

英文作成支援システムにおける 例文検索方式

3F-2

寺濱 幸徳* 小澤 邦昭** 小嶋 弘行* 絹川 博之*

*株式会社日立製作所 システム開発研究所

**株式会社日立製作所 コンピュータ事業部

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)~3)}。英文作成において、特に日本人の場合、例文参照機能は有効な手段と考えられる。これは、模範的な英語表現(例文)を事前に集め、実際に英文を書く時に、この例文の中から、役立つ表現を活用して英文を作成支援する機能である。効率的な例文参照を行なうためには、検索対象文に対して、意味的類似性を許容する例文検索を可能とすることが必要である。本報は、検索キーワードにより例文の意味を表すことを考え、検索キーワードの組合せ、あるいは、これと意味的に類似なものに言い換えるための類似表現辞書を利用する新しい例文検索方法を提案する。

2. 例文検索の考え方

検索対象文に対して、検索対象文の意味を表す複数の単語で構成した検索キーワードを用いて例文を検索する。この時、検索キーワード中の単語を一つでも含まない場合は、この例文の検索は不可能である。一方、検索キーワードを構成する単語の数を削減して例文を検索すると、検索する必要のない例文まで検索され、この検索例文中から、再度所望の例文を検索する必要性が生じる。そこで、検索キーワードを構成する単語の数を限定し、かつ意味的類似性も考慮できる効率的な例文検索方法を考える必要がある。

英文の意味は、主語、動詞、目的語を一つの組として一義的に表せると考えられる。これを検索対象文の意味を表す検索キーワードとなる複数の単語の組合せとして選定し、英文

の作成を支援するための日本語文の意味的な類似例文検索に適用することにした。

本報では、例文を構成する単語の中から、例文の意味を表す検索キーワード(主語、動詞、目的語)を設定し、意味的に類似な主語、動詞、目的語の組合せを格納した類似表現辞書を用いる例文検索方式について述べる。

3. 例文検索方式の概要

例文検索方式の手順について、この概要を以下に述べる(図1)。なお、検索対象となる文は、日本語文とし、例文データベースから検索された例文(日本語文)に対応した英文が最終的に検索されることになる。

- (1) 検索対象文から既存の構文解析技術を利用することにより、検索キーワード(主語、動詞、目的語)を抽出する。抽出した検索キーワードをもとに、例文が格納されている例文データベースより検索キーワードを含む例文を検索する。
- (2) 検索キーワードを含む例文が例文データベースに存在しない場合、検索キーワードに意味的に類似な単語の組合せを、図2に示すごとく類似表現辞書を用いて検索する。単語の組合せとしての類似表現辞書には、基準となる単語(基準語)を中心に、一般的に出現頻度の高い語(基本表現語)と、それに類似する単語が格納されている。これにより、同じ基本表現語に属する類似な単語の選択が可能となり、この類似な単語の組合せを用いて例文データベースから検索することができる。

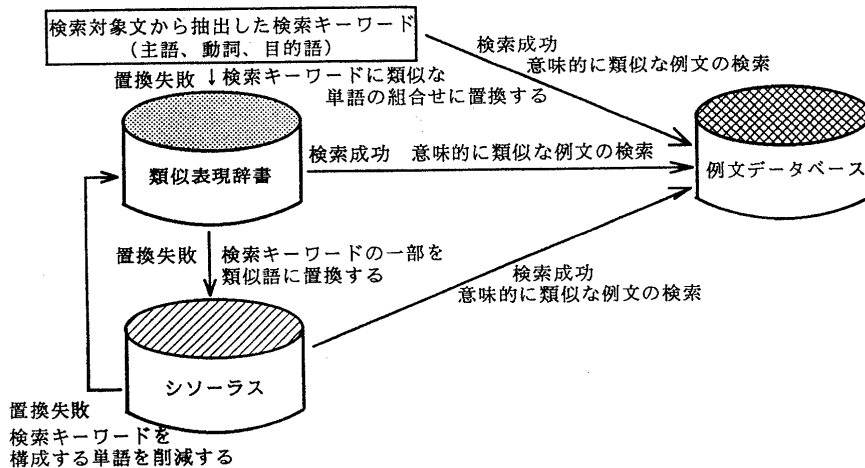


図1 例文検索方式の概要

Example Sentence Search Method in Computer-Aided English-Writing System
Yukinori TERAHAMA, Kuniaki OZAWA, Hiroyuki KOJIMA, Hiroshi KINUKAWA
HITACHI, Ltd.

