

例文をもとにした英文書作成支援システム

武田 明子[†] 古郡 延 治^{††}

例文を利用し、英文の文章作成を支援するシステムを構築した。文書作成支援ツールは、文章の中にある誤りを検出・訂正するために開発されてきた。しかし、文法やスタイル・チェックは日本人の英文作成にはあまり役立たない。われわれの英語は、これらのツールの及ばない独特の間違いを含んでいるからである。本システムは、書かれた文章の検査ではなく、データベースの中にある例文を使い、次の手順で、英文らしい英文を書くことを支援する。(1)日本語文を入力する。(2)キーワードを抽出し、それを例文検索用のキーワードに変換する。(3)キーワードから検索用の式をつくる。(4)データベースから例文の検索をする。(5)入力文を英文にするのに使える例文があれば、それを所期の文を得るために編集する。なければ、ユーザにキーワードの選択を要求し、(3)に戻る。本研究では、例文検索に項目表と事例ベース推論を利用した。その特長は、(1)意味処理を行わずに“含意”検索を実現していること、(2)学習により、検索を繰り返すに依り、検索効率が上がっていくことにある。われわれは、論文を書くときに必要な謝辞と著者紹介に関する文章例で実験を行い、システムの実用性を確認した。

A Sample-Based System for Helping Japanese Write English Sentences

AKIKO TAKEDA[†] and TEIJI FURUGORI^{††}

Writer's aid programs correct erroneous sentences in spelling, punctuation, syntax, grammar, usage, etc. However, the systems to check irregularities in English are not very helpful for the Japanese users of English. The primary problem they have is not in forms but in idiomatic ways of expressing ideas in English. Described here is a system that helps the Japanese write sound English sentences. It takes as input a Japanese sentence, extracts relevant keywords from it, converts them to their English equivalents, retrieves English sentences from a database, and presents them as the candidates for expressing what is in the Japanese sentence. The system may use one or more of the sentences retrieved to write or edit English, or ask with user-specified keywords for other examples when the ones presented are not suitable for use. We have tested the system in writing sentences for "Acknowledgments" and "About the Author", both necessary for writing academic papers. The results shows that, once sufficient amount of examples are taken for creating the database, the system is considered to be effective and helpful for the Japanese users of English.

1. はじめに

計算機による文書作成は、校正や推敲を支援するプログラムの開発によって補強されてきた^{1),2)}。英文の文書作成には、誤綴の訂正、主語と動詞の一致や冠詞と名詞の一致などを含む文法のチェック、用語や文の長さ、構文に関するスタイルのチェックを行う支援ツールなどがある^{*}。最近では、日本語文の推敲、添削、校正を行うプログラムの開発も進んでいる³⁾⁻⁶⁾。

本稿では、日本人の英文文書作成を支援する一システムの構築と、その評価結果を報告する。われわれは、ここで、文章の中にある誤りを検出・訂正するのではなく、データベース中の例文の検索を行い、その修正や合成、編集を繰り返すことによって所期の文章をつくっていくシステムの開発を試みる⁷⁾。

2. 文書作成支援ツール

一般に文書作成支援ツールは、対象言語のいかににかかわらず、ユーザが不用意に犯す間違いを訂正する

[†] 横河システムエンジニアリング株式会社
Yokogawa System Engineering Corporation

^{††} 電気通信大学電気通信学部情報工学科
Department of Computer Science and Information Mathematics, Faculty of Electro-Communication, University of Electro-Communications

^{*} 市場に出ているシステムに、RightWriter (Que Software), Grammatik IV (Reference Software International), Writer's Workbench (AT&T), Correct Grammar (Lifetree Software) などがある。

ことを目的に開発されたものである。そのうち、スペリング・チェッカは、英語の文書をつくるに当たっての有用性にすぐれ、それは、むしろ当然のこととして綴りの間違いを犯す日本人の英語ユーザにとって、ことのはか利用価値の高いものとなっている。これに対し、文法チェッカやスタイル・チェッカの有用性は、認められる場合でも、狭い範囲のものに限られている。

ソフトウェアの利用価値は、ユーザの特質に依存する。スペリング・チェッカは、だれにとっても便利なものの、日本人の英語ユーザには、なおかつ、その英語の特性を考慮した改良が望まれる⁹⁾。

