日英機械翻訳システム TWINTRAN の日英翻訳文法

西田収
シャープ(株) 情報商品開発研究所

Jiri Jelinek
元 シャープ(株)

吉見毅彦
シャープ(株)/ 神戸大学

田村直之
神戸大学大学院自然科学研究科

村上温夫
甲南大学理学部

1 はじめに

日本語の言語処理に必要な日本語文法に関し，文節モデルに基づいて大規模な文法が記述される一方で，理論言語学の側からJPSGなどの新しい枠組も提案されている．このような新しい枠組の中で実用に耐えられるような大規模で高品質な文法をいかにして記述するかが課題となっている．文法を系統立てて記述するためには，文法規則がどのようにして見つけ出されたのかという文法の構成過程を明らかにする必要があるが，そのような報告は少ない[6]．本稿では，文法を結合価モデルによる言語の分析法に基づいて構成する．これまでに，結合価モデルを日本語の動詞と助詞句の分析に応用した研究[1]などが報告されているが，従来の結合価による分析は動詞とその補語の関係の分析に止まるものが多かった．これに対し，本稿では，形態素，語，文など，あらゆるテキスト構成要素の分析を結合価モデルに従って行なう．すなわち，あらゆる構成要素が結合価を持ち，それが階層をなすとみなす．本稿のモデルでは，ある構成要素が持つ結合価を，それがとれる補語(その構成要素にとって必須あるいは可能である構成要素)の種類，個数，出現順序，および補語との結合の強さによって定義する．

自然言語処理システムは，機械翻訳や自然言語インターフェイスなど多様な分野に応用されるようになり，各分野毎に異なる性能が要求されてきている．あらゆる応用分野に適用できる汎用的な文法を近い将来に構築することは困難である．本稿の文法は，応用を日英機械翻訳に絞った文法であり，その使命は，日英機械翻訳における日本語解析に必要な情報を提供することにある．従って，あるテキスト構成要素を英訳する上でその構成要素が果たす役割そのものをその構成要素の機能とみなす．

応用を日英機械翻訳に限定しても，機械翻訳システムは多種多様な入力を処理する必要があるので，入力を解析するための文法は巨大で複雑なものにならざるを得ない．手間をできるだけかけずに，大規模な文法の開発や保守，改良を行なうためには，文法を簡潔に表現することが重要である．そこで本稿では，結合価モデルに基づいて分析した結果を記述する形式として，範疇文法の記法[7]を用いる．なお，本稿では範疇文法をただ表記法として利用しているのみで，日英機械翻訳システムTWINTRAN[3]の文法自体は範疇文法に関連する習慣，言語観を継承していない．また，文法のすべてを範疇文法の表記法で記述しているわけではない．現在，本文法を実装し，評価と改良を行なっている．

2 結合価モデルによる分析

結合価モデルによる言語の分析法は，Lees[5]らによって提案されており，本稿ではこれに従う．分析法の概要を以下に述べる．また，具体例は4節で挙げる．

品詞の分類には，構文上の置き換え可能性テストと意味上の置き換えテストの二種類を用いる．構文上の置き換え可能性テストは，あるテキスト構成要素を，その文法的な性質を変えないで他のどのような構成要素と置き換えられるかを調べるテストであり，これによって名詞，動詞などの品詞分類を行なう．また，意味上の置き換え可能性テストは，ある構成要素を，その意味を変えないで他のどのような構成要素と置き換えられるかを調べるテストであり，これによって助詞，助動詞などの機能分類を行なう．

あるテキスト構成要素がとれる補語の種類，出現順序，個数を規定する文法規則の作成には，連接可能性テスト，支配可能性テスト，拡張可能性テストを用いる．連接可能性テストは，ある構成要素がどのような構成要素に連接できるかを調べるテストである．支配可能性テストは，ある構成要素が他のどのような構成要素をその修飾語あるいは補語としてとれるかを調べるテストである．拡張可能性テストは，ある構成要素が別の複数の構成要素を同時にとれるかどうかを調べるテストである．

