

# 市況速報記事に対するテンプレート型日英翻訳の効果

5 J - 6 白井 諭<sup>\*1</sup> 細野明美<sup>\*2</sup> 野沢弥生<sup>\*2</sup> 木村知子<sup>\*2</sup> 阿部さつき<sup>\*2</sup> 内野 一<sup>\*1</sup>

\*1NTTコミュニケーション科学研究所

\*2NTTアドバンステクノロジ(株)

## 1はじめに

経済記事は類似の表現が繰り返し使用されるため、用例系の機械翻訳の適用が有利と考えられる。特に定型性が高い場合は、あらかじめ翻訳テンプレートを作つておくことにより、翻訳処理の効率化・高精度化を図ることが考えられる。例えば、英語の経済ニュースの場合、3割の文が100%正しく高速に翻訳されることが報告されている[加藤95]。

日本語の市況速報記事も定型性が高いため、同様の考え方が適用できる。テンプレート方式では、テンプレートの適用を厳密にすると表現の揺らぎに弱くなり、逆に柔軟性を持たせると誤適用の恐れが生じるという問題がある。

本稿では、翻訳テンプレートを用いる日英翻訳システムの構成を示すとともに、翻訳テンプレートの概要とヒット率を向上させるための改良項目について述べる。また、3種類の市況速報記事の数カ月分に対する適用実験の結果を報告する。

## 2市況速報記事の形式

本稿では、市況速報記事14週分(1995年6月～9月)のうち、東証外国部(1日1記事)、大証(1日4記事)、東証CB(1日3記事)を対象とした。これらの記事は日本経済新聞社のテレコンデータベースから取り出し、[高橋97]の方法により日英記事を対応付けた。記事の例を図1に示す。

市況速報記事を数カ月分を通してみると次のような特徴があることがわかる。

3種類の記事とも日本語・英語とも構成は同じで、見出しの後、市場の総括が1文、概況や背景が1～3文、個別銘柄の様子が2～3文書かれている。

日本語記事では、市場の総括には類似の表現が頻

(1995年7月6日)

◇大証大引け・ハイテクなど買われ3日続伸

後場寄りあとにかけて一時伸び悩む場面があつたが、大引けにかけてハイテク株や関西系の小型材料株を中心に買い進まれ、大証修正は大幅続伸。村田製、ローム、任天堂が一段高。オムロン、森精機、青山商、住友林、きんでん、ダイダン、參天薬も堅調。ホシデンはストップ高。兼松日産農、ルシアン、住友シチック、堺化学、紀州紙、松村組、明星工も人気付く。半面、日精化、サンスター、ロート、キムラタンは安い。

Osaka Stocks Cls: Surge to close higher

Stocks ended sharply higher Thursday. The morning advance stalled in early afternoon trading. But high-tech issues and small-cap, incentive-backed stocks based in Osaka drew active buying just ahead of the market close. The Adjusted Stock Price Average gained 264.55 points to finish at 16,485.18. Among the many gainers were Murata Mfg., Omron and Sumitomo Forestry. Nippon Fine Chemical, Sunstar and Rohto Pharmaceutical slipped.

(1995年7月7日)

◇大証大引け・急騰し1万7000円台回復

大証修正は前日比944円高と昨日1月31日(1255円高)以来の上げ幅を記録し、1カ月ぶりに1万7000円台を回復。日米同時の金融緩和が素直に好感され、主力のハイテク株から材料株までほぼ全面高。ローム、村田製が買われ、ホシデンは連日のストップ高。富士通ゼ、ルシアン、ダイワボウ、住友シチック、松村組、中外炉がにぎわい、小野薬、オートバックス、島精機、シマノ、青山商も上げた。半面、オムロン、堺化学は一服。

Osaka Stocks Cls: Soar to 17,000 level

Stocks on the OSE gained steadily throughout Friday, climbing above the 17,000 level for the first time in month. The easing of U.S. and Japanese monetary policies stimulated the buying of a wide variety of stocks, from high-tech issues to small-cap issues with specific incentives. The Adjusted Stock Price Average closed the day at 17,430.01 points, up 944.83. Rohm and Murata Mfg. were bought. Hosiden went limit-up. Meanwhile, Omron and Sakai Chemical Industry fell.

