

日本語慣用表現の分析と日英翻訳への適用

奥 雅博

NTT電気通信研究所

本稿では、日英翻訳の日本語解析に適用することを念頭において、慣用表現のうち1つの格要素と述語からなる述語相当語句とみなせる表現を対象とした慣用表現解析方式を提案する。慣用表現は、述語相当語句だけに限っても、科学技術文に対して約3～5%の割合で出現する。この割合は、決して無視できるものではなく、慣用表現を正確に解析することは、日本語を解析する上で重要な問題である。述語相当語句の慣用表現は、

- ・ 慣用表現の構成語の意味と全体の意味との関係
- ・ 構成語間の結合度

の2点に着目すると、次の3つに分類できる。

1. 慣用句
2. 強い語結合
3. 機能動詞結合

本慣用表現解析方式は、それぞれの分類の持つ特徴を用いることによって、述語相当語句の慣用表現の約90%を解析の対象とすることができる。

The Classification of Japanese Idiomatic Expressions and its Application in Japanese-to-English Machine Translation

Masahiro OKU

NTT Electrical Communications Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa 238, Japan

This paper presents analysis methods for Japanese idiomatic expressions which work as predicates. In Japanese-to-English translation, it is important to analyze Japanese idiomatic expressions because correct translations cannot be made by simply translating each component word. In this method, Japanese idiomatic expressions which work as predicates are classified based on the following two points:

- ・ The meaning relation between component words and the whole expression.
- ・ The strength of combination between component words.

This process produces three classifications.

1. Idioms.
2. Expressions with strongly bound component words.
3. Expressions with functional verbs.

Working with the features of each of these three classifications, this method can successfully analyze about 90% of all predicate idiomatic expressions in Japanese.

1. はじめに

言語族の異なる日本語から英語への翻訳においては、日本語の各単語に対する英単語（熟語）を並べるだけでは正しい訳語が得られないことが多い。これは、日本語の単語が英語の表現に対応するだけでなく、より広い範囲の表現が英語表現に対応するのが一因である。すなわち、日本語表現の構造自体が意味を持っているために、各単語が英語表現に対応するのではなく、全体として英語表現に対応するためだと考えられる。表現の構造と意味が最もよく結び付いているものに慣用表現がある。この慣用表現は、「油を売る」のように、各単語の意味を分析しても全体の意味が理解できない表現や、「背が高い」のように各単語の意味から表現全体の意味が理解できる表現など、様々である。

慣用表現は、述語相当語句だけに限っても、科学技術文に対して約3～5%の割合で出現する。この割合は、決して無視できるものではなく、慣用表現を解析することは、日本文を解析する上で重要な問題である。

本稿では、日本語において特に重要であると考えられる述語相当語句の慣用表現を取り上げ、その約90%が解析可能な方式について述べる。この方式は、表現全体の意味と各構成語の意味との関係、および、構成語間の結合度の両者に着目して分類した慣用表現の特徴を利用しており、本稿ではこの分類についても述べる。

2. 本稿で扱う慣用表現の範囲

慣用表現という語は、一般的には、2つ以上の単語の連結体であって、その結び付きが比較的固く、全体としてある意味を持つ表現と定義されている[1]。

このように定義された慣用表現は、文法的に見ると、それ自体が格言・ことわざのように文として働くもの（例：光陰矢のごとし）、体言として働くもの（例：何食わぬ顔）や用言として働くもの（例：油を売る）など多岐にわたっている。また、慣用表現の構成を見ると、名詞のみによる表現（例：山の神）、名詞と用言とによる表現（例：油を売る）など、種々の表現が存在する。

このように多岐にわたる慣用表現すべてを一様に扱うのは、困難であると考えられる。そこで本稿では、扱う慣用表現の範囲を次のように限定する。

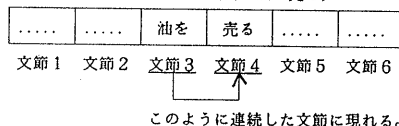
☆慣用表現 = 「日本語において、1つの格要素と述語からなり、述語相当語句であることとみなせる慣用表現」。

