自 然 言 語 処 理 103−16
（1994．9．16）

日本語構文解析システム「KNP」のハングル化とそれを用いた日本語から韓国語への対照分析

黄 道三　　黒橋 禎夫　　長尾 眞
京都大学工学部 電気工学第二教室

要旨

実際文では，非文ともいえるほどの文も使われる場合があり，文を限定されたルールによって解析するには限界がある．韓国語は語順が自由で，不規則が多い言語であるため，決まったルールでは解析できない場合が多い．最近，用例パターンとの類似性に基づいて日本語文を解析するシステム「KNP」が開発され，構文解析において96%という高い成功率を表している．これは，長い日本語文には並列構造になっている文節が多いので，これらを先にまとめることによって，長い文を簡単な構造に解析することができるという考え方に基づいている．日本語と韓国語とは構文構造と語彙形成形態とにおいて似ているので，KNPは韓国語にもうまく働くと考え，KNPのハングル化($hKNP$)を試みた．330文に対して実験を行なったところ，韓国語の長い文にも十分に有効であることがわかった．また，日本語180文とそれを訳した韓国語文を各々KNPと$hKNP$で構文解析し，日本語から韓国語への対照分析を行なった結果，すべての文において文節の順序が一致しており，74文(41%)の構文構造が完全に一致していることがわかった．

A Korean version of KNP and a Contrastive Study of Japanese and Korean Languages

Dosam Hwang　　Sadao Kurohashi　　Makoto Nagao
Department of Electrical Engineering, Kyoto University

Abstract

There is a limitation to the analysis of sentences using a restricted grammar rule, because ungrammatical sentences are often found in practical usage. Moreover, it is difficult to analyze Korean sentences, because Korean has flexible word order, and has many irregular rules in grammar. Japanese Analysis System, KNP, is presented and shows 96% in success rate. KNP is based on Analogy. As many parallel phrase structures usually exist in long Japanese sentences, KNP analyzes long sentences effectively and correctly by simplizing them in advance using analogy. Korean $KNP(hKNP)$ has been developed successfully by simply modifying KNP because of many common grammatical characteristics in Japanese and Korean. We present and evaluate some analysis results of $hKNP$ for 330 Korean sentences. Using KNP and $hKNP$, a contrastive analysis for 180 Japanese sentences and their Korean translations is also shown in this paper.

1 はじめに

多くの韓国語構文解析は文法記述において制約された文法規則に頼って研究が行なわれてきた．しかし，非文ともいえるほどの文も実際使われる場合があり，限定されたルールによって文を解析するには限界がある．韓国語は語順が自由で，不規則が多い言語であるため，決まったルールでは解析できない場合が多い．

最近，類似性に基づいて日本語文を解析するシステム「KNP」が開発され，構文解析において96%という高い正解率を表している．KNPは，文節間の類似度に基づいて並列構造の範囲を見つけ出し，簡単なヒューリスティックスを用いて依存構造を求める．そして，フレームの辞書を用いて格構造を評価することにより妥当な依存構造を選ぶという方法に基づいている[1].

これは，長い日本語文には並列構造になっている文節が多いので，これらを先にまとめることによって長い文を簡単な構造に解析することができるという考え方に着目している．韓国語文も長い文になればなるほど多くの並列構造を含んでいるので，局所的な解析だけでは長い文に対して精度のよい解析結果が得られることが期待できない．日本語と韓国語(日韓語)とは構文構造と語彙形成形態とにおいて似ているので，KNPは韓国語にもうまく働くだろうと考え，KNPのハングル化を試みた．類似度を求めるとき用いられるシソーラスも既に構築しているので[2]，KNPに対して簡単な変更を行なうだけでハングルKNP($hKNP$)を構築することができた．330文に対して実験を行なったところ，95%の正解率を示した．構文解析の結果は木構造の形で出てくるので，人目でも文の構造を便利に把握することができる．それで，日本語180文とそれを訳した韓国語文を各々KNPと$hKNP$を通して構文解析し，日本語から韓国語への対照分析を行なった．

2 $hKNP$の概要

KNPは形態素解析，並列構造解析，係り受け解析，格構造解析の4段階を通して文を解析する．格構造解析には格フレーム辞書を用いるが，韓国語の格フレーム辞書が開発されていないので係り受け解析までハングル化を行なった．

