2C−3　機械翻訳における格の不整合が生じた場合の推定処理

菊池　浩三，　小島　義弘

（株）富士通静岡エンジニアリング

1．はじめに

富士通では機械翻訳システムATLAS を開発し，これまでマニュアル・仕様書等，簡潔に表現されているものを中心に，翻訳能力の向上が図られてきた．

近年，論文・新聞記事の翻訳も試みているが，長文・省略・口語的表現等が用いられており，さらなる翻訳率の向上をはかる必要がある．しかしこのような文を厳密に解析しようとしても，１文単位の情報では限界があることが認識されてきた．

これまでは，省略等により係り先がなくなると，文全体が翻訳失敗となり，翻訳結果は何も表示されなかった．

今後，外国人によるスキャニングを考えてみても，構文的に曖昧な部分を保留しても翻訳結果を出力することは有効と考える．

2．問題点

翻訳失敗の原因としては

① 文書が長く，係り関係が複雑になる

② 省略，短縮表現により，係り先がなくなる．

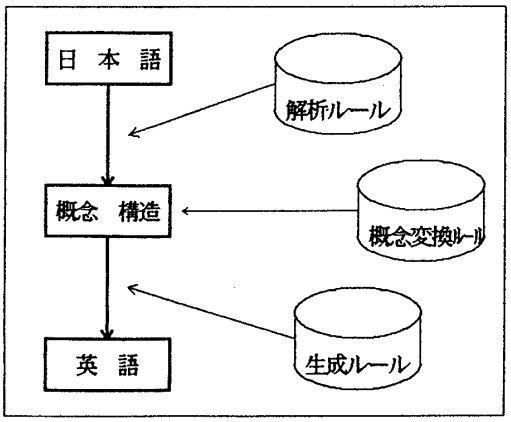
③ 辞書の属性不備

などがある．

ここでは，これまで翻訳に失敗していた文を翻訳するための，TEMPORARY-NODE,UNDEFINED-ARCを用いた処理について述べる．

3．TEMPORARY-NODE処理　（述語省略の対応）

最初にATLAS の処理の流れの概要を以下に示す．



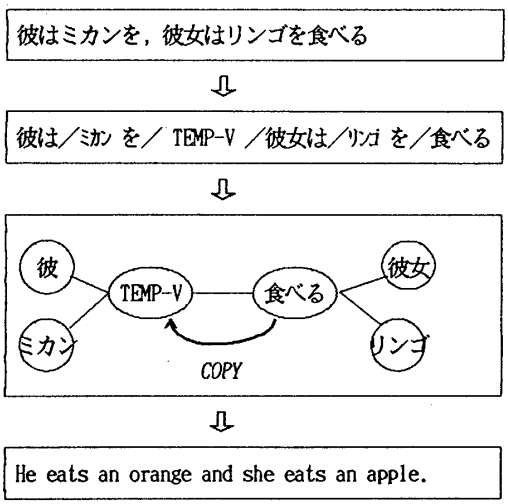
多くの機械翻訳システムがそうであるように，ATLAS においては，一つの語に同種の格は複数係ることはできない．したがって

「彼はミカンを，彼女はリンゴを食べる」

という文は「ミカンを」の係先がなくなり，解析不能になってしまう．

そこで，解析時にTEMPORARY-NODEを補い，概念構造を作成し，その後概念構造変換ﾌｪｲｽﾞにおいて真の動詞をコピーする．

処理の流れ



4．UNDEFINED-ARC 処理

UNDEFINED-ARC 処理とは，解析によって係り先を確定できない場合，最も蓋然性の高い語にUNDEFINED-ARC でかけて概念構造を作成し，なんらかの英文を訳出するための処理である．

【処理例１】：未定義格助詞相当語補正処理

タービン発電機を例として，制御性能の評価を行った．

上記の文では，「発電機を」が「例として」に係らず，他に「を格」をとっている「行った」に係ろうとしたため，「発電機を」の係り先がなくなり，解析に失敗した．

最初の解析に失敗すると，原文を分割し，再度解析す

Processing for Case Discrepancy

Kozo KIKUCHI, Yoshihiro KOJIMA

Fujitsu Shizuoka Engineering,Ltd.

る．２回目の解析において，はじめてUNDEFINED-ARC処理が作動する．

分割結果

(1) タービン発電機を　例として，
(2) 制御性能の　評価を行った．

分割の基準

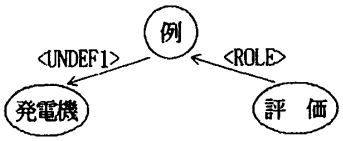
各文節内の属性に点数をあたえ，それを評価して決定する．

e.f.　格助詞，係助詞，接続助詞，名詞修飾語，カッコ，カンマ等

２回目の解析処理では，一定の句ごとに分割されるため，近くの語へと係るような解析が行われる．

最初の解析で係り先がなくなってしまった「発電機を」をUNDEFINED-ARC で「例として」にかける．

このように，最初の解析で係り先がなくなった語を，分割された同一の部分内でかけるとき，<UNDEF1>とする．



<UNDEF1>でかかっている場合，生成ではその先のノード全て（部分ネット）をカッコでくくり，先頭には ***を訳出し，解析結果が完全でないことを示す．

*** The control performance was evaluated as <turbine generator> example.

<＿＿＿> は語順等の不完全性を示す．

【処理例２】：複雑な並立の補正処理

福沢は，各国の歴史，議会政治のこと，経済・軍事，博物館などの説明をした．

上記の文では，「経済・軍事・博物館などの」が「説明をした」に<OBJECT>格で係っているので，「各国の歴史，議会政治のこと」を同じ格でかけようとすると解析失敗となる．

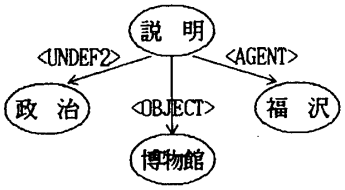
⇩

文　分　割

(1) 福沢は，
(2) 各国の　歴史，　議会政治の
(3) こと，
(4) 経済・軍事・博物館などの　説明をした．

最初の解析では，「議会政治の」の係り先がなくなった．分割後「議会政治の」は「こと」と結合し，(4)の「説明をした」にUNDEFINED-ARC でかかる．

このように同一分割部分にない主述語にかけるときは<UNDEF2>とする．



生成では<UNDEF2>で係っている場合，その先のノード（部分ネット）を独立して文頭に出してコロンでつなぎ，他のノードはその後に普通の生成と同様に文として訳出する．

*** History in each country and assembly politics: Fukuzawa explained economy, military affairs, schools, and museums, etc.

ここであげた２例も含め，翻訳失敗文は辞書整備・文法修正で大部分は対応できる．この処理では，辞書や文法でまだカバーされていない部分を後編集すれば使えるように訳出し，翻訳作業の効率化を図るためのものである．

５．評価

以上に示した処理方式を利用することにより，論文・新聞記事を対象とした翻訳失敗率はおよそ１／４に減少した．

６．おわりに

今後さらにこれらの処理条件を整備し，訳文出力率を１００％に近づけていきたい．

また，現在UNDEFINED-ARC でつながっている部分を，概念構造変換ﾌｪｲｽﾞにおいて既存のARC に決定していく処理も考える必要があると思われる．

参考文献

〔１〕信国：自然言語における長文分割方式
情報処理学会第３９回全国大会　4U-7（1989）