言語は最上位に意味を，最下位に文字(あるいは音声)を持つ体系であり，この体系は形態層，語彙層，構文層などのいくつかの層から成っている．これらの層の分析には，中断可能性テストを用いる．中断可能性テストとは，元の文の文法的正しさや意味，構成要素間の支配従属関係を変えずに，ある二つの構成要素の間にどのような構成要素を挿入できるかを調べるテストである．

3 日英対照分析

日英機械翻訳という応用のために，2節で述べたテストによる分析に加えて，英語との対照分析を行なう．具体的には，意味上の置き換え可能性テストを行なう際に，ある構成要素が，その意味を変えないで他の構成要素と置き換えられるかどうかを，それら二つの構成要素の英訳が同じになるかどうかによって判断する．また，何を語彙項目とするかは英訳を考慮して定める．すなわち，日本語だけを対象とするならば，別々の語彙項目として辞書に登録されるようなものでも，その英訳に照らし合わせて，まとめて登録した方がよい場合，本稿では

一つの語彙項目とみなす．例えば，「(NP) の電源が投入され」は分解せずに，このまま登録し，“SUBJ(NP) be ^ powered up” という英訳を与える[1]．このことから，TWINTRAN の文法は日本語文法ではなく日英翻訳文法であるとみなすほうが適切である．

4 文法記述

文法の記述には，$X \leftarrow Y + X\backslash Y$ のような範疇文法の記法を用いる．ここで，範疇 $X\backslash Y$ は直前の範疇 Y をとって範疇 X を作るような範疇を表す．$X\backslash Y$ のような範疇を導出範疇，X や Y のような範疇を基本範疇という．導出範疇 $X\backslash Y$ は基本範疇 Y を補語とすることを表しているので，このような記法を用いることによって，結合価が自然に表現できる．

4.1 範疇分類

2節で述べた置き換え可能性テストと中断可能性テストによって得られる基本範疇の分類 (13 種類) を表 1に，導出範疇の分類 (20 種類) を表 2にそれぞれ示す．

表 1: 基本範疇の分類

基本範疇	説明
TEXT	文章
SENT	文
VP	動詞句，形容詞句，形容動詞句
NP	名詞，名詞句
PP	助詞句
ATTR	連体詞，名詞修飾句，接頭辞
ATR2	形容動詞の語幹
ADV	副詞，副詞句
DEL	限定詞
QUANT	数量詞
NUM	数詞
ET	取り立て助詞
COOR	並列接続詞

範疇 VP は，学校文法でいう動詞，形容詞，形容動詞に相当する．NP は，助詞「が」が後に続くことのできるものである．ATTR は NP を修飾するものである．ADV と PP は共に VP を修飾するものであるが，前者は修飾できる VP が限られない (動詞型依存性がない) のに対し，後者は限られる (動詞型依存性がある)．ATR2 は，形容動詞の語幹である．ATR2 と NP との違いは，「が」や「を」が直後につくことができるかどうかである．例えば，「普段」という語は，「が」や「を」が直後につくことはできないので，ATR2 に分類される．これに対し，「普通」には「普通が」というの用法もあるので，NP と ATR2 の両方に分類する．また，「肯定応答」のような複合語は，通常，名詞「肯定」と名詞「応答」の合成として分析されるが，ここでは，英語との対照から「肯定」を ATTR に分類する．

範疇 NP\NP は前に NP を取って NP を作る範疇 (例えば，「数式処理」の「処理」など) である．VP\VP は前に VP を取って VP を作る範疇 (「遊ばせる」の「せる」など) である．従来の分類での助動詞に相当する．

[1]見出し語の ‘(NP)’ の部分は実際には登録されていない．英訳において，‘SUBJ(NP)’ は NP が主語になることを意味し，記号 ‘^’ はある種の副詞がこの位置に挿入されうることを意味する．