形態素解析システムは韓国科学技術院で開発したハングルtaggerを用い，類似度計算のために用いるシソーラスは「分類語彙表」の分類体系に基づいて作成した韓国語シソーラス[2]を利用した．日韓語は構文構成と単語形成において似ているので，KNPのハングル化に際してKNPの解析手順やアルゴリズムを変更する必要はなかった．しかし，単語形成の形態において異なっている部分のため，文節分割規則と係り受け規則を変更し，$hKNP$は図1のように構成した．斜線の部分がKNPのハングル化において変更した部分である．

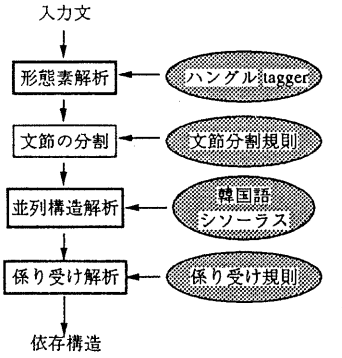


図1: $hKNP$の構成

2.1 文節の分割

形態素解析によって入力韓国語文を形態素単位に分割し，各形態素の品詞，さらに活用する語の場合には原形を決定する．本研究では構文解析に関心があるので，形態素解析結果に誤りのある部分は人手で書き直した．次に分割された単語列を文節[*1](自立語とそれに続く付属語)にまとめる．ここでは，意味的まとまりという観点から「本用言＋語尾＋補助動詞」はひとつの文節にまとめた．また，「할 수 있다(することができるorできる)」，「에 관하다(に関する)」，「에 있어서(において)」などは複数の文節になっているが，意味的まとまりとみてひとつの文節にまとめた．

2.2 並列構造解析

2.2.1 並列構造の推定

1文中の並列する部分は何らかの意味において類似していることが多い．そこで，類似性という観点から並列構造を検出する．

文節間の類似度の計算

文節間の類似度を品詞の一致，文字列の一致，シソーラスによる意味的な近さなどによって定義し，全ての文節間について計算する．図2の三角行列の(i,j)要素の数字はi番目とj番目の文節の類似度のポイントを示す．

並列キーの設定

並列キーはKNPと同じ基準にしたがって名詞並列，

[*1] 韓国語では語節または句に対応する．

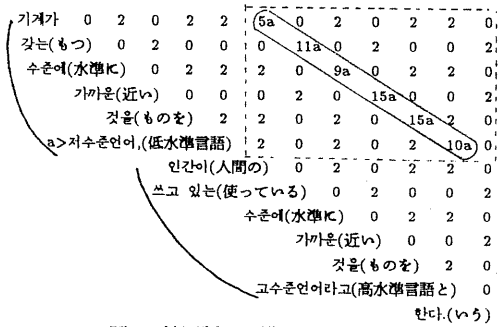


図 2: 韓国語並列構造の推定の例

表 1: 並列キーを示す付属語

名詞並列	,와(と) 도(も) 나(や) 랑(や) 든가(か)
	인지(か) 라든가(とか) 동시에(かつ)
	만(이) +아니라(だけ (で) はなく)
	및(および) 또는(または) 혹은(あるいは)
述語並列	것에 +대해서(の+に対して) 든가(とか) 고(し) 며(し)
	나(が) 지만(が) 데(が) 거늘(が) 되(が)
	뿐만(이) +아니라(だけ (では) なく) 및(および)
	여(して) 서(して) 또는(または) 혹은(もしくは)
部分並列	및(および) 또는(または) 혹은(もしくは)

述語並列，部分並列に定義し，並列構造の存在を示す文節を見つけるのに使われる韓国語の付属語を表1のようにまとめた．KNP での付属語に対応する韓国語の付属語を与えたが，「や，し，が」などについては各々複数の韓国語の付属語が対応しているので，それらすべてを並列キーを示す付属語として扱った．

