

スベル補完機能を組み込んだ英文作成支援システムの試作

酒匂 孝之

木下 聡

{takayuki.sako,satoshi.kinoshita}@toshiba.co.jp

(株) 東芝 研究開発センター 情報・通信システム研究所

1 背景

近年、オフィスや家庭での計算機の普及により計算機を使って文書を作成する機会が増えてきている。また、インターネットの発達によって、海外へメールを出したり、書類を作成するなど、英語の文書を作成する機会も増える傾向にある。そのため、日英機械翻訳アプリケーションや様々な英文作成支援アプリケーションが開発され市販されている。

しかし、日英機械翻訳アプリケーションは日本語文書をアプリケーションが直接、英文に翻訳するものであり、そこにユーザの介入する余地は少ない。そのため、翻訳された結果がユーザの意図したものとは異なるという場合があり、英語に精通したユーザにとってはかえって修正に手間がかかる場合もある。また、多くの英文作成支援アプリケーションは、予め用意された表現を組み合わせる文書を作成するものが多く、それ以外の表現を使用したい場合は普通のテキストエディタを使用して、英文を作成する場合と変わらない。そこで[赤峯 91]では日本語を入力して、逐次英語に変換していく手法が提案されている。この手法は予め日本語のテキストが存在しない場合は有効であるが、予め日本語のテキストが存在する場合でも、一度、日本語を入力してから変換しなくてはならず、手間がかかる。

また、[Foster96]は英仏語間の単語補完手法を提案している。これは予め与えられた対訳コーパスを使って、ある文脈中で、ある英単語が仏語に変換される確率を求め、入力される毎に、その尤もらしい仏語を補完するというものである。しかし、この手法では予め大量の対訳コーパスを用意する必要が

ある点と、集めたコーパスの性質によってその後の補完に影響が出る点が問題と考えられる。

2 英文作成支援システム

そこで我々は、システムが予め電子化されている日本語文書を利用して、ユーザの英単語の入力を補完し、補完候補を提示するスベル補完機能を組み込んだ英文作成支援システムを作成した。

本英文作成支援システムは以下の特徴を持つ。

- 入力原文に適したスベル補完
- 簡単なキー操作による補完が可能
- 日本語の読みによる補完機能

予めシステムは翻訳辞書を利用して日本語文書を形態素解析し、日本語文書に出てくる単語の訳語情報を取得する。ユーザが英単語の一部、もしくは日本語の読みを入力した段階で訳語情報を検索し、適切な訳語の候補（補完候補）をユーザに提示する。

本システムを利用することにより、文書入力のために行われるキー入力の回数が削減されることが確認された。

以下、上記英文作成支援システムの概要とその評価について述べる。

2.1 画面構成

図1に本英文作成支援システムの画面構成を示す。本システムは原文表示画面、訳文表示画面、情報表示画面の3画面で構成される。

原文表示画面には英文作成の元となる日本語文書（原稿）を表示させる。現在ユーザが作成している英文と対応する日本語文は、フォーカス文として他の文と区別をするために色を変えるなどの方法を使って強調する。

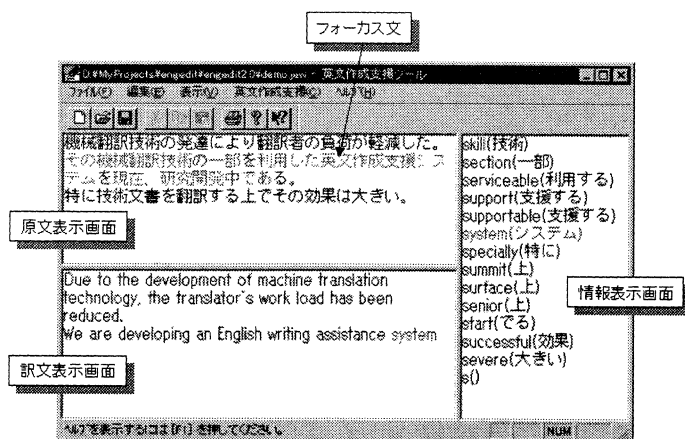


図 1 画面構成

訳文表示画面はユーザが実際に英文文書を作成する画面であり、ユーザの入力はこの画面に表示される。また、補完した単語の第一候補もこの画面のカーソルの位置に表示される。

情報表示画面にはユーザの指示によってシステムが検索した補完候補を表示させる。ユーザは情報表示画面に表示された補完候補をマウス、カーソルキー等を用いて選択することができる。