表 2: 導出範疇の分類

範疇類	導出範疇
VP 類	VP\VP,VP\NP,VP\ATR2,VP\SENT
PP 類	PP\NP,PP\VP,PP\SENT,PP\TEXT
NP 類	NP\NP,NP\VP,NP\SENT
ATTR 類	ATTR\ATR2,ATTR\NP,ATTR\SENT
ATR2 類	ATR2\NP,ATR2\VP
ADV 類	ADV\ATTR,ADV\ATR2,ADV\NP,ADV\SENT

4.2 句構造規則

範疇文法の記法を用いて文法を記述すると，いくつかの文法規則は自明である．すなわち，導出範疇 $X\backslash Y$ が存在すれば，$X \leftarrow Y + X\backslash Y$ という文法規則が存在する．ここで，範疇 VP を作る二つの規則 VP $\leftarrow$ VP + VP\VP と VP $\leftarrow$ SENT + VP\SENT を比較する．前者は，「走ら + せる」のような構造を表す規則であり，後者は，「彼が生きている + か分かりません」のような構造を表す規則である．両者の違いは，VP\VP が VP を一つだけしか取らないのに対し，後者は二つ以上の VP を含むような SENT を補語に取れることである．例えば，「歩き走らせる」の「せる」のような VP\VP は「走る」のみを修飾する．一方，「歩き走ることができる」の「ことができる」のような VP\SENT は，「歩く」と「走る」の両方の VP を修飾する．このような分析は，2節で述べた拡張可能性テストを用いて行なう．

4.3 動詞に対する補語の機能分類

動詞に対する補語の機能 (格) を分類するために，補語に意味上の置き換え可能性テストを適用する．意味が変わるかどうかは，同じ英訳になるかどうかで判断する．例えば，「京都へ行け．」における助詞「へ」と「京都に行け．」における助詞「に」は，共に “Go to Kyoto.” のように，前置詞 “to” に英訳することができるので，動詞「行く」に対して同じ機能を持つ．このようにして分類した VP の補語の機能を表 3に示す．なお，ここで行なった補語の機能分類と 4.4節で述べる動詞型の分類は相互に依存するものであり，一方の分類が完了した後に他方の分類ができるものではない [2].

本稿での補語の機能分類は，深層ではなく表層での分類である．表層で分類を行なう理由は，日英機械翻訳システムの日本語解析に必要な情報を提供することを目的とする限りにおいては，表層表現をより深い層へ写像する必要はないからである．例えば，「先生が生徒を誉める．」と「生徒が先生に誉められる．」という二つの文において，深層では，前者に現れる助詞「が」は動作主，後者に現れる「が」は被動作主という異なる機能分類になる．これに対し本稿では，「先生が」も「生徒が」も英語で主語となることから，共に SUBJ という機能分類とする．

4.4 動詞の結合価 (動詞型) の分類

動詞がどのような補語をとれるかによって分類した動詞型の一覧を表4に示す．表4は，例えば，動詞型19の動詞がSUBJ，AI，PSTをそれぞれ高々一つとれることを表している．また，動詞型1，12，15，17には補語の集合が二通りずつある．