また，*KNP* では「して」を並列キーとして扱わなかったが，hKNP では「して」に対応する「하여，해서」に対して並列キーとして扱った．これは，「하여，해서」の後ろに複数の用言があるとき，その係り先を決めるのは難しい．ところが，「하여，해서」は原因・結果を表す場合が多く，そのとき対応する文節は類似性をもっているので，並列キーとして定義することによって本並列構造の抽出方法でうまく解析されると考えた．しかし，「名詞+에 대해(서)(に対して)」，「名詞+에 의해(서)(によって)」のように名詞句の付属語のときには並列キーとして扱わなかった．

並列構造の範囲の推定

そして，並列キーの前後で，類似度の総和が最も大きい文節列の組を求める．これは，図2の点線の範囲内において，一番下の行の 1 つの要素から出発して点線の範囲内の一番左の要素までの左上方向への要素の並び (パス) の中でポイントの和が最も大きいパスを求めることに対応する．この処理は DP マッチングの手法で行なう．そのようなパスが求まれば，そのパスの左側の対応する文節列と下側の対応する文節列が並列であると推定する．図2の例文では，実線で囲まれた要素の並びが最高得点を得るパスで，このパスに対応する 2 つの文節列，「기계의(機械の) ～저수준언어(低水準言語)，」と「인간의(人間の) ～고수준언어라고(高水準言語と)」が並列であると推定する．

2.2.2 並列構造の整理

長い韓国語文には 1 文中に複数の並列構造が複雑な関係で存在することが多いので，前節の処理で得られた並列構造に関する情報を以下のようにまとめることにより，文を簡単化する．

兄弟関係の整理 兄弟関係にある並列構造を 1 つの並列構造にまとめる (図 3(a))．

親子関係の整理 範囲的に親を複数もつ並列構造について親同士の関係を調べることにより直接の親を決める ((図 3(b))．

違反関係 正しい並列構造の重なりではなく，並列構造の検出の誤りによって生じたものであり，並列構造の再検出を行なう．((図 3(c))．

これらの処理の具体例を図 3(a) の例文 (以後，例文 1 とよぶ) について示す．この例文の場合，前節の処理で次の 3 つの並列構造が検出される．

- CS1 [표제(表題)，] - [저자(著者)，]
- CS2 [저자(著者)，] - [주제등의 主題などの]
- CS3 [주제등의(主題などの) ～재편성하고(再編成し)，] - [색인(索引) ～기록하여 둔다(記録しておく)．]

このうち CS1 と CS2 は兄弟関係にあるので，これらを 3 つの部分からなる 1 つの並列構造 CS'([표제(表題)，] - [저자(著者)，] - [주제등의(主題などの]) にまとめる．さらに，CS2 と CS3，すなわち CS' と CS3 が親子関係にあるので，CS3 の前部を「표제(表題)，저자(著者)，」の部分を含むように延長する．これらの処理の結果，図 4に示すように，文の大まかな構造を得ることができる．

2.3 係り受け解析

前章で説明したように，文内の並列構造が決定されると文の大まかな構造がとらえられたことになり，文を簡単化した形で扱うことが可能となる．この結果，簡単化された各部分に対して単純な係り受け解析をするだけで，従来の構文解析よりも頑強な，すなわち長い文に対しても有効な構文解析を行なうことができる．