2. 2 スペル補完機能

図 2 は本システムの特徴であるスペル補完機能の処理の概略を示す。システムはユーザが作成する英文に対応する日本語文（フォーカス文）を、機械翻訳用の辞書を利用してあらかじめ形態素解析し、訳語リストとしてシステム内部に保管する。この訳語リストは見出し語、訳語、読みの情報によって構成される。次にユーザが何かキー入力を行った場合、システムは訳語リストを検索し、第一の補完候補を訳文表示画面のカーソルの位置に、全ての補完候補を情報表示画面に出力する。ユーザは訳文表示画面に表示された補完結果でよい場合は、そのまま文字入力を続け、他の候補を選びたい場合は、情報表示画面に表示された補完候補を選択することによって英単語を入力することができる。

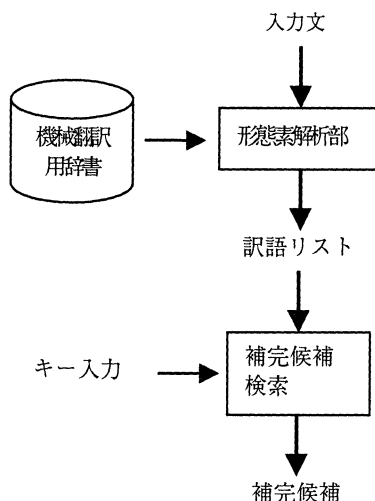


図 2 スペル補完機能概略

2. 2. 1 英単語の一部を利用したスペル補完

ここでは英単語の一部を入力したスペル補完機能について述べる。この機能は入力したい単語のスペルを途中まで入力し、補完指示を与えることによって、システムが残りのスペルを補う機能である。本補完機能の概略を図 3 に示す。

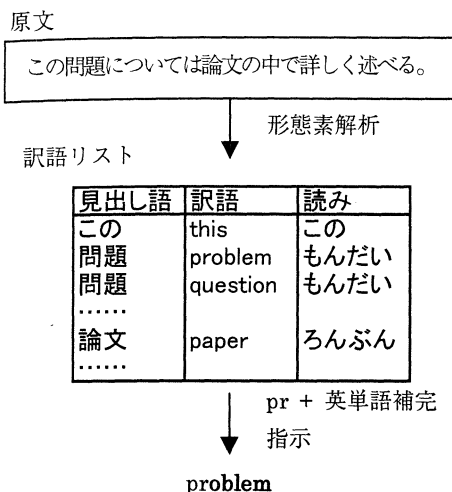


図3 英単語の一部を利用した補完

例えば、ユーザが「問題」の訳語として “problem” を入力しようとしたとする。ユーザが “pr” と入力した段階で英単語補完指示を与えるとシステムは訳語リストを検索し、訳語が “pr” で始まる単語を検索する。その結果、システムは “problem” を見つけ、残りの “oblem” を補完して訳文表示画面に表示する。本来、“pr” で始まる単語は “program” “probability” など多数存在するが、2. 2で示した通り、予め作成する英文文書の元となる日本語文書を形態素解析して訳語リストを作成し、その訳語リストを検索することによって補完処理を行っているため、上記の補完が可能になる。

もちろん、入力単語によっては一意に決まらない場合も存在する。例えば、前出の例ではユーザが “p” と入力した段階で補完指示を与えた場合、補完候補として “problem” の他に “paper” などが見つかる。そこで第一候補として、“problem” が訳文表示画面に表示され、“problem” “paper” など全ての候補が情報表示画面に表示される。もし、ユーザが入力しようとした語が “paper” である場合は次候補ボタンやマウス等を使用して訳語を選択することによって補完候補を確定する。

2. 2. 2 日本語の読みを利用したスペル補完

2. 2. 1で示した、単語のスペルの一部を入力したスペル補完では、ユーザは少なくとも英単語の一部（先頭部分）を知っていなくてはならない。しかしながら、入力する単語そのものが分からない場合も少なくない。そこで、第二の入力支援機能として、英単語に変換したい日本語の読みを入力することによってスペルを補完する機能を導入した。本補完手法の概略を図4に示す。