表 3: 動詞の補語の機能分類

機能分類	説明
AG	日本語では助詞「に」や「によって」で表され，英語では前置詞 "by" で表される．使役態の動詞がこの機能の補語をとると，この補語は英語では目的語になることがある．動詞型 (表4) が12または15である動詞がこの機能の補語をとると，この補語は英語では主語になることがある．
AI	日本語では助詞「で」や「によって」で表され，英語では前置詞 "by means of" や "by"，"with" で表される．
CP	日本語では助詞「と」で表され，英語では前置詞 "with" で表される．ここで，"with" は道具や付加ではなく，同伴を意味する．
IO	間接目的語．日本語では助詞「に」で表され，英語では前置詞 "to" や "for"，"at" で表される．
LT	日本語では助詞「まで」で表され，英語では前置詞 "until" や "up to" で表される．
ML	日本語では助詞「までに」で表され，英語では前置詞 "by" で表される．
OBJ	直接目的語
PAC	日本語では助詞「で」や「において」を，英語では前置詞 "in" や "on"，"at" を場所名詞に付加した形で表される．
PST	日本語では助詞「に」や「において」を，英語では前置詞 "in" や "on"，"at" を場所名詞に付加した形で表される．
PTR	日本語では助詞「を」を，英語では前置詞 "over" や "through"，"across" を場所名詞に付加した形で表される．
QA	変化を表す動詞 (動詞型7，8) によってもたらされる物事や特性を表す．例えば，「彼は医者になった．」は "He became a doctor." と英訳される．
QO	引用．日本語では助詞「と」で表される．
SOBJ	動詞が動詞型12，15または一部の動詞型の可能態である場合，日本語では助詞「が」で表され，英語では目的語になる．例えば，「鳥 (に) は羽がある．」は "Birds have feather." と英訳される．
SUBJ	主語．
TG	目標．日本語では助詞「に」や「へ」で表され，英語では前置詞 "to" や "for"，"at" で表される．

表 4: 動詞型

動詞型	補語	例
1	SUBJ,OBJ,QA,CP,AI,PAC,LT,ML	思う
	SUBJ,OBJ,QO,TG,CP,AI,PAC,LT,ML	
2	SUBJ,OBJ,QO,TG,CP,AI,PAC,LT,ML	書く
3	PST,LT,QO	ある ("it says" "it reads")
4	SUBJ,OBJ,IO,CP,AI,PAC,LT,ML	探す
5	SUBJ,OBJ,TG,CP,AI,PAC,LT,ML	植える
6	SUBJ,OBJ,CP,AI,PAC,LT,ML	救す
7	SUBJ,QA,CP,AI,PAC,LT,ML	終る
8	SUBJ,OBJ,QA,CP,AI,PAC,LT,ML	削る
9	SUBJ,TG,CP,AI,PAC,LT,ML	急ぐ
10	SUBJ,IO,CP,AI,PAC,LT,ML	違う
11	SUBJ,CP,AI,PAC,LT,ML	遊ぶ
12	SUBJ,SOBJ,CP,AI,PST,LT,ML	ある ("have")
	SOBJ,AG,CP,AI,PST,LT,ML	
13	SUBJ,IO,CP,AI,PST,LT,ML	留まる
14	SUBJ,CP,AI,PST,LT,ML	ある ("there be")
15	SUBJ,SOBJ,IO,CP,AI,PST,PAC,LT,ML	助かる
	SOBJ,AG,IO,CP,AI,PST,PAC,LT,ML	
16	SUBJ,PTR,TG,CP,AI,PAC,LT,ML	走る
17	SUBJ,IO,CP,AI,PAC,LT,ML	青い
	SUBJ,AI,PST	
18	SUBJ,AI,PST,QA,PAC,LT	等しい
19	SUBJ,AI,PST	嫌だ

4.5 態の変化

動詞の態 (動詞型) は，「させる」などの助動詞や「てもらう」などのある種の補助動詞が接続することによって変化する．本稿では，支配可能性テストの結果に基づき，能動態，受動態 (pas)，使役態 (caus)，間接受動態 (pin)，可能態 (pot) を区別する．表4に示した動詞型は，動詞が能動態である場合の分類である．表4の各動詞型がどのような態に変化でき，変化した態ではどのような補語をとれるかを表5に示す．表5において，例えば1-causは動詞型1の使役態を意味する．

4.6 層の分類

文を構成する層は語彙層，形態層，構文層に分けられる．層の区別は中断可能性テストによって行なう．動詞と補語の結合力は動詞が属する層によって異なり，動詞が形態層に属する場合の方が，構文層に属する場合よりも強い．従って，形態層の動詞の補語は省略できないが，構文層の動詞の補語は，省略が可能であり，出

