

5S-7

英文ビジネスレター 推敲支援プロトシステムの開発

小澤 邦昭 絹川 博之
(日立製作所 システム開発研究所)

高橋 成佳 神成 広之
(日立京業エンジニアリング) (日立マイコン)

1. はじめに

外国語で文書を作成するという人間の知的作業を、計算機で支援する研究を進めている¹⁾²⁾。今回、そのうちの主要機能の一つである英文推敲支援機能を取り上げ、プロトタイプシステムを開発した。このシステムは「日本人英語」から「ネイティブの英語」への推敲支援を目的とし、日本人の書いた英文に含まれる誤りを見つけ、誤りの理由と修正方法をメッセージとして出力する。推敲対象文書として、今回は英文ビジネスレターを選んだ。

2. 推敲支援プロトシステムの概要

本システムでは、次の誤りを推敲の対象とする。

(1) 基礎的文法事項に関する誤り

丹念に辞書を引けば防ぐことのできる文法上の誤り、およびケアレスミスの文法上の誤りを、ここでは基礎的文法事項の誤りと呼ぶことにする。本システムで取り上げる誤りの種類は、次の通りである。

- ・冠詞と数 (例: equipments ⇒ equipment)
- ・動詞型 (例: assure you to do ⇒ that節)
- ・主語と述語の一致 (例: the division are ⇒ is)

(2) 英語慣用表現に関する誤り

文法的には誤りではないが、ネイティブ・スピーカーにとっては不自然な表現を、ここでは英語慣用表現に関する誤りと呼ぶことにする。誤り例を次に示す。

- ・It was my great pleasure to meet you ⇒ a
- ・We would be very happy if you could inform us
⇒ greatly appreciate it

なお、本システムでは、推敲支援機能の他に、英語慣用表現の推敲ルールを、ユーザが追加登録できるエディタを設けてある(図1参照)。

3. 文法事項の推敲支援処理

文法事項の推敲を行うには、構文解析が必要となる。従来研究されてきた構文解析は、文法的に正しいテキストが入力されることを前提としている。このため、文法規則や辞書に記載されている制約を利用しながら構文解析を行うことが可能である。これに対し、推敲支援処理では、文法的誤りを含むテキストが入力されることを前提としなければならない。このため、文法規則の一部や辞書記載の制約の一部を意図的に緩めて(最初から緩めておく方法と、正常に解析できなかったときに緩める方法がある)、従来の構文解析技術を利用する方法が採られてきた^{3)~5)}。

ここでは、文法規則や制約をできるだけ使わないことを狙って、品詞の並びからテキストで使われている文型を解析し、この文型が文法規則や辞書情報に違反しないか調べる方法を採用した。文型解析の処理手順の概要を次に示す。なお、形態素解析は既に終えているものとする。

- (1) 品詞選択ルールにより、多品詞解消を行う。
- (2) 述語動詞(句)に着目して、その直後で文を区切る。各区分において、述語動詞(句)の主語を見つける。
- (3) 述語動詞(句)の後の品詞の並びから文型を決める。
- (4) 接続詞、関係詞の有無等から、述語動詞(句)が主節、従属節、関係詞節のいずれに属するか決める。

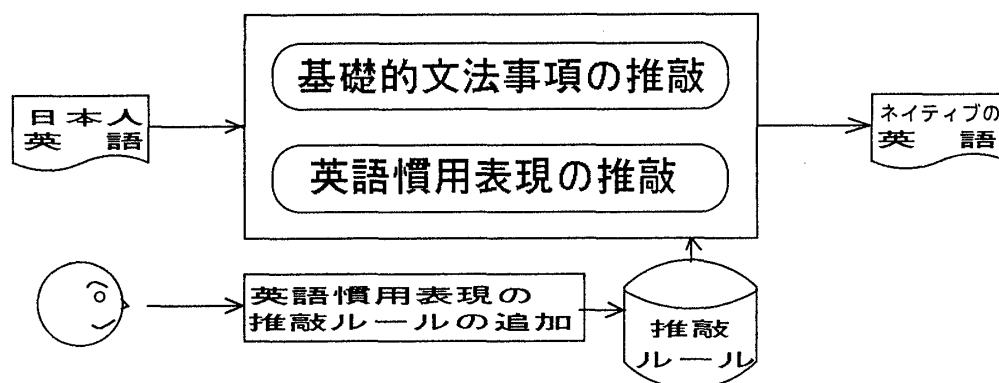


図1 英文推敲支援プロトシステム

Prototype Development of Computer Aided English Text Improving System for Business Letters
Kuniaki OZAWA, Hiroshi KINUKAWA (Hitachi), Shigeyoshi TAKAHASHI (Hitachi Keiyo Engineering),
Hiroyuki KANNARI (Hitachi Microcomputer Engineering)

文型解析処理の例を次に示す。

We are proceeding several steps required by the Japanese Government.