- (3) 検索キーワードを含む例文が例文データベースに存在せず、類似表現辞書を用いても例文が検索出来ない場合、検索キーワード中の基準語をシソーラスにより言い換える。再び類似表現辞書を用いて類似な単語の組合せに言い換え、この単語の組合せを用いて例文データベースから検索することができる。
- (4)(2), (3)の検索手段を用いても、検索文に類似な例文が検索できなかった場合、検索キーワードとなる単語の数を削減したものをを用いて、例文データベースから検索する。

4. 提案した例文検索方式による例文検索例

下記の文を検索対象文例として、意味的に類似な例文の検索実行例を示す。なお、類似表現辞書および例文データベースの作成にあたっては、市販の辞書^{*)}、および市販の書籍^{*)}を用いた(図3)。

検索対象文:

「彼が選挙に落ちるとは、だれも予想しなかったことだ。」

この検索対象文から構文解析により、検索キーワードを抽出すると、

(主語, 動詞, 目的語)=(彼, 落ちる, 選挙)

となる。次に、例文データベースより抽出した検索キーワードを含む例文が存在するかどうかを検索する。ここで、検索が失敗したとすると、類似表現辞書により、検索キーワードと意味的に類似な他単語の組合せに置換する手順を実行する。置換した結果、検索キーワードに類似な他単語の組合せとして、例えば、

(主語, 動詞, 目的語)=(*, 敗退する, 選挙)

が検索キーワードとして得られたとする。

ここで「*」は、主語を限定しないことを示す。

したがって、置換された単語の組合せを含む、次の例文が例文データベースより検索される。

検索結果:

「先月の総選挙で社民党政権が敗退した。」

対応する英文データ:

"In the general election in Sweden last month, the Socialist Democratic Labor Party government was defeated."

類似表現辞書番号	0 0 1		
基準語	選挙		
基本表現語: 落ちる	敗れる	落選する	敗退する
例文番号: ---	---	0 4 8	0 4 2
する	行う	運営する	
---	---	0 4 5	
利用する			
0 4 3			
動き出す			

注) ---は、該当する例文が例文データベース上に存在しないことを表す。

図2 類似表現辞書

以上の検索方法で明らかのように、1語のシソーラスでは、必ずしも結び付かない「落ちる」と「敗退する」を、「選挙」と組合せ、2語で検索キーワードを定義することにより、意味的に類似な例文の検索が可能となることを示した。

5. 例文検索方式の検討

本報で示した例文検索方式は、意味的に類似な例文を検索しようとする場合に、従来の1語のシソーラスによる検索だけではなく、検索対象文の表す意味を表現した検索キーワードの利用を図った。さらに、類似表現辞書から検索キーワードに類似な他単語の組合せに置換することにより、例文データベースから検索対象文に意味的に類似な例文も検索できるものであり、本方式が有効である見通しを得た。

ここで、検索キーワードの類似表現を格納した類似表現辞書については、一般的に十分に整理されているものではない。したがって、今後この類似表現辞書を容易に構築できるための仕掛けを考えることが、課題としてあげられる。さらに、例えば、検索対象文が受身形の場合、まず、能動形に変更してから構文解析により検索キーワードを抽出するなどして、文体を統一した上で例文を検索することが、効率的な検索方式として肝要と考える。

なお、検索キーワードだけでは、どうしても例文を検索できない場合には、検索対象文に関連した出典分野などの情報を付加することにより、ある程度関連した例文を検索できるようにすることも有効と考えられる。

6. おわりに

意味的に類似な例文を検索する場合、不必要な例文まで検索してしまうことを防ぐために、本報での例文検索方式を提案した。本検索方式は、以下の特徴を有する。

(1) 意味を代表する複数の語の組合せにより、検索キーワードを定義した。

(2) 類似表現辞書により検索対象文に意味的に類似な例文を検索できる。

さらに、文体を統一した上での検索キーワードの設定や、類似表現辞書の容易な構築方法の必要性など検索方式の効率化について、示唆を得た。

参考文献

- 1) 小澤邦昭ほか: 借見学会第38回全国大会予備集, pp.220-221 (1989)
- 2) 小澤邦昭ほか: 借見学会第40回全国大会予備集, pp.482-483 (1990)
- 3) 小澤邦昭ほか: 借見学会第41回全国大会予備集(3), pp.181-182 (1990)
- 4) 小澤邦昭, 寺津幸雄ほか: 借見学会第42回全国大会予備集(3), pp.112-113 (1991)
- 5) 林 史典: 国語基本用語辞典, 教育社 (1986)
- 6) 朝日新聞論説委員会: 天声人語 '91年の号(第87集), 原書房 (1992)

例文番号	例文内容
0 4 2	先月の総選挙で社民党政権が敗退した。 In the general election in Sweden last month, the Socialist Democratic Labor Party government was defeated.
0 4 3	改革そっこのけで総選挙に利用されてはね。 They are using it as a tool in the race for the presidency of the governing party. (it = discussion on political reform)
0 4 4	法案の廃案や国会の閉会と重なって自民党の総選挙が動き出している。

図3 例文データベース