表 5: 動詞の態の変化

動詞型	補語
1-caus	SUBJ,OBJ,AG,QO,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,AG,QA,CP,AI,PAC,LT,ML
1-pas	SUBJ,AG,QO,TG,CP,AI,PAC,LT,ML
	SUBJ,AG,QA,CP,AI,PAC,LT,ML
1-pin	SUBJ,OBJ,QO,AG,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,QA,AG,CP,AI,PAC,LT,ML
1-pot	SUBJ,SOBJ,QO,TG,CP,AI,PAC,LT,ML
	SOBJ,AG,QO,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,QO,TG,CP,AI,PAC,LT,ML
	SUBJ,SOBJ,QA,CP,AI,PAC,LT,ML
	SOBJ,AG,QA,CP,AI,PAC,LT,ML
	SUBJ,OBJ,QA,CP,AI,PAC,LT,ML
2-caus	SUBJ,OBJ,AG,QO,TG,CP,AI,PAC,LT,ML
2-pas	SUBJ,AG,QO,TG,CP,AI,PAC,LT,ML
2-pin	SUBJ,OBJ,QO,AG,TG,CP,AI,PAC,LT,ML
2-pot	SUBJ,SOBJ,QO,TG,CP,AI,PAC,LT,ML
	SOBJ,AG,QO,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,QO,TG,CP,AI,PAC,LT,ML
4-caus	SUBJ,OBJ,AG,IO,CP,AI,PAC,LT,ML
4-pas	SUBJ,AG,IO,CP,AI,PAC,LT,ML
4-pin	SUBJ,OBJ,IO,AG,CP,AI,PAC,LT,ML
4-pot	SUBJ,SOBJ,IO,CP,AI,PAC,LT,ML
	SOBJ,AG,IO,CP,AI,PAC,LT,ML
	SUBJ,OBJ,IO,CP,AI,PAC,LT,ML
5-caus	SUBJ,OBJ,AG,TG,CP,AI,PAC,LT,ML
5-pas	SUBJ,AG,TG,CP,AI,PAC,LT,ML
5-pin	SUBJ,OBJ,AG,TG,CP,AI,PAC,LT,ML
5-pot	SUBJ,SOBJ,TG,CP,AI,PAC,LT,ML
	SOBJ,AG,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,TG,CP,AI,PAC,LT,ML
6-caus	SUBJ,OBJ,AG,CP,AI,PAC,LT,ML
6-pas	SUBJ,AG,CP,AI,PAC,LT,ML
6-pin	SUBJ,OBJ,AG,CP,AI,PAC,LT,ML
6-pot	SUBJ,SOBJ,CP,AI,PAC,LT,ML
	SOBJ,AG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,CP,AI,PAC,LT,ML
7-caus	SUBJ,OBJ,QA,AG,CP,AI,PAC,LT,ML
7-pot	SUBJ,QA,CP,AI,PAC,LT,ML
7-pin	SUBJ,AG,QA,CP,AI,PAC,LT,ML
7-caus	SUBJ,OBJ,AG,QA,CP,AI,PAC,LT,ML
8-pas	SUBJ,AG,QA,CP,AI,PAC,LT,ML
8-pin	SUBJ,OBJ,QA,AG,CP,AI,PAC,LT,ML
8-pot	SUBJ,SOBJ,QA,CP,AI,PAC,LT,ML
	SOBJ,AG,QA,CP,AI,PAC,LT,ML
	SUBJ,OBJ,QA,CP,AI,PAC,LT,ML
9-caus	SUBJ,OBJ,TG,AG,CP,AI,PAC,LT,ML
9-pot	SUBJ,AG,TG,CP,AI,PAC,LT,ML
9-pin	SUBJ,AG,TG,CP,AI,PAC,LT,ML
10-caus	SUBJ,OBJ,IO,AG,CP,AI,PAC,LT,ML
10-pot	SUBJ,AG,IO,CP,AI,PAC,LT,ML
10-pin	SUBJ,AG,IO,CP,AI,PAC,LT,ML
11-caus	SUBJ,OBJ,CP,AG,AI,PAC,LT,ML
11-pin	SUBJ,OBJ,CP,AG,AI,PAC,LT,ML
15-pot	SUBJ,SOBJ,IO,CP,AI,PST,PAC,LT,ML
	SOBJ,AG,IO,CP,AI,PST,PAC,LT,ML
16-caus	SUBJ,PTR,AG,TG,CP,AI,PAC,LT,ML
	SUBJ,OBJ,AG,TG,CP,AI,PAC,LT,ML
16-pas	SUBJ,AG,TG,CP,AI,PAC,LT,ML
16-pin	SUBJ,PTR,AG,TG,CP,AI,PAC,LT,ML
16-pot	SUBJ,SOBJ,TG,CP,AI,PAC,LT,ML
	SOBJ,AG,TG,CP,AI,PAC,LT,ML
	SUBJ,PTR,TG,CP,AI,PAC,LT,ML

