用言句の並列構造について述べる。第3.3節では、それぞれの埋め込み関係について述べる。

3.1 体言句の並列構造

体言句の並列構造に用いられる標識は、「又は」、「若しくは」、「及び」、「並びに」がある。「かつ」は、体言句の並列構造を示す標識ではない。「あるいは」は、他の法律では、「又は」の代わりに用いられることがあり、ここでの分類にある通り、体言句を並列しないということではないと考える。「又は」と「若しくは」は、体言句だけでなく、用言句をも表すのであるので、頻度2で用言句を並列する「あるいは」は、これらに準ずるものと考えられる。英文のウィーン統一売買法を見ると、「又は」、「若しくは」、「あるいは」は、「or」で表され、「及び」、「並びに」は、「and」で表される。

「or」で表される「又は」、「若しくは」、「あるいは」の違いは、それらのあいだの埋め込み関係にある。それらが示す並列構造の埋め込み関係は、

又は ⊃ 若しくは ⊃ あるいは

となる。例えば、

この義務の違反には、一定の期間物品が通常の目的若しくはある特定の目的に適するという保証 若しくは一定の品質若しくは特性を保持するという保証の違反が含まれる。

という文では、「若しくは」によって並列構造が示される「通常の目的若しくはある特定の目的」および「品質若しくは特性」が「又は」によって示される並列構造の中に埋め込まれている。なお、埋め込みを持たない並列構造では「又は」を用いる。1重に埋め込まれた並列構造には「若しくは」を、2重に埋め込まれたものには「あるいは」を用いる。したがって、「又は」の頻度が最も高くなる。

同様に、「and」で表される「及び」と「並びに」の違いも、それらのあいだの埋め込み関係にある。それらが示す並列構造の埋め込み関係は、

及び ⊆ 並びに

となる。例えば、

この契約は、売買契約の成立並びに売買契約から生ずる売主及び買主の権利及び義務のみを規律する。

という文では、「及び」によって並列構造が示される「売主及び買主」および「権利及び義務」が「並びに」によって示される並列構造の中に埋め込まれている。なお、埋め込みを持たない並列構造では「及び」を用い、埋め込みがあるときは、埋め込むほうに「並びに」を用い、埋め込まれるほうに「及び」を用いる。「or」を意味する接続詞とは、逆の用い方である。

また、「or」を意味する接続詞が示す並列構造と「and」を意味する接続詞が示す並列構造との埋め込み関係が考えられるが、「及び」の示す並列構造が「又は」の示す並列構造の中に埋め込まれるだけであり、その他の埋め込み関係はない。「あるいは」は、ともかくとして、「及び」が示す並列構造は、他の並列構造を内部に埋め込まない。したがって、「若しくは」や「あるいは」の示す並列構造が「及び」の示す並列構造を埋め込むことはあっても、その逆はない。さらに、「又は」の示す並列構造が「並びに」(したがって「及び」による並列構造を含む)の示す並列構造を埋め込む可能性はあるが、逆に「並びに」が示す並列構造によって、「又は」(したがって「若しくは」「あるいは」による並列構造を含む)が示す並列構造が埋め込まれる可能性がある。

3.2 用言句の並列構造

用言句の並列構造に用いられる標識は、「又は」、「若しくは」「あるいは」、「かつ」がある。「及び」と「並びに」は、用言句の並列構造を示す標識でない。前節と同様英文のウィーン統一売買法を見ると、「又は」「若しくは」「あるいは」は「or」で表され、「かつ」は「and」で表される。

「or」で表される「又は」、「若しくは」、「あるいは」の違いは、それらのあいだの埋め込み関係にある。それらが示す並列構造の埋め込み関係は、

又は ⊇ 若しくは ⊇ あるいは

となる。例えば、

買主が不適合を発見し又は発見すべきであった時より以前に、物品の全部又は一部が通常の営業過程で買主により売却され、又は通常の用法で消費若しくは改変された場合。

という文では、「若しくは」によって並列構造が示される「消費若しくは改変される」が「又は」によって示される並列構造の内部に埋め込まれる。なお、埋め込みを持たない並列構造では「又は」を用いる。1重に埋め込まれた並列構造には「若しくは」を、2重に埋め込まれたものには「あるいは」を用いる。したがって、「又は」の頻度が最も高くなる。これらの接続詞の用い方は、前節のものと同しいが、並列される内容は異なる。

「and」で表される「かつ」は、「or」を意味する接続詞が示す並列構造をすべて埋め込む。理論的には、「又は」の示す並列構造を埋め込めば、その他の「or」を意味する接続詞による並列構造をすべて埋め込む。よって、