文法チェッカやスタイル・チェッカは、だれをユーザと見るかで評価が分かれている。文法やスタイル・チェッカの能力は一般にまだ低い⁹⁾。それは(学生の書いた文)の分析能力を欠くから無用であるという見方もある¹⁰⁾。(熟達者の論文や記事に対しても役立つたないので、)だれにとって利用価値があるのか疑問を呈している人もある¹¹⁾。しかし、簡単な間違いしか発見できず、間違いを見つけ損なったり、正しいことを間違いと指摘したりすることがあっても、規範的な英語を書くためには文書作成上で有益なツールであるとの評価もある¹²⁾。

大方の日本人の英語ユーザは、内容はともかくとして、英語圏の学生以下の英文を綴る。われわれは、文法ほかに、訳語選択の問題や、慣用表現(いいたいことを英語でどう表現するか)の問題をかかえている。その特徴を次の文で考えてみよう。

My salary is very cheap.

[私の給料はとても安い。]

They are high degree programs.

[それらは高度なプログラムである。]

My English teacher said to me to write
your opinion.

[英語の先生は私に意見を書けといった。]

日本人は、この類の文をよく書く。しかし、その不規則性(間違い)の検出には、パターン照合をもとにした文法チェッカも、構文分析を行って検査しているといわれる文法チェッカも、何の役にも立たない。(実際、筆者が Correct Grammar を使って行った実験では、日本人の書いた他の多くの文とともに、これらの文もすべて正当なものと判断された。¹³⁾この事実や、文法チェッカが長文や構造の複雑な文には適切に機能しなくなることを考えると、日本人の英語ユーザには、書きたいことを英語でどう書く(表現する)

かを示唆してくるシステムのほうが実用性に優れたツールとなるのではないと思われる。

3. システムの概要

3.1 システムのながれ

正しい英語を文法や単語の知識だけで書くのは困難である。そこで、手紙文を書くときでも、論文を書くようなときでも、われわれは適切な例文を見て、それにならった英文を書くことが多い。(よい英和辞書の条件のひとつに、適切な用例を含んでいるかどうかがある。)本システムは、端的に言えば、この過程を体现するものである。

図1に本システムの概略を示す。ここで、中央の実線で囲んだところがシステムの主要部分である。その中の破線で囲んだ部分で英文作成と、その管理および出力を行う。

検索部は例文の検索をする。レイアウト処理部ではユーザのつくった例文の出力を行う。保存部はユーザのつくった文書をユーザファイルとして保管する。

破線の外にある登録部では例文として使う英文を集める。ここで検索の対象となる例文のデータベースがつくられる。

3.2 例文の検索

本システムの中心は例文の検索過程にある。図2は、ユーザが例文を検索しながら英文書を作成する過程を示す(図中の事例データベースについては4.3節を参照)。データベース中の例文は、文書の種類により一様ではないためと、検索の効率を考慮して、ユーザの定めるカテゴリ別に分類されて蓄えられる。ユーザはまず、作成しようとする文書のカテゴリーを選択する。書きたいと思うことを日本語文で入力すると、システムは、日本語文からキーワードを抽出、それを等価の英文のキーワードに変換した上で例文の検索を行

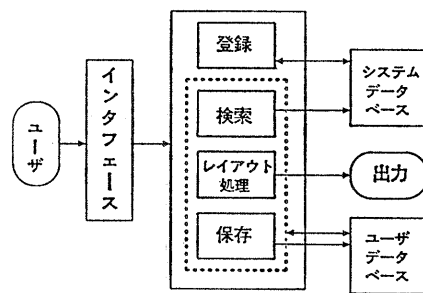


図1 システム全体の構成図
Fig. 1 Organization of the system.

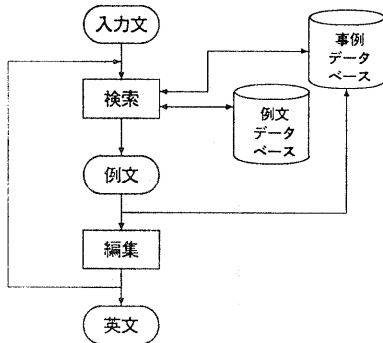


図 2 英文作成の流れ

Fig. 2 Outline of retrieving and tailoring a sample sentence.