この限定によって、慣用表現の構成は、「名詞＋助詞＋用言」に決まる。

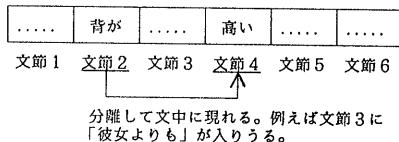
3. 慣用表現の分類・特徴

慣用表現を分類するのに、構成語間の結合度、および、慣用表現の構成語の意味と表現全体の意味の関係の2点に着目する。

(1) 結合度の強い慣用表現（例：油を売る）



(2) 結合度の弱い慣用表現（例：背が高い）



＜図1： 結合度の違いによる慣用表現の現れ方＞

構成語間の結合度とは、慣用表現を構成する語の文中での現れ方である。慣用表現の文中での現れ方の例を図1に示す。図1に示すように、結合度が強い表現においては、慣用表現を構成する語が文中に連続して現れ、結合度が弱い表現においては、慣用表現を構成する語間に他の語句（副詞や別の格要素など）が入った形で現れやすい。

慣用表現の構成語の意味と表現全体の意味の関係とは、各構成語の意味がどの程度表現全体の意味に反映しているかである。慣用表現には、「油を売る」のように、各構成語の意味を合成しても表現全体の意味が出てこないものから、「背が高い」のように、各構成語の意味を合成することによって全体の意味を理解できるものまでが存在する[2]。

本稿では、上記の2点に着目して、慣用表現を次の3つに分類する。

- 慣用表現
- (1) 慣用句（例：油を売る）
 - (2) 強い語結合（例：背が高い）
 - (3) 機能動詞結合（例：実験を行う）

以下に、各表現の持つ特徴を示す。

(1) 慣用句

慣用句は、構成語間の結合度が強く、構成語の意味を合成しても表現全体の意味を表さない表現である。

- (a) 構成語間に他の要素（副詞、格要素など）が非常に入りにくい（結合度が強い）。
 - ・油を売る → ×油を スタンドで 売る
- (b) 構成語の名詞が、自由に連体修飾を受けない。
 - ・ほぞをかむ → × 大きな ほぞをかむ
- (c) 表現自体が、文法的意味（様相など）を含んでおり、その文法的意味を除くと意味が変化する。
 - ・花を持たせる → ×花を持つ（別の意味）

(2) 強い語結合

強い語結合は、構成語間の結合度がさほど強くなく、間に副詞や他の格要素が入りやすい。また、表現全体の意味は、構成語それぞれの意味に基づいて導き出される。

- (a) 構成語間に他の要素が入る。

・効率が良い → 効率が 非常に 良い

(b) 構成語の名詞が連体修飾を受ける。

・背が高い → 彼の 背が 高い

(3) 機能動詞結合 [3]

機能動詞結合は、動作名詞と機能動詞とが結びついた表現である。機能動詞結合は、構成語間の結合度が比較的高く、構成語である動作名詞が表現全体の意味のほとんどの部分を担っている。機能動詞とは実質的な意味を動作名詞に預けて自らはもっぱら文法的な機能を果たす動詞をいう。機能動詞は本来の実質的な意味を失っている。

(a) 表現全体が動詞だけのときは違う独自の格支配をみせる。

・連絡を取る → 彼と 連絡を取る

(「取る」だけでは必須格のト格をとれない)

(b) 表現全体が文法的意味(様相など)を持つ。

・影響を受ける → 影響される(受身)

(c) 構成語間に他の要素が入りうる。

・実験を行う → 実験を 地下室で 行う

(d) 動作名詞が連体修飾を受け、表現内容を豊かにする。

・影響を受ける → 彼の 影響を受ける

上記の各表現の持つ特徴から、結合度の強い表現ほど、全体の意味に対する各構成語の持つ意味の寄与が小さくなることがわかる。

4. 日英翻訳への適用

日本語慣用表現を含む文を英語に変換することを考えると、慣用表現は複数の語が共起することによってその意味、すなわち訳語が決まるため、単純に単語に訳語を与える方法では意味が変化してしまい、正しい訳語が生成できないことが多い。特に、述語相当語句の慣用表現は、慣用表現を構成する語だけでなく、他の格要素にも影響を及ぼすことから、文中にこういった表現が存在しているかどうかを解析することは、日英翻訳において翻訳の質を高めるという点からも重要な問題である。しかしながら、慣用表現は従来のシステムではほとんど扱われておらず、わずかにM u pプロジェクトに見られる程度である[4]。