かつ ⊇ 又は ⊇ 若しくは ⊇ あるいは

と埋め込み関係を表すことができる。例えば、

物品の不適合が、売上の知り又は知らないはずはあり得なかった事実で、 若しくは、 売主がそのことを買主に対して明らかにしなかった事実に関連するときは、 売主は、第38条及び第39条の規定を援用することができない。

という文では、「又は」によって並列構造が示される「知り又は知らないはずはあり得なかった」が「かつ」によって示される並列構造に運込まれている。「かつ」は、用言句の並列にしか用いられない。また、その他の並列構造に埋め込まれることはない。

3.3 多重の並列構造

他の並列構造を埋め込む並列構造(これを多重の並列構造と呼ぶことにする)については、第3.1節および第3.2節でもそれぞれ体言句の並列構造および用言句の並列構造という枠内に限ってだが示してきた。本節では、これらをまとめた埋め込み関係について述べる。接続詞による並列構造は293あるので、接続詞による並列構造を持つ文は、平均 $293/166 = 1.765\dots$ の並列構造を持つ。したがって、並列構造を持つ文は、2重以上である頻度が高い。よって、より緻密に並列構造間の埋め込み関係を明らかにすることの効果は大きい。

下の図に体言句や用言句の並列構造を示す接続詞の埋め込み関係を示した。

あるいは」や「及び」は、他の並列構造を埋め込まない並列構造の標識である。これに対して、「and」を意味し、用言句の並列のみに用いられる接続詞である「かつ」は、体言句の並列構造に用いられる接続詞をほとんどその並列構造内に埋め込む。例外は、「並びに」である。原因は頻度がすくないためで、「並びに」が特殊であるためではない。

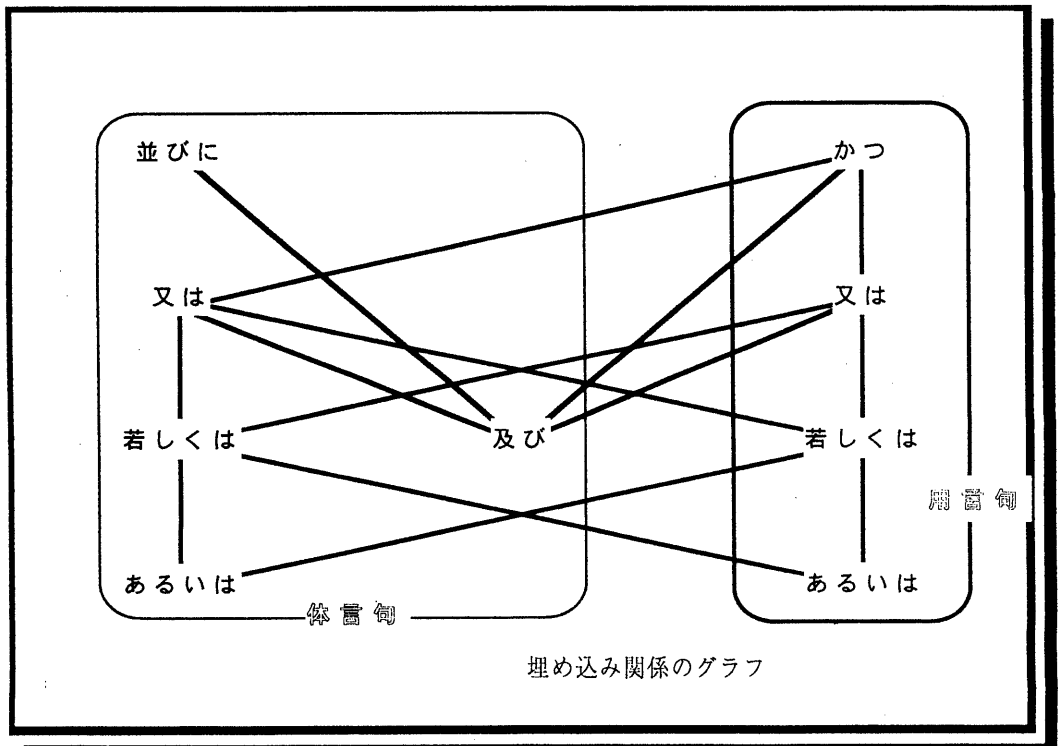
「並びに」が示す並列構造に唯一埋め込まれる「及び」が示す並列構造も「かつ」による並列構造には埋め込まれる。例えば、

申入れは、物品を示し、かつ、明示又は黙示によりその数量及び代金を定め又はこれを定める条項を含む場合には、内容が十分に確定しているものとする。

という文においては、「及び」によって並列構造が示される「数量及び代金」が「かつ」によって示される並列構造に埋め込まれる。さらに、「又は」によって示される用言句の並列構造は、「及び」によって示される体言句の並列構造を埋め込む。例えば、例文の中の