う。(すでに作成済みの文章の加筆や訂正ならば、保存部を通してユーザファイルから文書をもってきた上で以上のことを行う。)

4. 例文検索の方法

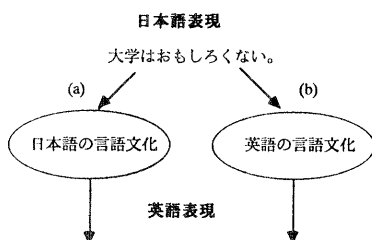
例文検索の性能は、大量の例文から適切な例文を選ぶ過程の実装の仕方によって左右される。われわれのシステムの検索の基本理念は、「日本語文(表現)から核となる意味(アイデア)を取り出し、それと対応する英文のアイデアに変換することによって日本語に対応する英語文(表現)を検索する」というものである。

4.1 日本語表現と英語表現の対応

図3は、日本語表現と英語表現の対応を示すものである。われわれは、しばしば、(a)の過程を経て母国語を他の言語表現にしようとする。そこで、日本人が英語を書くと、

The university is not interesting.

のような意味不明の直訳調の英文をつくってしまう。日本語の意味を体現し、それに対応した英文にするた



The university is not interesting. Lectures at the university are not interesting.

図 3 日本語表現と英語表現の対応

Fig. 3 Relation between Japanese and English expressions.

めには、(b)のような過程を経た英文を書くことが望ましい。そうしないと、「私の給料はとても安い」に対し、

My salary is very low.

I earn a very low salary.

I work for a very low salary.

のような英文を書くことはできない。しかし、この過程を経た英文を書くには、日本語と英語の双方の言語に通じている必要が出てくる。

4.2 キー概念による例文検索

日本語の表現に対応する英語表現を、われわれは亀田¹⁴⁾にあるキー概念の考え方を使得て行う。単語の意味する概念は、キー概念によって表すことができる。たとえば、「感謝」、「深謝」、「謝意」などは、/感謝/というキー概念で表せる。(以下、かき括弧でくくった用語は単語を、スラッシュでくくった用語はキー概念を表す。)この考え方を使得、本システムでは、以下の手順で日本語からアイデアを得、そのアイデアに対応する英文の例文を得る。

- (1) 日本語から(漢字の連続またはカタカナの連続からなる)キーワードを切り出す。
- (2) 切り出した単語をもとにして、いくつかのキー概念を抽出する。
- (3) (2)で得たキー概念を組み合わせて検索式を作成し、それらを画面に表示する。
- (4) ユーザは、表示されたものの中から使用する検索式をひとつ選択する。
- (5) システムは、ユーザが選択した検索式を使得、実際に例文を検索し、結果を画面に表示する。
- (6) 検索の結果、目的の例文を得られなかった場合は(3)に戻る。

以下に、(1)から(6)の過程を実例によって示す。次の日本語文を考える。

「本論文をまとめる上で、Xとの貴重な議論が参考になりました。」

まず、ステップ(1)で、この文からキーワードとして、

「本論文」、「上」、「貴重」、「議論」、「参考」が得られる^{*}。次にステップ(2)により、キー概念、/論文/、/有益/、/議論/

を得る。(ステップ(2)の過程では、後で説明する対

* 現在のところ、キーワードの切り出しは、辞書を使った最長一致によって行っている。

応表を使う。) ステップ(3)では、ここで得たキー概念をもとに、検索式を作成し画面に表示する。この例では、以下のような検索式が表示され、ステップ(4)でユーザはこの中から次の1.の検索式を選択したとしよう。

1. /論文/, /有益/, /議論/
2. /論文/, /有益/
3. /論文/, /議論/
4. /有益/, /議論/
5. /論文/
6. /有益/
7. /議論/

ステップ(5)では、ステップ(4)でユーザが選択した検索式を、実際の検索に使用できる形(キーワード式)に変換する。そのためにシステムはまず、ユーザが選択した検索式に含まれるキー概念をそれぞれ意味的に対応する英単語に変換する。(この過程でも後述の対応表を使用している。)