現順序もほぼ自由である．「である」や「だ」のようなコピュラ文を作る動詞は，構文層と形態層の二種類の補語を同時にとる．「A は B である．」という文において，「である」は構文層の補語「A は」と形態層の補語「B」をとる．構文層の補語「A は」は省略可能であり，「A は」と「である」の間に別の語句を挿入することができる．これに対し，形態層の補語「B」は省略することができず，「B」と「である」の間に他のテキスト構成要素を挿入することもできない．

5 おわりに

本稿の日英翻訳文法には，次のような特徴がある．

1. 結合価モデルに基づく体系的な文法である．
2. 日英機械翻訳のための応用指向の文法である．
3. 範疇文法の記法を用いて簡潔に表現されている．

これまでに文節モデルに基づいて大規模な文法がいくつか記述されている．文節モデルでは，結合価を表すための枠組みを持たないため，ある文節が別の文節に係るかどうかを決めるために，非交差などの規則以外は意味に頼る必要がある．また，文節の係り受けによる構造は，「意味に素直でない」と言われる．例えば，「桜の花が咲く」という文を文節モデルに基づいて分析すると「桜の」が「花が」に係ることになる [4]．一方，結合価モデルから見ると，「が」が「桜の花」を補語に取るという分析になる．後者の方が文の意味を素直に表しており，意味の合成や別の言語への翻訳を行う場合には，結合価に基づく分析を用いる方が処理が容易になる．

参考文献

[1] 石綿敏雄, 荻野孝野. 結合価から見た日本文法. 水谷静夫, 石綿敏雄, 荻野孝野, 寿来直子, 草薙裕（編）, 文法と意味 I, 朝倉日本語新講座 3, pp. 81–134. 大修館書店, 1974.

[2] J. Jelinek. Distributional and Functional Analysis of Japanese Verbs. In Rosemary, J. Yates, editor, *Sheffield Studies in Japanese: 1*, Scientific and Technical Japanese Series 031, pp. 102–127. Centre of Japanese Studies, University of Sheffield, 1979.

[3] J. Jelinek, O. Nishida, T. Yoshimi, N. Tamura, and H. Murakami. TWINTRAN: A Japanese-to-English Machine Translation System based on Text-Wide Grammar. PROLOG 産業応用シンポジウム論文集, pp. 59–63, 1992.

[4] 小池清治. 大学生のための日本語文法. 有精堂出版, 1987.

[5] R. B. Lees. The Grammar of English Nominalisations. *International Journal of American Linguistics*, Vol. 26, No. 3, 1960.

[6] 佐野洋, 福本文代. 日本語の述部階層構造に基づく形態論的な文法規則の記述法. 自然言語処理, Vol. 3, No. 3, pp. 3–29, 1996.

[7] H. Uszkoreit. Categorial Unification Grammar. In *Proceedings of the 11th International Conference on Computational Linguistics (COLING)*, pp. 187–194, 1986.