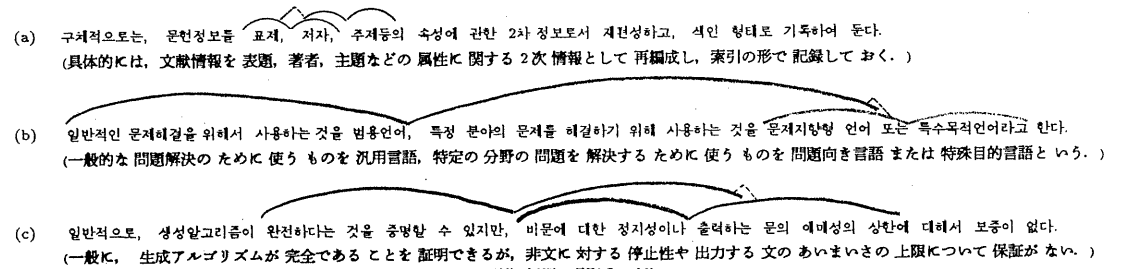


図 3: 並列構造間の関係の例

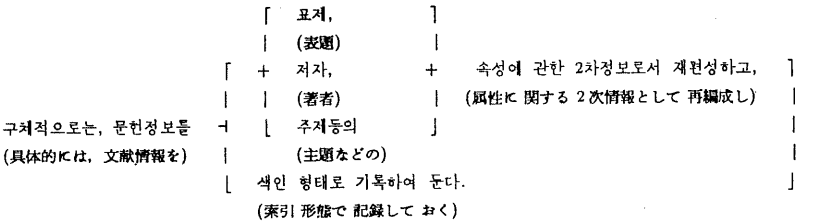


図 4: 文の簡単化

2.3.1 係り受け解析の概要

依存構造木では，木の節点には受けの文節を，その子の節点には係りの文節を配置する．韓国語の場合も，各文節はそれよりも後ろの文節に係るので，最終的には文末の文節を根節点とする依存構造木が得られる．

係り受け解析は，まず並列構造に関する部分について行なう．複数の並列構造が入れ子構造 (親子関係) をなしている場合は，その中の最も深いレベルにある並列構造から順に解析していく．そして，最後に文全体の解析を行なう．

並列構造の前部と後部はそれぞれ意味的なまとまりをなしており，それぞれを 1 つの依存構造木にまとめることができる．そこで，並列構造を表すためにその 2 つの依存構造木の根節点 (依存構造木の表示では "<P>" を付加) を子の節点とする新たな節点 (依存構造木の表示では "PARA" と表記，並列節点とよぶことにする) を作り，その上のレベル (親の並列構造，それがない場合には文全体) の係り受け解析では，この並列節点のみが扱われるようにする．並列節点は，係りの文節となる場合には終点文節の内容を継承し，受けの文節となる場合には並列キーと終点文節の内容を継承する．

たとえば図 4に挙げた例文 1 では，まず並列構造 [표제, 저자, 주제등의(表題，著者，主題などの)] の各部分が処理対象となる．この各部分は 1 文節なので係り受け解析の必要はなく，単にそれらをまとめた並列節点が作

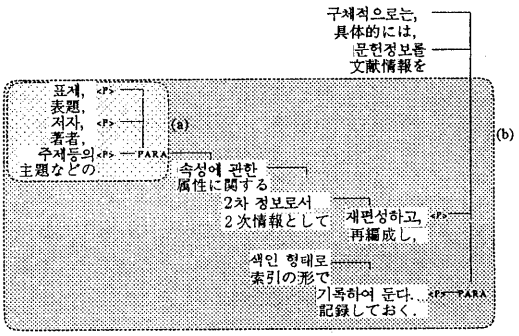


図 5: 係り受け解析の例 (例文 1)

られる (図 5(a))．次に，並列構造 [표제(표題), ～기억하여 둔다(記憶して おく).] の前部と後部を係り受け解析してそれぞれを依存構造木にまとめ，さらに並列節点を作る (図 5(b))．そして最後に文全体を係り受け解析し，文全体の依存構造木を得る (図 5)．

2.3.2 一定範囲内の文節列の係り受け解析

一定範囲内の文節列に対して次のような優先規則によって決定論的に動作する解析を行なう．

優先規則による決定論的解析

- 解析すべき範囲について後ろから順に各文節の係り先を決定する．
- 各文節の係り先は，係り受け関係が相互に交差しない条件 (非交差条件) をみたし，かつ係り得る文節のうち

表 2: 優先係り受け規則

係り先の文節	係っていく文節
名詞	冠形詞(連体詞), 助詞「의(の)」, 活用語の冠形形(基本形・タ形)
動詞	活用語の副詞形・従属法叙述形(連用形), 副詞, 助詞
叙述形の形容詞・形容名詞*2 (イ・ナ形容詞)	活用語の副詞形・従属法叙述形(連用形), 副詞, 助詞(「을(を)」・「에(へ)」以外)
冠形形の形容詞・形容名詞 (連体形のイ・ナ形容詞)	副詞, 助詞「가(が)」・「의(の)」・「에(に)」・「보다(より)」
副詞形・従属法叙述形の形容詞・形容名詞 (連用形のイ・ナ形容詞)	副詞
名詞+指定詞(判定詞)	活用語の副詞形・従属法叙述形(連用形), 副詞, 助詞(을(を)以外)

*2「하다(だ)」がついて動詞または形容詞になる単語.