ここでは、上記の分類・特徴を利用して、慣用表現を含む文を解析する方式について述べる。本方式は、分かち書き、文節間係り受け解析がすでに終了していることを前提とし、慣用表現を中心とした結合価構造を用いて入力文を解析することを基本としている。

4.1 事例調査

1つの格要素と用言とからなる述語相当語句の慣用表現を日英翻訳に適用するにあたって、実際の文中で、慣用表現がどのような現れ方をするか、特に、構成語である名詞(以下、構成名詞という)への修飾語句はどういう語句で

表1: 慣用表現の調査結果

(1) 慣用句・強い語結合について(全件数を100としたときの相対値)

		新聞	科学分野
慣用句・強い語結合	単純な慣用句・強い語結合	75.5	82.8
	構成名詞が複合名詞のもの	2.4	3.4
	構成名詞への修飾語句が「名詞+助詞」のもの	14.2	8.6
	構成名詞への修飾語句が連体詞のもの	1.2	0
	構成名詞への修飾語句が形容詞文のもの	形容詞単独	2.0
		形容詞に修飾があるもの	0.4
	構成名詞への修飾語句が動詞文のもの	動詞単独	0
		動詞に修飾があるもの	4.3

(2) 機能動詞結合について(全件数を100としたときの相対値)

		新聞	科学分野
機能動詞結合	単純な機能動詞結合	34.8	22.0
	構成名詞が複合名詞のもの	20.7	23.9
	構成名詞への修飾語句が「名詞+助詞」のもの	30.7	18.8
	構成名詞への修飾語句が連体詞のもの	1.4	2.2
	構成名詞への修飾語句が形容詞文のもの	形容詞単独	7.6
		形容詞に修飾があるもの	2.4
	構成名詞への修飾語句が動詞文のもの	動詞単独	0
		動詞に修飾があるもの	2.4

(3) 慣用表現の出現頻度

	新聞	科学分野
慣用句・強い語結合	2.1%	0.6%
機能動詞結合	3.1%	3.0%

あるかを調査した。調査対象は、新聞記事約12,000文、教科書・論文誌などの科学技術分野約9,000文である。この結果を表1に示す。表1においては、慣用句と強い語結合を同一に扱っている。これは、慣用句は結合度が強く、強い語結合はそれほどでもない、全体の意味と各構成語の意味の関係が異なるなどの違いはあるものの、各構成語が共起して初めて全体としての意味が成立するという点で、慣用句と強い語結合の両者が共通しているためである。

慣用表現は、慣用表現全体(慣用句・強い語結合と機能動詞結合の両者)を対象とすると、全文に対して約3~5%の割合で出現する。この割合は多くはないものの無視できるものではないと考えられる。

表1より、各表現について次のことが言える。

(1) 慣用句・強い語結合

表1より、慣用句・強い語結合のほとんどは構成名詞を修飾する語句が存在しない「単純な慣用句・強い語結合」である。構成名詞を修飾する語句が存在する表現には、表1に示すように、形容詞文、動詞文、名詞+助詞の文節、連体詞の4種類からの修飾を受けるものがある。形容詞文、動詞文は、埋め込み文として構成名詞を限定し、連体詞も

また構成名詞の持つ意味を限定する。しかし、これらの表現は、英語との対応を考えると、構成名詞の訳語を限定する形で翻訳する必要がある。これに対して、名詞＋助詞の文節は、慣用句・強い語結合に対して格要素となる場合が多く、英語において主語や目的語などの重要な役割を果たす。

前者は、出現率も小さいので、本稿では、表1における「単純な慣用句・強い語結合」と「構成名詞への修飾語句が名詞＋助詞のもの」の2つを扱うことができる方式を提案する。本方式は、両者を扱うことによって、慣用句・強い語結合全体の約90%を解析の対象とすることができる。

(2) 機能動詞結合

機能動詞結合についても、(1)と同様のことが言えるが、構成名詞が複合名詞である場合(例：研究開発を行う)がかなり存在する。この表現を扱うには、まず、複合名詞を構成する語基間の関係を認定し[5]、その後、主名詞と機能動詞との関係を解析する必要がある。また、構成名詞を修飾する語句として形容詞単独のものが比較的多い。この形容詞単独の部分は、英語において副詞1語に対応するものがほとんどである。