/論文/ → paper, draft
 /有益/ → useful, helpful
 /議論/ → discuss

これらの単語を同じキー概念を表す単語どうしはORで、異なるキー概念どうしはANDで組み合わせる。また、複数形や過去形などを同時に検索できるように、単語の後ろにアスタリスクを付け、前方一致を行うようにする。この例で、ユーザの選択した検索式(キー概念の組み合わせ)から作成された検索用のキーワード式は次の通りである。

{/論文/, /有益/, /議論/} →
 (paper* | draft*)
 & (useful* | helpful*)
 & discuss*

さて、ステップ(2)とステップ(5)の実行のために、システムは二種類の対応表を使用している。対応表とは、キー概念と日本語の単語の対応およびキー概念と英単語の対応をもたせた表である。表1は、文書カテゴリを、論文を書くときに使う謝辞にとった例で作成した対応表の初期値の一例である。(ユーザが入力した日本語文によって、システムが適切なキーワードを抽出できない場合は、ユーザにキーワードを直接要求することによって検索を行う。)

ユーザの入力した単語がすでに対応表に登録されている場合、システムはそれを使って対応表から検索用の式をつくり、例文を検索する。単語が対応表に登録

表1 キー概念と単語の対応表
 Table 1 Key concepts and corresponding words in English.

(a) 日本語単語との対応例
 (a) An example for Japanese words.

キー概念	対応する単語
/論文/	「論文」「本稿」「本論文」
/有益/	「有益」「貴重」「有効」
/助言/	「助言」「意見」
/感謝/	「感謝」「深謝」「謝意」
/研究/	「研究」「本研究」
/議論/	「議論」「討論」
/システム/	「システム」
/実装/	「実現」「開発」「実装」「作成」
/査読者/	「査読者」
/協力/	「協力」

(b) 英単語との対応例
 (b) An example for English words.

キー概念	対応する英単語
/論文/	paper, draft
/有益/	useful, helpful
/助言/	comment, suggest
/感謝/	thank, grateful, gratitude
/研究/	research
/議論/	discuss
/システム/	system
/実装/	implement
/査読者/	referee, reviewer
/協力/	cooperate

されていない場合には、和英辞書を使い、例えば、ユーザが「指導」を入力したとすると、それに対応する英単語“guidance”を得る。この単語から検索用の式“guidance*”をつくり、例文データベースから例文を検索する。そのあとで、システムはこの単語を新しく対応表に登録する。これで、この用語は以後の検索に使用可能となる。(対応表に不要な単語が登録されないようにするため、新しいデータの登録は、個々のユーザの確認を取りながら対話的に行われる。)

対応表はあらかじめ用意するが、データの登録は不十分なものであってもよい。システムを使用しているうちにデータが追加されていくからである。実際、対応表に追加登録される単語は、ユーザが検索に使用した単語なので、対応表は検索に必要な単語のみを登録していくことになる。

ステップ(1)から(5)によって検索した結果、適切な例文を得られなかった場合は、ステップ(3)に戻って別の検索式を選択し、検索をやりなおす。

画面に表示されたいずれの検索式を使っても検索が

うまくいかない場合、ユーザは直接単語を入力し、検索式を作りなおす。例えば上の例で、ユーザが「意見」という単語を入力すると、システムは、この単語を/助言/というキー概念に変換し、先にシステムが日本語文から抽出した三つのキー概念/論文/、/有益/、/議論/とあわせた四つのキー概念をもとに、ステップ(3)以降の処理を行う。

4.3 類似事例の利用

4.2 節で述べた検索方式では、場合によっては最適な検索結果を得るまでに何度も検索を繰り返す必要がある。また、画面に表示された大量の英文の中から適切なものを選びだす作業は、ユーザにとって大きな負担となる。

日本語を英訳するときのことを考えてみると、意味的に類似した日本語文は、文の形や使われている単語が違っていても、結果的には同じような英文に訳されることが多い。以前に意味の類似した日本語文を英訳した経験がある場合は、そのとき作成した英文を参照することで簡単に英文を作成できると考えられる。

