2E−4

外電経済ニュースの英日機械翻訳

加藤　直人*　鎌田　雅子**　相沢　輝昭*

*ＮＨＫ放送技術研究所　**（株）漢字情報サービス

１．はじめに

毎日大量に入ってくる外電に対応するために，機械翻訳のニーズが高まっている．外電で送られてくるニュースは政治，経済，スポーツ等多岐にわたっており，使われる語や言い回しも様々である．経済ニュースに関してみると，

・特殊な文型を持っている．
・高品質の翻訳を得るためには特別な日本語訳が必要である．

等の特徴がある．これらは機械翻訳にとって困難な問題であり，実際に翻訳率も低い．（我々の翻訳システムの場合，約２０％である．）しかし，これらの特徴を分析して利用すれば翻訳率の向上が期待できる．

本稿では、外電の経済ニュースに的を絞り，その特徴を利用した，外電経済ニュース英日機械翻訳システムについて述べる．

２．経済ニュース翻訳システム概要

ＡＰ電は一日３５０ほどのニュースが入電し，経済ニュースはそのうち５０ほどである．各ニュースには必ずタイトルがついており，そのニュースの特徴を表している．なかでも表ニュースでは例えば「ゴルフの結果」，「株の取引価格」など，経済ニュースでは「金価格」，「日本市場」など，毎日固定したタイトルが使われる．したがって，タイトルによって自動的に表ニュースや経済ニュースを選別することができる．

図１に外電経済ニュース翻訳システムの処理の流れを示す．

本システムでは始めにタイトルによって，表ニュース，経済ニュース，一般ニュースの３種類にニュースを分類する．表ニュースは表翻訳ルーチンに進み，英単語を日本語に一対一に単純に置き換える処理のみが行なわれる．（今回は，表ニュースの中でも経済に関係するものしか扱わなかった．）一般ニュースは従来からの機械翻訳システムで処理される．

English to Japanese machine translation system for economic news of AP wire service
Naoto KATOH*, Masako KAMATA** and Teruaki AIZAWA*
*NHK Science and Technical Research Laboratories
**Kanji Information Service Co., Lit.

経済ニュースに関しては以下で説明する３つの処理を行ない，いずれかの処理で翻訳された場合には，結果を出力し以降の処理を中止する．

処理１は，主に数量表現や時間表現のみが変化するような定型文を翻訳の対象とおり，形容詞や副詞などが１つでも付加されたような文は翻訳できない．そこで処理２では若干の柔軟性を持たせるために，定型文に完全に一致しない場合でも翻訳できるように文法や辞書を作成した．これらいずれの処理でも翻訳されなかった文は，処理３で一般の文を処理する過程と全く同じ翻訳が行なわれる．

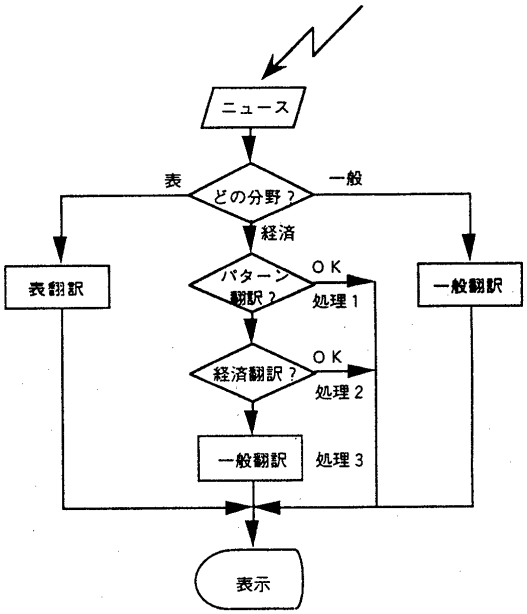

図１　外電経済ニュース英日機械翻訳システム

３．各翻訳処理

各処理における諸元を表１に示す．

３．１　処理１の諸元

経済ニュースには定型文が非常に多く，このような英文はその日本語の訳文を与えておき，単純に数量表現や時間表現等の変化部分のみを置き換えるだけで翻訳できる[1]．そこで，出現頻度が高い経済ニュース文のうち上位４００文に対して日本語訳を人手で与え，定型パターンの英語テンプレートと日本語テンプレートを自動的に作成した．テンプレートの数を増やせばカバーできる文の数も増える．