以上のことから、機能動詞結合としては、表1における「単純な機能動詞結合」、「構成名詞への修飾語句が名詞＋助詞のもの」、「構成名詞が複合名詞のもの」、および、「構成名詞への修飾語句が形容詞単独のもの」の4種類を扱うことができる方式を提案する。この4種類を扱うことによって、本方式は、機能動詞結合全体の約85%を解析の対象とすることができる。

4.2 慣用句・強い語結合の解析方式

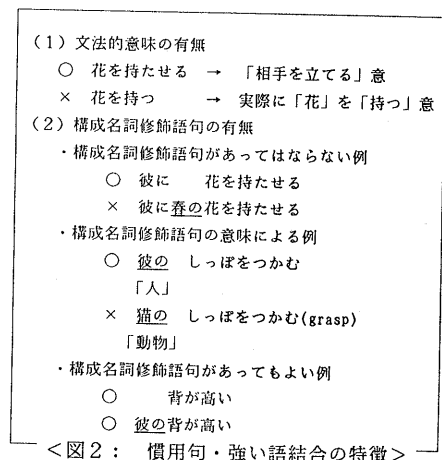
文中に存在する表現が慣用句・強い語結合であるためには、構成語すべてが揃っていて、かつ、両文節が係り受け関係で結ばれている必要がある。しかし、これだけでは、慣用句・強い語結合として認定することはできない。これは、2章で述べた特徴によるものである。

4.2.1 解析上の問題点

- (1) 慣用句・強い語結合の表現自体に、様相などの文法的意味を持つものが存在し、これらについては文法的意味を含めて慣用句・強い語結合として認定しなければならない(図2-(1))。
- (2) 慣用句・強い語結合の中には、構成名詞を修飾する語句(ここでは、名詞＋助詞のもののみを対象としている：以下、構成名詞修飾語句という)の存在によって意味が変化したり、構成名詞修飾語句の持つ意味によって慣用句・強い語結合になるものやないものがある(図2-(2))。

4.2.2 解析方式の概要

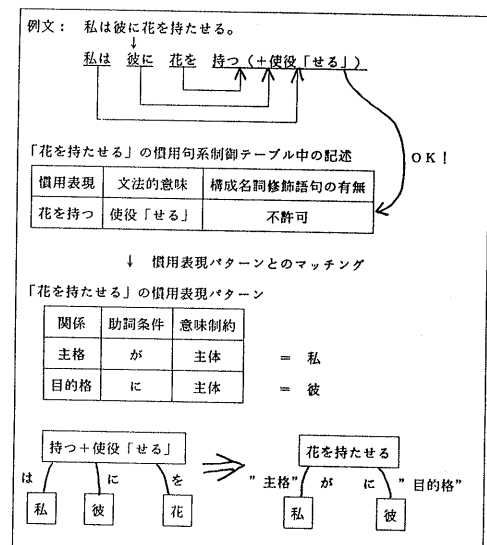
慣用句・強い語結合は、1つの格要素と用言とからなる



が、この両者をまとめて、あたかも1つの用言であるかのように考えると、慣用句・強い語結合を中心とした結合価構造を与えることができる。上記の問題点(1)、(2)は共に、文中の表現が慣用句・強い語結合として使われているかどうかを決める条件であると考えられる。

ここでは、これらの条件を慣用句系制御テーブル中に記述し、条件を満たしたときにのみ、慣用表現の結合価パターン(慣用表現パターン：慣用表現とその支配要素となる格要素との関係を記述したもの[6])を検索する。慣用句系制御テーブルの記述内容は、次の2点である。

- ・ 表現が含まなければならない文法的意味。
- ・ 構成名詞修飾語句を取るかどうか、取る場合にはその主名詞の意味カテゴリ。



<図3： 例文に対する慣用句・強い語結合の解析の概要>

慣用句・強い語結合の解析の概要を慣用表現「花を持たせる」を含む文「私は彼に花を持たせる」を例にとって説明する。

まず、慣用句系制御テーブルに記述されている条件を入力文が満足するかどうかを調べる。図3に示すように、慣用表現「花を持たせる」は用言「持つ」に文法的意味＝使役「せる」を要求するが、構成名詞修飾語句の存在を認めていない。この条件によって、「私は花を持つ」や「彼に春の花を持たせる」といった表現は、慣用表現ではないと判定することができる。