本システムでは、ユーザが作成した英文と、そのもともとなった日本語文をひとつの事例として、事例データベースと呼ばれるデータベースに登録していき、新しく日本語文が入力された際、以下の手順で過去の類似事例を検索し、ユーザに提示する。

- (1) 入力文からキー概念を抽出する。
- (2) (1)の結果を利用して、事例データベース中の日本語文と入力文との「類似度」と「一致度」を計算する。
- (3) (2)の結果から、入力文と事例との類似性を評価し、類似性の高い事例から順に表示する。

ステップ(1)の処理は、4.2 節で述べたものと同じである。ステップ(2)の「類似度」と「一致度」は、以下の式によって求められる。

$$\langle \text{類似度} \rangle = k/m$$

$$\langle \text{一致度} \rangle = k/n$$

k : 一致するキー概念の数

m : 入力文から抽出したキー概念の数

n : 事例の日本語文から抽出したキー概念の数

例えば、入力文から抽出したキー概念が {/査読者/ /論文/、/助言/} で、事例データベース中のある事例から抽出したキー概念が {/論文/、/助言/} だった場合、これら二つの類似度は $2/3=0.67$ 、一致度は $2/2=1.00$ である。類似度の値が大きいほど、入力文と

類似した事例であるとする。類似度が同じ値の場合は、一致度の高いものほど類似していると評価する。

システムが、入力文から類似事例を検索する様子を見てみることにする。例えば、次の日本語文を入力したとする。

「本稿を執筆するにあたり種々議論していただいた X, Y に感謝します。」

システムはまず、この文から次のようなキー概念を抽出し、

/論文/、/議論/、/感謝/

これをもとにして類似事例を検索する。その結果、以下のような事例が得られる。ここでは、各事例について、検索された事例の日本語文と英文、日本語文から抽出されたキー概念、入力文と事例との類似度と一致度を示す。

1. 「また、貴重な討論を絶やさないう Y, Z 両氏に心より感謝します。」

キー概念 : {/有益/、/議論/、/感謝/}

類似度 : $2/3=0.67$ 、一致度 : $2/3=0.67$

作成文 : The author also would like to thank Y and Z for several useful discussions that have helped us formulate the ideas expressed in this paper.

2. 「議論に加わっていただいた Y 研究所員の皆さん、特に、有益な助言をいただいた X2, X3 に感謝します。」

キー概念 : {/議論/、/有益/、/助言/、/感謝/}

類似度 : $2/3=0.67$ 、一致度 : $2/4=0.50$

作成文 : I wish to thank all those members of Y Laboratory for many stimulating discussions, and especially thank X2 and X3 for their helpful comments.

3. 「また、本論文の構想をまとめる上で、同研究所内で貴重な議論がなされました。」

キー概念 : {/論文/、/構想/、/有益/、/議論/、/研究所/}

類似度 : $2/3=0.67$ 、一致度 : $2/5=0.40$

作成文 : We would like to thank members of Y Laboratory for several useful discussions that have helped us formulate the ideas expressed in this paper.

表示された事例の中から適切な例文がある場合には、それを利用して英文を作成する。この例では、ユーザは上記の中から事例 3. を選択し、英文の一部

を修正して次のような文を作成した。

We would like to thank X and Y for several useful discussions that have helped us formulate the ideas expressed in this paper.

検索結果の中に利用できる例文がない場合には、4.2 節で述べた通常の検索方式によって例文を検索し、英文を作成する。

5. 英文の作成例

本システムは、実験データとして、二つのカテゴリに属する例文を使っている。それらは謝辞(Acknowledgments)に関するものと著者紹介(About the Author)に関するものである*。

以下に、一ユーザがシステムを実際に作動させてつくった文書例を示す。謝辞を例にとる。いま、次の日本語の謝辞を英語で書きたいとする。

「本研究を進めるにあたり、助言していただいた X1 に深謝します。また、システム的设计・実装に精力的に携わっていただいた X2, ならびに X3, X4 に感謝します。」

この日本語文をもとに、システムが英文書作成の支援をする過程は次の通りである。この処理は 4.3 節で説明した例と同様なので、説明は省略する。

Sentence 1:

単語: 「研究」, 「進」, 「助言」, 「深謝」

index: {/研究/, /助言/, /感謝/}

類似事例:

1. 「特に、X2, X3 は著者たちと共同で本論文の研究をされたのでここに謝意を表したい。」

index: {/共同/, /論文/, /研究/, /感謝/}

類似度: $2/3=0.67$, 一致度: $2/4=0.50$

作成文: The authors appreciate many helpful conversations and discussions with X and members of Y Laboratory.