最も近い文節とする(係り得る関係は表2に示す).

ただし, 助詞「가(が)」・「를(を)」を伴う文節は, それぞれ同じ述語に2つ以上は係り得ない.

以下の3つのタイプの文節に対しては, 特別の規則を用いる.

1. 区切り文節に対する規則

条件を示す述語(「~하면(すれば), 」など)と因果関係を示す述語(「~하기 때문에(するので), 」など)が読点を伴っている場合, この種の述語を区切り文節とし, その係り先は係り得る最も近い文節ではなく最も遠い文節とする.

2. 読点と伴う文節に対する規則

読点を伴う文節は, 係り得る最初の文節を飛び越して2番目の文節に係るとする. ただし, 次の場合にはこの規則は適用しない.

- 係り得る最初の文節が上述の区切り文節である場合.
- 「… ~(한, 인) 것이 많다. ((する, である)ことが多い.)」, 「… ~(이다, 하다) 고 생각한다. ((である)と考える.)」などの表現で下線部分が係り得る最初の文節である場合.
- 係り得る最初の文節が並列構造をまとめた並列節点の文節である場合.
- 係り得る文節が解析対象範囲内に1つしかない場合.

3. 助詞「는(は)」を伴う文節に対する規則

助詞「는(は)」は文を伴う文節は次のような規則を用いることにした.

表 3: 韓国語文の係り受け解析結果の評価

		出現数	正解	誤り	正解率
韓国語文1	並列文節	309	265	44	86%
	用言に係る文節	1208	1156	52	96%
	体言に係る文節	859	845	14	98%
	合計	2376	2266	110	95%
韓国語文2	並列文節	148	131	17	89%
	用言に係る文節	1067	1042	25	98%
	体言に係る文節	254	222	32	87%
	合計	1469	1395	74	95%
総計		3845	3661	184	95%

- 助詞「는(は)」を伴い読点を伴わない場合:
  最も近い区切り文節に係るとする.
- 助詞「는(は)」と読点を伴う場合:
  最も遠い区切り文節に係るとする. ただし, その文節に付属語が全て同じである文節が係っていて, かつ, その前にも区切り文節が存在する場合には前の区切り文節に係るとする.

3 実験結果と評価

実験は, 岩波情報科学辞典, 日本科学技術情報センタ(JICST)発行の抄録文, サイエンス(Vol.17,No.12「科学技術のためのコンピューター」), の各60文を韓国語に訳した文(韓国語文1)と韓国語の新聞, 文法自習書, 論文抄録の各50文(韓国語文2), 合計330文に対して行なった.

3.1 定量的評価

330文の解析結果について人手で評価を行なった. 文節を, 並列文節(並列キーを含んでいる文節), 用言に係る文節と体言に係る文節に分類し, 並列キーについては正しい終点文節が検出されているか, 各々正しい係り先が求まっているかを調べた(表3). 韓国語文1と韓国語文2に対して各々文節単位で95%の正解率を表している. 1文内の全ての並列構造と係り受け関係が正しく解析され, 文全体に対して正しい依存構造木が求まったものは330文中207文(正解率63%)であった. 韓国語では読点があまり使われないので, 2.3.2項の「区切り文節に対する規則」と「読点を伴う文節に対する規則」が適切に働かず, KNPの正解率(66%)より若干低い正解率となった.

3.2 解析の誤り

KNPで起こった以下の解析エラーはそのまま$hKNP$でも出てきた.