例文「私は彼に花を持たせる」は、この両者を満足しているため、慣用表現「花を持たせる」を含むと判定できる。

次に、慣用表現パターンを検索し、入力文とのマッチングを、格要素の持つ助詞と主名詞の持つ意味カテゴリの条件をチェックすることによって行う[6]。この結果、例文では、「主格＝私」、「目的格＝彼」と認定し、木構造を図3に示すように変換する。

以上のようにして、慣用表現「花を持たせる」を入力文の中から抽出し、他の格要素との関係を解析することができる。

4. 3 機能動詞結合の解析方式

機能動詞結合は、動作名詞と機能動詞とが結合した表現であるため、その数は動作名詞の数と機能動詞の数の積となる。この数は非常に大きなものであり、1つ1つの機能動詞結合について解析規則を記述するのは、實際上、不可能である。

そこで、機能動詞結合の特徴（表現全体の実質的な意味は動作名詞が担っていて、機能動詞は文法的意味を持っているにすぎない）を利用して、図4に示すように、機能動詞結合を含む文を動作名詞の用言形を述語とした文に変換する。この方式は、解析規則の数を、機能動詞の数程度に抑えることができるという利点を持つ。

A国が B国の 支配を 受ける
↓ (機能動詞「受ける」：受身)
A国が B国に 支配される
(A国を B国が 支配する(+受身))

<図4： 機能動詞結合の解析>

4. 3. 1 解析上の問題点

- (1) 同じ字面を持つ機能動詞であっても、その文法的意味が異なる場合がある(図5-(1))。この違いは、機能動詞と結合している動作名詞が自動詞的であるか、他動詞的であるかなど、機能動詞というよりも動作名詞の持つ特徴に依存している。
- (2) 動作名詞の用言形を述語とした文に変換する際には、もともと機能動詞を修飾していた文節、および、動

- (1) 複数の文法的意味を持つ機能動詞の例(受ける)
 - ・支配を 受ける → 支配される(受身)
 - ・感動を 受ける → 感動する(基本態)
- (2) 助詞の変化の例(与える)
 - ・Bに注意を与える → Bに注意する
- (3) 助詞の変化の例(受ける)
 - ・Bの支配を受ける → Bに支配される
- (4) 本動詞として使用されるもの
 - ・機能動詞「集める」 → 受身
 - ・署名を 集める → × 署名される(受身)
 - ○ 「集める」は本動詞
- (5) 文法的意味の違うもの
 - ・自動詞化する動作名詞+「与える」 → 使役
 - ・影響を与える → × 影響させる
 - ○ 影響する

<図5： 機能動詞結合の特徴>

作名詞を修飾していた文節を、新しい述語(動作名詞の用言形)への修飾として捉え直す必要がある。このとき、図5-(2)、(3)に示すように助詞の変換が必要となる。

- (3) 形式的には、機能動詞結合「動作名詞+機能動詞」の形をしているが、実際には文中において本動詞として働いているものや、機能動詞と結合する動作名詞の字面によっては、その機能動詞の持つ変換規則からはずれ、独自の交換規則を要求するものがある(図5-(4)、(5))。

4. 3. 2 解析方式の概要

機能動詞結合は、その特徴に応じて、動作名詞の用言形を述語とした文に変換して、その持つ結合価パターンを用いて解析するが、上記の問題点(1)は、変換の際に用いるべき規則を限定するものであり、問題点(2)は、変換する際の規則として捉えることができる。

ここでは、これらの情報を機能動詞結合制御テーブルに記述し、この情報を用いて入力文を変換する。機能動詞結合制御テーブルの記述内容は次の4点である。

- ・動作名詞の意味カテゴリ。
- ・動作名詞が自動詞化するか、他動詞化するか。
- ・助詞の変換情報。
- ・例外テーブル名(後述)。

問題点(3)については、機能動詞の字面だけでなく動作名詞の字面も定まっており、個々について変換規則は異なっている。従って、これらの例外に変換規則を与えるために、例外テーブルを用意する。例外テーブルには、各機能動詞ごと、各動作名詞ごとに以下の内容を記述する。

- ・本動詞として使われているかどうか。

- ・機能動詞の持つ文法的意味。
- ・助詞の変換規則。

機能動詞結合の解析の概要を機能動詞「受ける」を含む文「A国がB国の支配を受ける」を例にとって説明する。図6に示すように、「受ける」は複数の用法を持つため、まず、どの用法で使われているのかを決定する。例文中の機能動詞結合「支配を受ける」における動作名詞「支配」の用言形「支配する」は他動詞であるので、図6より文法的意味＝「受身」、助詞の変換「の→に」であることがわかる。また、「感銘を受ける」の場合には、動作名詞の用言形「感銘する」は自動詞であり、動作名詞としての意味カテゴリ＝「感情」であるので、文法的意味＝なし、助詞の変換「からの→に」の情報が得られる。