2. 「日頃ご指導いただく X 氏には有益なご意見をいただき深謝します。」

index: {/指導/, /助言/, /有益/, /感謝/}

類似度: $2/3=0.67$, 一致度: $2/4=0.50$

作成文: The author wish to express his gratitude to X for kind guidance and his

most helpful comments.

3. 「議論に加わっていただいた所員の皆さん、特に、有益な助言をいただいた X2, X3 に感謝します。」

index: {/議論/, /有益/, /助言/, /感謝/}

類似度: $2/3=0.67$, 一致度: $2/4=0.50$

作成文: The authors appreciate many helpful conversations and discussions with X and members of Y Laboratory.

ユーザはこの中から事例 2. の例文を選択し、そのまま利用した。

Sentence 2:

単語: 「システム」, 「設計」, 「実装」,

「精力的」, 「携」, 「感謝」

index: {/システム/, /開発/, /感謝/}

類似事例:

1. 「ご討論いただいた X1 氏, プロトタイプシステムの開発に協力いただいた X2, X3 に感謝いたします。」

index: {/議論/, /システム/, /開発/,

/協力/, /感謝/}

類似度: $3/3=1.00$, 一致度: $3/5=0.60$

作成文: I am grateful to X1 for extensive discussions, to X2 and X3 who have contributed to the development of the prototype system.

2. 「このシステムの実現には X 君と Y 君の協力を得た。」

index: {/システム/, /開発/, /協力/}

類似度: $2/3=0.67$, 一致度: $2/3=0.67$

作成文: The author would like to acknowledge the contributions of X and Y in the design and implementation of this system.

ユーザはこのうち事例 2. の英文にまず次のような変更を加えた。

would like to → also would like to

X and Y → X2, X3 and X4

その結果、ユーザが最終的につくった文書は次のものである。

The author wish to express gratitude to X1 for kind guidance and his most helpful comments. The author also would like to acknowledge the contributions of X2, X3 and X4 in

* この二つを選んだのは、例文がとりやすかったからである。実験には、それぞれについて、およそ 100 ずつの文章（一つの文章は短いもので数文、長いものでは十数文からなる）を使っている。データは諸種の学術雑誌からとった。

the design and implementation of this system.

文書作成例をもう一つ、実際の著者紹介欄にあったものをあげる。この例では、事例ベースに類似した事例がなかった場合の処理の様子を示そう。ただし、ここでも詳細は略し、1)システムが入力文から切り出した単語、2)システムが抽出したキー概念、3)例文の検索に成功した検索式、4)利用した例文と、最終的に作成した文書だけを示す。

入力文：

「1977年X大学A学部B学科卒業。1979年同大学院修士課程修了。1980年Z研究所入所。1989年X大学A学部B学科助教授となり、現在に至る。自然言語処理、論理プログラミングの研究に従事。」

Sentence 1-2:

単語：「大学」、「学部」、「学科」、「卒業」

「大学院」、「修士課程」、「修了」

キー概念：{/大学/, /学部/, /卒業/, /修士/}

検索式：</大学/, /卒業/, /修士/>

例文：A graduate of Y University, X holds a M. A. degree in Mathematics.

Sentence 3:

単語：「年」、「研究所」、「入所」

キー概念：{/研究所/, /入所/}

検索式：</研究所/>

例文：In 1980 he joined the XX Corporation as a member of the research planning staff.

Sentence 4:

単語：「年」、「大学」、「学部」、「学科」、

「助教授」、「現在」、「至」

キー概念：{/大学/, /学部/, /教授/, /現在/}

検索式：</大学/, /学部/, /教授/, /現在/>

例文：An instructor for 10 years, Allen is currently an Associate Professor in the Computer Science Department at the University of Rochester.