物品を示し、かつ、明示又は黙示によりその数量及び代金を定め又はこれを定める条項を含む

という部分では、「及び」によって並列構造が示される「数量及び代金」が「又は」によって示される並列構造に埋め込まれる。よって、「及び」が示す並列構造は、「並びに」が示す並列構造(体言句)、「又は」が示す並列構造(体言句と用言句の双方)、「かつ」が示す並列構造(用言句)により埋め込まれることになる。



さらに、"or"を意味する「又は」「若しくは」「あるいは」は、体言句および用言句のどちらにも用いられるので、それらのあいだの埋め込み関係を考慮しなければならない。基本的には、埋め込まれるものが体言句であろうが用言句であろうが関係なく、例えば、

前号の規定が該当しない場合において、契約が、特定物~~又は~~特定在庫品中から抽出されるべき不特定物~~若しくは~~製造あるいは生産される不特定物に関するものであり、~~かつ~~、契約締結時に、両当事者が、物品が特定の場所に存在し又はそこで製造あるいは生産されることを知っていた場合 — その場所で物品を買主の処分にあつて委ねること。

という文において、「あるいは」によって並列構造が示される「製造あるいは生産される」は、用言句であるが、「若しくは」により並列構造が示される「特定の在庫品中から抽出されるべき不特定物若しくは製造あるいは生産される不特定物」に埋め込まれる。これは、さらに「又は」「かつ」により示される並列構造に順に埋め込まれる。このように、埋め込みの順序さえ正確であれば、体言句の中に用言句を埋め込むことは可能である。

法律文には、さまざまな言語的制約とそれに基づく言語的特徴がある。特にここで述べた、並列構造が埋め込み関係を持つときの標識の用い方は、その構造を明確にするために極めて有効である。

4 言語モデル

前節で述べた法律文の並列構造の分析に従い、言語モデルを付録に示す。並列構造を3つのカテゴリーに分類している。上から順に、用言句に関する並列構造、体言句に関する並列構造、サ変動詞に関する並列構造をそれぞれの素性構造が表す。

この言語モデルは、構文解析に先んじて、形態素解析の出力と単一化される。3つの素性構造の適用の順序は、サ変動詞の並列構造、体言句の並列構造、用言句の並列構造、である。

これにより、並列構造が品出する法律文は、正しくかつ効率的に解析することができる。

5 おわりに

前節で述べた言語モデルを利用することによって、並列構造に関する処理を可能にし、かつ、その負担を軽減することができる。また、対象言語が法律文であるために、並列構造の頻度は高い。したがって、結果として法律文の処理に対するその効果は倍増する。

今後は、この言語モデルを詳細化し、従来の言語モデルと組み合わせることによって、さらに正確かつ効率的な処理を可能とする言語モデルを構築していきたい。

参考文献

- [1] 国際的動産売買契約に関する国連条約(ウィーン統一売買法)
- [2] 野村浩郷, 自然言語処理の基礎技術, 電子情報通信学会, 1988
- [3] 岩本, 野村, 法律文の自然言語処理について, 情報処理学会, NL83-2, 1991
- [4] Iwamoto, H., K. Nagno, H. Nagai, H. Nomura, "Analysis of Sentence Endings and its Application to Linguistic Model for Law Sentences", International Symposium on Natural Language Understanding and AI, 1992
- [5] 岩本, 法律文の並列構造とその言語モデルの研究, 九州工業大学 情報工学部 平成2年度修士論文, 1993
- [6] 長尾, 辻井, 田中, 石川, 科学技術論文における並列句とその解析, 情報処理学会, NL36-4, 1985
- [7] 石川, 辻井, 長尾, 英語並列句の処理とその曖昧さについて, 情報処理学会, 自然言語処理研究会, NL50-2, 1987
- [8] 田村, 田中, 意味解析に基づく並列名詞句の構造解析について, 情報処理学会, 自然言語処理研究会, NL59-2, 1987
- [9] 村田, 等位構造および二重現象と機械翻訳, 情報処理学会, 自然言語処理研究, NL71-7, 1989
- [10] 黒橋, 長尾, 長い日本語における並列構造の推定, 情報処理学会, 自然言語処理研究会, NL86-2, 1991
- [11] 武田, 英日機械翻訳システムにおける並列関係の検出, 情報処理学会, 自然言語処理研究会, NL91-2, 1992
- [12] 武藤, 中川, 日本語の名詞句と等位接続詞「と, や, か」の形式意味論に関する考察, 情報処理学会, 自然言語処理学会研究, NL96-5, 1993

Appendix