例文「A国がB国の支配を受ける」は、上記の情報によって、図6のように変換される。この後、文法的意味＝「受身」による表層格変換が施され、変換後の用言「支配する」の持つ結合価パターンとのマッチングが行われる[6]、[7]。

以上の解析により、入力となった構造を図6に示すように変換し、関係を付与する。このとき、もともとは名詞句の一部であった名詞を変換後には格要素として捉え直す。

構成名詞が複合名詞である場合にも、上記と同様にして解析することができる。但し、本解析の前に複合名詞を構成する語基間の関係を認定する必要がある[5]。

なお、構成名詞修飾語句が形容詞単独である場合には、

一律に変換後の文に対する副詞要素として捉え直す(例：大きな影響を受ける → 大きく影響される)。

5. おわりに

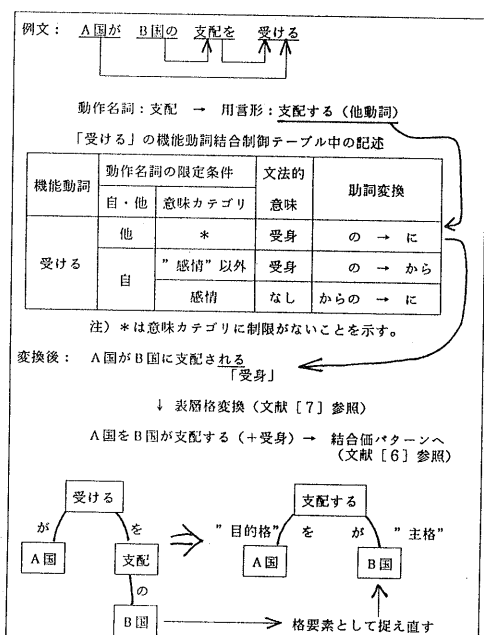
本稿では、1つの格要素と用言からなる述語相当語句とみなせる慣用表現を、構成語の意味と全体の意味との関係、および、構成語間の結合度に着目して分類し、各々についてその特徴を述べた。さらに、これらの特徴を使って慣用表現を含む日本文を解析する次の2つの方式を提案した。

両方式とも、結合価構造をもとに日本文を解析するものであるが、その前に制御テーブル中の種々の情報を用いて、その表現が文中で慣用表現として使われているかどうかを判定するという特徴を持っており、述語相当語句の慣用表現を含む文を正しく解析できる。

本方式では、制御テーブル中の情報は、ほぼ単文に閉じた範囲の情報のみを用いている。このため、深い意味解析や単文間にわたる情報を利用しなければ正しく解析できない文(例：彼は油を売ることを仕事にしている)についても、文中に慣用表現(例文については「油を売る」)が存在すると解析してしまい、正しい解析結果が得られないという問題がある。今後、上記の問題や表1において対象外とした慣用表現(構成名詞修飾語句として形容詞文や動詞文が存在するものなど)についても解析可能な方式について検討する予定である。

【参考文献】

- [1] 宮地裕編：「慣用句の意味と用法」明治書院(1982)
- [2] 村木新次郎：「慣用句・機能動詞結合・自由な語結合」日本語学 Vol.4 p.15 (1985)
- [3] 村木新次郎：「日本語の機能動詞表現をめぐって」国立国語研究所報告65研究報告集(2)(1980)
- [4] 長尾真他：「機械翻訳における訳語選択と構造変換過程」情報処理 Vol.26 No.11 p.1261 (1985)
- [5] 石崎雅人：「日本語複合名詞の解析」情報処理学会第35回全国大会 講演論文集 1T-1 (1987)
- [6] 林良彦：「結合価構造に基づく日本文解析」情報処理学会自然言語処理研究会資料 62-6 (1987)
- [7] 河合敦夫：「日英翻訳システムALT-J/Eにおける様相、時制の処理」情報処理学会第34回全国大会 講演論文集 5W-2 (1987)



<図6： 例文に対する機能動詞結合の解析の概要>