表 4: 解析の失敗例

例文 1:
...『「해석과정의 제어가 자유로운 액티브차트해석법과 단일화에 의거한 어휘통사적 문법적인 틀이라고 할 수 있는」HPSG』를 채용하고 있다.
...『「解析過程の制御が自由な アクティブ・チャート解析法と 単一化に基づいた語い・統辞的な文法的 枠組みである」HPSGを』採用している.
例文 2:
실제로, 필자등은「이것을 사용하여 중력상호작용이 지배하는」천체의 운동에 대해 고정도의 고속수치계산을 할 수있는 디지탈 오레리라고 하는 전용컴퓨터를 제작하고 있다.』
実際, 筆者たちは『「これを 使って, 重力相互作用が 支配する」天体の運動について, 高精度で高速の数値計算ができるディジタル・オレリーという専用コンピューターを製作している. 』
例文 3:
「다른『사회나 민족을』 구별하는 중요한 특질이」 된다.
「異なる『社会や 民族』を 区別する 重要な 特質と」 なる.
例文 4:
『어느 나라 어느 정부에도』
どんな 国 どんな 政府にも ...

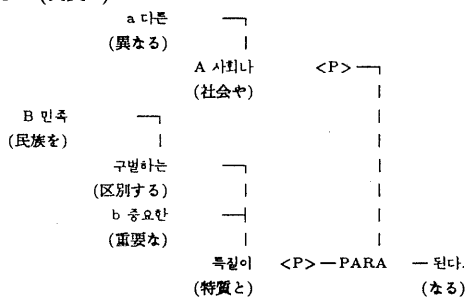
- 並列構造の範囲推定の誤り
- 述語的働きをするハダ名詞 (サ変名詞) の未処理による誤り
- 連体修飾の連続する部分での誤り

日本語では, 少し離れた文節に係っていくことを示すとき, 読点を多く使っているが, 韓国語では, 読点をあまり使っていないので, これに関する文節に解析の誤りが多く起こった. 誤って係り受け解析した 184 文節に接続詞または副詞節の後ろの読点の未使用によるものが 45 文節であるので, これらに対し解決できれば文節の正解率が 96% に, 文の正解率は 72% に上がる.

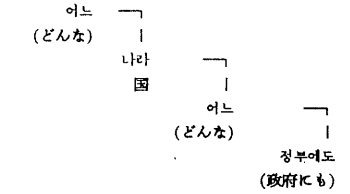
係り受け解析に失敗した具体例を以下に示す. 下線部分は並列キーの文節および誤って推定した範囲においてそれと並列する文節, 「」が誤って推定した範囲, 『』が正しい範囲である.

- 並列構造の範囲推定の誤りによるもの.

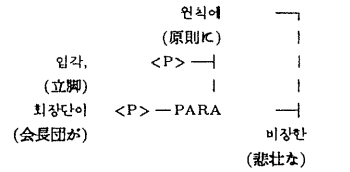
1. シソーラスへの専門用語の未登録語による誤り
   現在, 意味的類似性は分類語彙表のみを用いて与えているが, これに加えて専門用語のシソーラスなどが利用できれば並列構造推定の精度が向上すると考えられる. 例文 1 では「액티브차트해석법(アクティブ・チャート解析法)」と「HPSG」の類似度により大きなポイントが与えられれば正しい構造が推定できる.
2. 文節長のバランスがとれていない並列構造においての誤り
   並列キーが文のはじめであればあるほど, その後ろに並列のキーの文節と対応する可能性のある文節がたくさんあるので, 並列構造の推定は難しくなる. 例えば例文 2 のように文のはじめの連用中止が文末の述語と対応するような場合の解析は, バランスのとれた並列構造でないために非常に困難である.
3. 付属語の類似度
   下の図で a と b 文節に同一の付属語 (ㄴ) があるため, その文節間に高い類似度が与えられ, 正しい並列構造の範囲推定に失敗した. それによって誤った依存構造を生成している. A の「사회나(社会や)」と B の「민족(民族を)」とが並列構造になるのが正しい (例文 3).



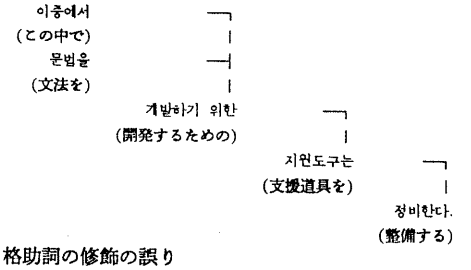
4. 読点なしで名詞が並べている句に対しては並列構造の推定ができなかった (例文 4).