図1 日英の市況速報記事の例

出し、個別銘柄の様子は極めて定型的であり、共にテンプレート化が期待される。これに対して、概況や背景は文が長く表現も変化に富むため、テンプレート化は難しい。また、類似表現には助詞の省略や揺らぎ、繰り返しや語尾の違いなどがあり、これらに対応することにより適用性を高めることができる。一方、英語記事には日本語記事にはない情報の付加や表現のバリエーションの多さが目立つ。機械翻訳では不足情報の補充は基本的に不可能である。

また、日英の記事は直訳的には対応していない。しかし、意訳的なものを含め日本文と対応しそうな英文を集めて眺めてみると、標準的な訳しが推定できることがわかった。そこで、当面は英語表現のバリエーションへの対応を諦め、標準訳を決定した後、それを翻訳テンプレート化した。

## Effects of Template-based Japanese-to-English Machine Translation for Stock Market Reports

Satoshi SHIRAI<sup>\*1</sup>, Akemi HOSONO<sup>\*2</sup>, Yayoi NOZAWA<sup>\*2</sup>, Tomoko KIMURA<sup>\*2</sup>, Satsuki ABE<sup>\*2</sup> and Hajime UCHINO<sup>\*1</sup>

<sup>\*1</sup>NTT Communication Science Laboratories  
and <sup>\*2</sup>NTT Advanced Technology Corporation

### 3 テンプレート型日英翻訳の方法

図1のような記事をテンプレート方式により翻訳する場合、「企業名」と「数字」が変数候補となる。このうち、企業名には独自の省略形が用いられるため、単純な文字列切り出しでは誤りが発生する恐れがある。また、図1の2日の記事には「この日の安値圏」という記述があるが、これらも変数化できるならテンプレートの適用性が増すと考えられる。

そこで、形態素解析を行なった後、翻訳テンプレートと照合し、変数が单一単語なら対訳辞書引きで対応するが、複数単語（名詞句と複合語に限定）なら変数部分をルール型日英翻訳により英訳させることとした。形態素解析、ルール型日英翻訳にはALT-J/E[池原91]を使用した。形態素解析により単語には意味属性が付与されるため、一般名詞と固有名詞の同形など紛らわしい表現に対する翻訳のテンプレートの適用を制御することも可能になった。

なお、テンプレート方式をあらゆる文に適用するのは現実には困難である。実際の翻訳システムではルール型翻訳方式と併用することになる[白井97]。

### 4 翻訳テンプレートの作成

2節で述べたように、日本文と標準的な英文の対訳集から、人手により翻訳テンプレートを作成した。表1にテンプレートの作成の経過を示す。ただし、95年6月の速報記事は1週間分だけである。

表1から、記事の種類によってテンプレートの適用性にはかなり差があるが、テンプレート作成後のヒット率は記事の種類ごとにほぼ安定しており、作成率も徐々に減っていくことがわかる。

しかし、8月の記事では「夏休み」「盆休み」のような語を含む文で、9月の記事では中間期決算報告絡みの文でアンマッチが多く発生した。このような年に1度のイベントがほかにもいくつか考えられるので、市況速報に対する翻訳テンプレートは1年単位で整備する必要があると考えられる。

また、東証CBのテンプレート作成前のヒット率が8月はあまり上がってないが、7月と比較して用語の違いや助詞の揺らぎが目立った。おそらく日本語記事の記者が交代したものと思われる。しかし、基本的には既存のテンプレートの修正により対応できたので、その後はテンプレートを作成する際に様々なバリエーションを考慮するようにした。