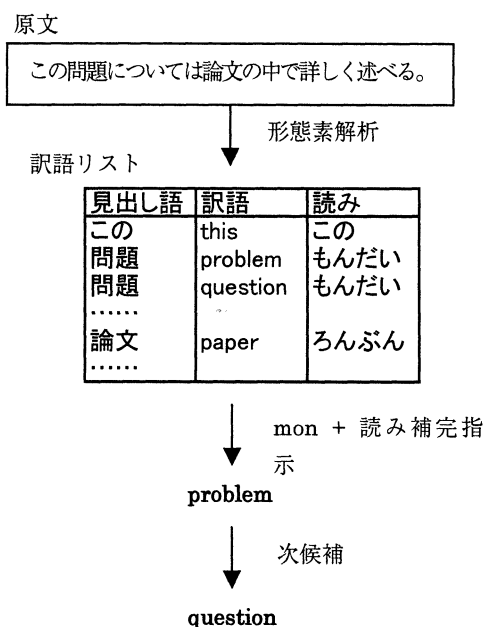


図4 日本語の読みを利用した補完

例えばユーザが「問題」の訳語を入力したい場合、訳文表示画面に「mon」と入力し、読み補完指示を与えると、システムは訳語リストを検索し、読みが「もん」から始まる単語を検索する。この場合は “problem” “question” が見つかり、“problem” が訳文表示画面に、“question” を含んだ全ての候補が情報表示画面に表示される。もし、ユーザが入力しようとした語が “question” である場合は、情報表示画面に表示された訳語を選択することによって

文書	単語数	文字数	キー打鍵数	削減率	補完単語数	補完単語率
特許抄録(7文書)	1653	9481	7719	18.6%	336	20.3%
一般文書(8文書)	1804	10679	9278	13.1%	376	20.8%

表 1：実験結果

単語が補完される。

3. 評価結果

本英文作成支援システムの有効性を調べるために、キー入力削減率と補完単語率を調査した。評価方法は、まず、初めに日英対訳テキストを用意した。次にその日本語テキストをシステムに読み込みませ、英文テキストを作成する英文の見本としてスペル補完機能を使いながら入力し、キー打鍵数を算出した。なお、評価に使用したテキストは特許抄録の対訳テキスト7文書（計1653単語）、一般文書の対訳テキスト8文書（計1804単語）の2種類で、1文書に対して2人の被験者に入力させ、平均打鍵数を算出した。評価は入力すべき単語が予め決まっているので、英単語のスペルの一部を入力した場合の補完のみを対象とした。

3. 1 キー入力削減率

まず、キー入力削減率を求めた。普通に入力した場合、入力ミスがないと仮定すると、キー打鍵数はテキストの文字数（スペースを含む）と一致するので、この文字数を基にして、キー削減率を算出した。

$$\text{キー削減率} = \frac{\text{テキスト文字数} - \text{キー打鍵数}}{\text{テキスト文字数}}$$

なお、本システムの打鍵数は補完指示を与えるためのキー入力、及び候補を選択するために押されたキー入力も含んでいる。

表1に実験結果を示す。表から分かるように、特許抄録では18.6%、一般文書では13.1%のキー入力削減された。特許抄録のキー削減率が一般文書に比べて高いのは、文字数の多い専門用語が多いためである。

3. 2 補完単語率

次にユーザがテキスト中の何単語に対して、単語

補完指示を与えたかを示す、補完単語率を求めた。

その結果、特許抄録、一般文書ともにユーザは約20%の単語に対して補完指示を与えていることが分かった。補完単語率が予想より低かったのは冠詞や入力コストの低い文字数の少ない語については補完操作に頼るより自分で入力した方が早いと判断したためと考えられる。単語数の少ない語については補完候補に入れない等の操作を加えることによって、ユーザが候補を選択する操作が減り、更に使いやすいシステムとなると考えられる。

4. おわりに

本報告では、スペル補完機能を組み込んだ英文作成支援システムを提案した。また、実験により、スペル補完機能を用いることによって、ユーザの文書入力効率が向上することを確認した。

今後は補完候補の絞り込みの手法や提示方法、原文文書がない場合の入力支援方法（日本語分野での支援方法としては、かな漢字変換における入力予測の研究[市村 99]等がある）について検討をし、ユーザにとって更に使いやすい英文作成支援環境を提供するよう研究を進める。

5. 参考文献

[赤峯 91] 赤峯,奥村,村木:日本語入力による英文作成支援,情報処理学会,第 43 回全国大会, 2H-10, 3-205,(1991).

[Foster96] George Foster, Pierre Isabelle, Pierre Plamondon : Word Completion: A First Step Toward Target-Text Mediated IMT", COLING 96, Vol.1, pp.394-399, (1996).

[市村 99] 市村,斎藤,木村,平川:予測に基づく入力支援機能を備えたかな漢字変換システムの開発, 情報処理学会研究報告,99-NL-129,pp.63-70(1999)