(1) "proceeding" は名詞と動詞(ing形)の多品詞であるが、「be 動詞の直後の ing 形」に関する品詞選択ルールにより、述語動詞と判定する。また、「直前が名詞、直後が by なら過去分詞」のルールにより、"required" を述語動詞ではなく、過去分詞と判定する。"steps" は形容詞 "several" の直後なので、名詞と判定する。

(2) 述語動詞 "proceed" の前方に位置する品詞の中から、代名詞主格の "We" を主語と判定する。

(3) 述語動詞の後に、主語でなく、かつ前置詞句の要素でない名詞があるので、文型は S ("We"), V ("proceed"), O ("steps") と判定する。

(4) 述語動詞は一つなので、主節に属すると判定する。

文型解析処理(3)により、動詞 "proceed" は、テキストでは他動詞として使われていることが分かる。一方、辞書に記載されている動詞型には他動詞の型はない。よって、"proceed" の使い方は誤りであると判定し、「"proceed" の後ろに "with" 等の前置詞が必要」とのメッセージを出力する。

4. 英語慣用表現の推敲支援処理

英語慣用表現の推敲支援処理では、誤り表現を予め推敲ルールとして登録しておき、このルールとテキスト中の表現との照合を行う。照合は、ルール中に設定した検索キーワードを利用して行う。照合が成功したときには、ルールに対応した推敲メッセージを出力する(図2参照)。

本処理で留意すべき点の一つとして、例外処理が挙げられる。推敲ルールとの照合が成功しても、場合によ

ては誤りと見なせない場合もでてくるからである。例えば、"It is my pleasure to V(動詞の原形)" は誤り表現である。この表現において、"my" を使うと、横柄な感じを与えるので、ネイティブ・チェックは通常、"a" に書き換える。ただし、V に相手の利益になる意味の動詞 ("help", "assist", 等) が来れば、"my" を使っても誤りではなくなる。この種の動詞が使われているときには、推敲ルールとの照合が成功しても、誤りと見なさない例外処理が必要となる。

5. プロトシステムの規模

本プロトシステムは、単語数約5万語の辞書を備えている。この辞書は、英日機械翻訳システム HICATS/EJ (Hitachi Computer Aided Translation System) の基本辞書をもとに作成した。英語慣用表現の推敲ルールは、1000ルールを目標に作成中である。プログラム量はC言語で、約20キロステップである。

6. おわりに

最近、「日本人英語」に関する書籍の出版が増えており、文法的にも語法的にも正しい英語を使おうという機運が高まってきている。本プロトシステムは、このようなユーザの要求に、計算機を利用して応えようという試みである。今後は、この分野の研究が一層活発化し、実用システムが開発されることが期待される。

参考文献

- 1)小澤邦昭ほか: 情報学会第38回全国大会予備集, pp. 220-221(1989)
- 2)小澤邦昭ほか: 情報学会第40回全国大会予備集, pp. 482-483(1990)
- 3)河合敦夫ほか: 情報学会論文誌, vol.25, No.6, pp. 1072-1079(1984)
- 4)Ikuo Kudo, et al.: Proc. of COLING'88, pp.341-347
- 5)中野幹生ほか: 情報学会NLP研究会報告72-2, pp.(1)-(8),(1989)

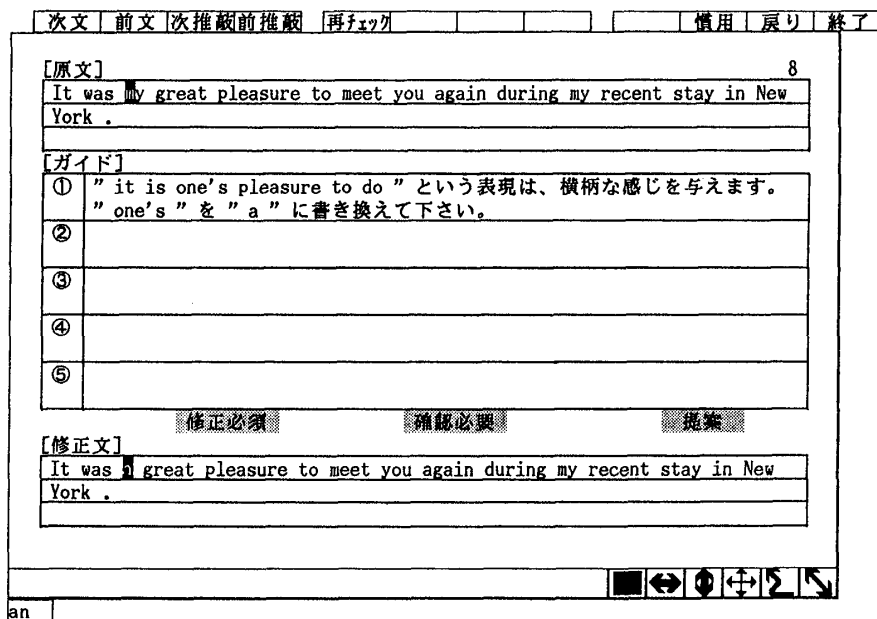


図2 推敲メッセージ出力画面例