図2に翻訳テンプレートの例と、それが適用される日本文および訳文を示す。

### 5 おわりに

本稿では、翻訳テンプレートの適用による日本語の市況速報記事の英訳において、翻訳システム構成

表1 翻訳テンプレートの効果

東証外国部

年月	原文数	作成前	作成後	作成数	累積数	作成率
95.6	58文	0.0%	89.7%	20件	20件	34.5%
95.7	221	35.3	82.4	42	62	19.0
95.8	274	55.1	85.8	38	100	13.9
95.9	236	61.1	81.4	29	129	12.3

大証

年月	原文数	作成前	作成後	作成数	累積数	作成率
95.6	96	0.0	60.4	34	34	35.4
95.7	462	29.9	60.8	105	139	22.7
95.8	525	37.1	56.8	68	207	13.0
95.9	429	44.5	58.7	46	253	10.7

東証CB

年月	原文数	作成前	作成後	作成数	累積数	作成率
95.6	131	13.0	45.0	25	25	19.1
95.7	522	18.6	42.0	82	107	15.7
95.8	555	17.3	42.0	80	187	14.4
95.9	532	37.8	46.8	66	253	12.4

作成前 = テンプレート追加前のヒット文数 / 原文数 (未知データ試験)

作成後 = テンプレート追加後のヒット文数 / 原文数 (既知データ試験)

作成率 = 作成ルール数 / 原文数 ※いずれも月ごとの集計

(A370009

((("売買/商い") ("/"が/) ("成立し") ("た") (1 \* "1610") {"銘柄"}) ("の"))  
("うち") {"("、")} ("値上がり/値上がり銘柄") {"("が/は")} (2 \* "1610")  
("、") ("値下がり/値下がり銘柄") {"("が/は")} (3 \* "1610") ("、")  
("変わら") {"("ず")} {"("が/は")} (4 \* "1610") {"("て")} ("、") ("比較でき")  
("ず") {"("が/は")} (5 \* "1610") {"("だつ")} {"("た")}) ("。")  
("Among the 1 " issues changing hands, " 2 " rose, " 3 " fell, " 4 " remained unchanged and comparisons unavailable for " 5 ".")

<グラフ表示> 売買の成立した1銘柄のうち、値上がり銘柄は2、  
値下がり銘柄は3、変わらずは4で比較できずは5だった。

<入力> 売買の成立した34銘柄のうち、値上がり11、値下がり10、変わらず4、比較できずは9だった。

<翻訳> Among the 34 issues changing hands, 11 rose, ten fell, four remained unchanged and comparisons unavailable for nine.

### 図2 翻訳テンプレートの例とその適用

と翻訳テンプレートの作成見通しについて述べた。現在は人手により翻訳テンプレートを作成しているが、適用性をなるべく高く保ちながら、誤適用の防止を図ったからである。今後は、作成済みテンプレートに基づいた条件記述の整理や、統計処理による変数化支援により[内野97]、テンプレート作成の効率化を図っていく予定である。

### 参考文献

- [池原91] S. Ikebara, S. Shirai, A. Yokoo & H. Nakaiwa: Toward an MT system without pre-editing --Effects of new method in ALT-J/E--, In Proc. of MT SUMMIT '91, pp.101-106
- [加藤95] 加藤：定型パターンを含む文の機械翻訳手法，情報処理学会論文誌, Vol.36, No.9, pp.2081-2089
- [白井97] 白井, 松島, 井上, 松尾, 矢部, 大山: 市況速報記事を対象とした日英翻訳システムの構成, 情報処理学会第55回全国大会, 5J-5, Vol.2
- [高橋97] 高橋, 白井, 大山, 渡邊, 上田: 日英新聞記事の記事対応コーパス自動作成, 言語処理学会第3回年次大会, D1-4, pp.127-130
- [内野97] 内野, 白井, 池原: 離散共起表現データを用いた単語のグループ化, 言語処理学会第3回年次大会, C1-7, pp.107-110