その例の一部を図２に示す．

☆入力（英文とその日本語訳）
英文
"Earlier, in Hong Kong, gold closed at 419.85 dollars an ounce, up from 417.65 dollars"
日本語訳
「それより先香港では、金は417.65ドルから上昇し、1オンス419.85ドルで引けた」

★出力（英語テンプレートと日本語テンプレート）
S→PAT1 REG PAT2 UNTEXP PAT3 UPDW PAT4 MUNTEXP
「それより先#2#では、金は#8#から#6#し、#4#で引けた」
（ここで#n#は右辺第n番目の項に対応する日本語を表す．）
PAT1 → Earlier , in 「」
REG → Hong Kong 「香港」
PAT2 → , gold closed at 「」
UNTEXP → 419.85 dollars an ounce 「1オンス419.85ドル」
PAT3 → , 「」
UPDW → up 「上昇」
PAT4 → from 「」
MUNTEXP → 417.65 dollars 「417.65ドル」

図２　テンプレート作成例

テンプレートを作成する際に，単純に数字のみを置き換えずに，数量表現を変数にしたのには
・原文とは異なった数量表現でも翻訳できるようにする．
・数字のみにすると，同じ数字が１つの英文中に２度出現した場合，どの数字がどの日本語訳に対応するか曖昧になる．
という理由による．

３．２　処理２の諸元[2)]

ここで使用する文法ルールと辞書は，'91年４月23日～25日のＡＰ電３日分（約４５０文が含まれている）を用いて作成した．経済ニュース用に文法ルールと辞書をチューニングするために使った，この３日分のデータを，以下では学習データと呼ぶ．解析できる文の種類を少なくすることによりルール数を少なくすることができ，またルール上で前置詞句の係り先もかなり特定できたので，構文的曖昧性を抑えることができた．しかし，システムの制約上，このすべての文を翻訳することができるように文法ルールや辞書を作成できたわけではない．また，従来の文法ルールでは正確に構文解析できる文でも，経済用文法ルールで誤って解析されるという問題も生まれてくるであろう．

３．３　処理３の諸元

処理３は従来の翻訳システムである．辞書は，処理２で使用した経済専門辞書の名詞のみを含む大規模のものを用い，文法ルールも一般のものである．一般文も処理３と同様に処理される．

表１　各翻訳処理の諸元

	翻訳方式	文法	辞書	処理速度
処理１	解析：パターンマッチング 生成：代入	対訳パターン 約400 局所解析用 約60	1,400 (局所解析用) 1,900 (対訳データ)	5秒／文
処理２	トランスファー方式	約500	基本語 57,000 経済語 31,000	10秒／文
処理３	トランスファー方式	約2,500	基本語 57,000 専門語 111,000	20秒／文

４．評価

学習データと非学習データ（'91年３月７日のＡＰ電，約１６０文）に対する処理別処理率と翻訳率を表２に示す．ここで評価にはMuプロジェクトの方法を用い，評価値上位２つを正解とした．

表２　処理過程別処理率と翻訳率（単位％）

		処理１	処理２	処理３	TOTAL
学習	処理率	29.1	61.3	9.6	100
	翻訳率	100	70.1	10.2	73.0
未学習	処理率	31.2	57.5	11.3	100
	翻訳率	100	58.7	22.2	66.3

学習データでは約３０％の文が処理１で翻訳され，約６０％の文が処理２で翻訳されている．翻訳率は処理１では当然のことながら１００％，処理２では約７０％である．全体の翻訳率は約７３％と高い．

学習データでも非学習データでも各処理における処理率はほとんど同じである．非学習データでは学習データに比べると処理２の翻訳率が１０％低くなっているのに伴い，全体の翻訳率も約１０％下がっているものの，従来の翻訳率に比べ格段に向上している．処理３に残った文は非常に長いものが多かった．

５．おわりに

外電経済ニュースの機械翻訳について述べた．経済ニュースは特有な文型を持っている場合が多く，その特徴を使うことによって従来より翻訳率が格段に上がった．

今後はスポーツニュース、一般のニュースも同様な方法が適用できないかどうかを検討し，外電ニュースの翻訳率向上をめざす．

［参考文献］

１）加藤，浦谷「定型パターンを含むニュース文の抽出とその英日機械翻訳」情報処理学会第44回全国大会(1992)

２）相沢，鎌田「ＡＰ電経済ニュースの英語解析用文法」情報処理学会第45回全国大会(1992)