- 下の図で「입각(立脚)」は用言の役割をしているが, 名詞として認識されて, 「원칙에(原則に)」が「비장한(悲壮な)」に係っている. これは, 例 2 と同じ理由から並列構造の推定にも失敗している.

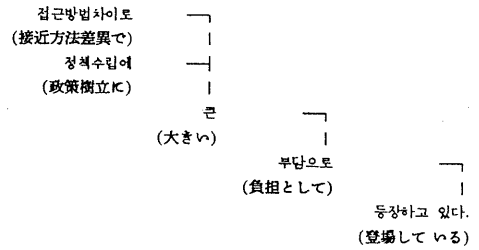


- 日本語では接続詞と副詞節の後ろに読点を打って最後の述語節を修飾するのを表しているが，韓国語では読点を使わない方が多い．例では，「이중에서」が「정비한다.(整備する)」に係っていくのが正しいが，「이중에서」の文節に読点がないのですぐ後ろの述語を修飾することになって誤った依存構造になった．



- 格助詞の修飾の誤り

この研究では，格構造解析を行なっていないため，以下のように格助詞の文節とその述語節との間に正しい係り受け関係ができていない．



この実験で対象としたのはかなり長い文であり，1文当たりの文節数もかなり多い(平均 11.7)．このような長い文に対して一意に依存構造を決定しようとしたのであるから，文単位でみた解析の正解率 63% は決して低くないと考えられる．また，誤った係り受け関係を含む文についても，接続詞，または副詞節での誤りが多く，並列構造，条件・因果関係を示す表現，助詞「는(は)」による文の主題の提示，など文のおおまかな構造に関する部分は多くの場合正しく解析されており，そこからなんらかの情報を取り出すことは十分可能である．

4 日韓語文の対照分析

4.1 日韓語文の構文構造の対照分析

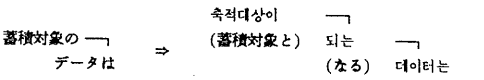
実験は，岩波情報科学辞典，日本科学技術情報センタ(JICST)発行の抄録文，サイエンス (Vol.17,No.12「科学技術のためのコンピューター」)，の各60文，合計180文の日本語文とそれを韓国語に訳した文とを各々 KNP と $hKNP$ で解析して行なった．係り受け解析を通して出てきた日韓語文の構文構造解析木に対して部分的に誤った部分は人手で修正したうえ，構文構造の一致度を調べた．180文中，33文(19%)は解析失敗文，あるいは修正しにくい文として構文構造の一致度を判断できなかった文である．74文(41%)に対しては構文構造が完全に一致しており，その中で係り受け解析が正しく行なわれているのは57文(32%)であった．73文(40%)は一致していなかった．

表 5: 日韓語文の構文構造一致度

一致	不一致	部分的不一致	解析失敗・エラー
74文(41%)	49文(27%)	24文(14%)	33文(19%)

不一致文の中で24文(13%)は文全体の構造に影響を与えない部分的な構造変化によるものであった．その例を以下に示す．

1. 単語の変化：副詞句が副詞節の形に変ったため，構造が不一致している．
   - 「区切られて いないため」 ⇒ 띄어쓰기가 되어 있지 않기 때문에(分かち書きが されて いない ため)
   - 「の」が「이 되다(が なる)」のように「助詞＋動詞」の形に変った．



2. 読点の省略，単語の削除 (図 6)：日本語文では読点が多く使われているが韓国語文ではなくなっている．これは，韓国語文は分かち書きして書くので，読点を使わなくても文を区切りやすいからだと思う．以下の例では，「両方において，」の読点が韓国語ではなくなり，「処理の途中の段階で」での「の」がなくなっている．

また，180文に対して韓日語文の文節の並び方を調べてみたが，すべての文節が同じ順序で並べていた．

4.2 日韓語の語彙の対照

ここでは日韓語文の構文構造に対して主に対照分析を行なったが，$hKNP$ の解析結果は構文構造木の形で出力されるので，日韓語の語彙に対しても効率的に対照分析を行なうことができた．これに対して以下に述べる．

1. 単語意味の省略：「効率よく」に対して「효율 좋게(効率よく)」に訳しても正しいが，「効率」の中に「よい」という意味が含まれているので，「よい」という単語に対して単に「있다(ある)」に訳された．
   - 効率 よく ⇒ 효율 있게(効率 あるように)
2. 単語の削除：「するかどうかを」の中で「するか」に「どうするか」という意味が含まれているので，「ど

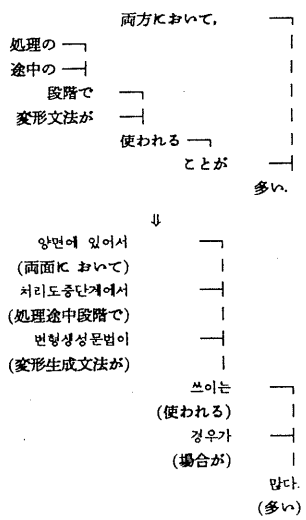


図 6: 日韓語文の対照 - 読点の省略，単語の削除 -

うか」という単語は訳されていない場合もある．また，「の」と読点は訳されない場合が多い．

- するか どうか を ⇒ 하는 가를(する かを)
- 「の」および読点の省略: 日本語で「に対して」，「において」などの文節の後ろに読点を頻繁に使っているが，韓国語ではこの文節に対応する「에 대해서」，「에 있어서」に読点をあまり使っていない．また，韓国語では「名詞＋の＋名詞」のパターンで，「の」を省略することができる．しかし，これに関する省略の規則は知られていない．

3. 単語の挿入および変化 : 上記とは違って新しい単語が加わったり，同義語であるが品詞の異なる単語に対応している．「の」が「으로 하는(とする)」のように「助詞＋動詞」に変化された．
   - 目標 の アドレス ⇒ 목표로 하는어드레스(目標 とする アドレス)
   - そして ⇒ 이렇게 하여서(このように して)
   - サ変名詞への変化
     与えられた ⇒ 부여된(付与された)
   - 決定的と ⇒ 결정적인 것과(決定的である ことと)
4. 相の変化
   - もつ ⇒ 갖고 있는(もって いる)
   - 異なっている ⇒ 다르다(異なる)

5 おわりに

本稿では，簡単な変更だけで KNP のハングル化ができ，その $hKNP$ を実験してみることによって KNP で使った方法が韓国語文に対しても有効に働くことを示した．また，同一の方法で解析する構文解析システムを用いることができ，日韓語文の対照分析が効率よく行なわれるようになった．自動的に構文構造木が図の形で出力されるので，人目でみながら便利に文の対照を行なうがことができた．とくに，いままでは人手で文を解析しそれを構文構造木に表したので，長い文の場合非常に手間がかかった．そのため，主に簡単な短い文に対して文の対照が行なわれてきたが，このシステムによって長い文に対しても対照ができると思う．

実際の例文を対象にした対照分析では日韓語は文の構造が非常に一致しており，とくに文節の語順は完全に一致していることが確認された．したがって文構造の変化を行なわず単純な語節または句の変換によって日韓機械翻訳システムを作成することができると考えられた．

しかし，ここで使った形態素解析システムの成功率が良質ではないため，性能の良い形態素解析システムの開発が望まれる．$JUMAN$ のハングル化によって形態素解析の分かち書きの単位もできるだけ同一にすることも考える必要があると思う．今後，韓国語格フレーム辞書を開発し KNP での格解析モジュールもハングル化を行なうことも望ましいと思う．

謝辞

本研究で日本語例文の翻訳に協力して頂いた早稲田大学の宋永淋氏に感謝します．ならびに韓国語の形態素解析システムを提供して頂いた韓国科学技術院の崔 紀善教授に感謝します．

参考文献

[1] Sadao Kurohashi, Makoto Nagao: KN parser: Japanese dependency/case structure analyzer, *SNLR: International Workshop on Sharable Natural Language Resources*, Nara Institute of Science and Technology, Nara, Japan, 10 - 11 August, (1994).

[2] 黄 道三，長尾 真: 分類語彙表からの韓国語シソーラスの作成，情報処理学会論文誌，Vol.35, No.2, pp210-221, (1994).

[3] 黒橋 禎夫，長尾 真: 長い日本語文における並列構造の推定，情報処理学会論文誌，Vol.33, No.8, pp1022-1031, (1992).