Sentence 5:

単語：「自然言語処理」、「論理」、「プログラミング」、「研究」、「従事」

キー概念：{/研究/, /従事/}

検索式：</研究/, /従事/>

例文：Since 1976, X has been engaged in research on distributed processing; in this area his main interests are oper-

ating systems, programming languages, and algorithms.

作成文書：

A graduate of X University, Taro Yamada holds a M.Sc. Degree in engineering. In 1980 he joined the Z Laboratory. He is currently an Assistant Professor in the Computer Science Department at the University of X. He has been engaged in research on natural language processing and machine translation.

図4はシステムの動作中の画面例である。ここでは、上部がユーザ用の編集画面、下部がシステム画面となっている。この図は、上の文書作成例のうち、4番目の文を作成する途中で、システム画面に表示された検索結果の中から、ユーザが1番目の例文を選択し、ユーザ画面にコピーした状態を示したものである。

6. 現状と評価

本システムは、C言語と Emacs Lisp を用い UNIX ワークステーション上に実装されている。ユーザとのインタフェース部には Emacs エディタを使っている。ユーザはこのエディタにより文書の作成、例文の検索を行う。プログラムは、C言語で約5,000行、Emacs Lisp が、コンパイル後で約1,000行である。

本システムの検索方式では、キー概念と組合せキー概念を用いることにより、意味処理を行わずに意味を

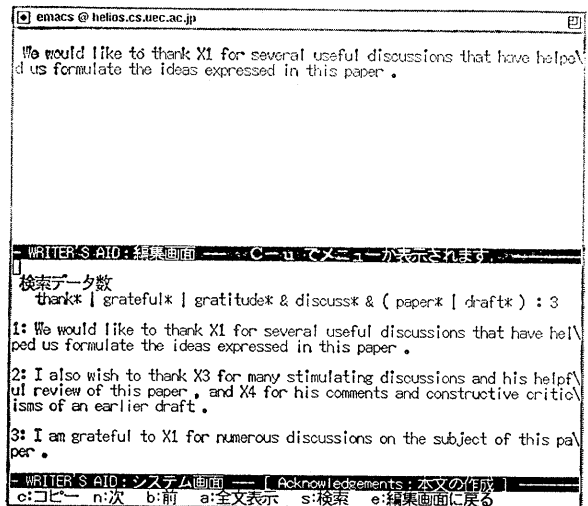


図4 システムの画面例

Fig. 4 A screen for retrieving sentences.

反映した検索を行えることのほか、事例ベース推論の過程で検索結果を学習する機能をもたせているため、検索を繰り返すほど検索効率があがるという長所がある。

一般に、キーワードを使っての情報検索には、再現率（抽出したデータの中にある必要データの、抽出したデータの中にある必要データと抽出できなかった必要データとを加えたものに対する割合）と適合率（抽出したデータの中にある必要データの、抽出できなかった必要データと、抽出した不必要なデータとを加えたものに対する割合）の問題がある¹⁵⁾。対応表を利用した検索は、入力文と意味的に関係の深い単語だけを探索的に反映させるため、シソーラス等を用いた通常の方式に比べて不要な例文が検索されにくく、そのことが適合率の向上につながっている。ユーザにとっては、適合率が高いほど、画面に表示された多くの例文の中から必要な例文を選び出す際の負担が少なくなる。

一方、本システムは、多くの文書検索システムと同様、キーワードの抽出方法にもっと意を注ぐ必要がある。入力となる日本文からのキーワードの抽出は、システムの能力と直接に結びつく。どの単語をキーワードにするかという問題に加えて、日本語の文章では、言語特有なものとして、キーワードの切り出し（複合語の分割）の問題（例えば「修士課程」は、「修士」と「課程」か「修士課程」か。）が加わる^{16),17)}。システムは、入力文にキーワードがない場合には、適当なキーワードをユーザに要求した上で次段階に進むようになっているが、往々にしてユーザの意に反した例文を検索してしまう。しかし、この難点をなくすためには、文中からのキーワードの選択法を改良する方向とともに、グローバルやキーワードの抽出方法を考える必要もあると思われる。たとえば、次の文章の最初の文からは、図3にあげたような英文例の検索はむずかしい。

大学はおもしろくない。授業も友だちとのつきあいもつまらない。教師は教育には興味がないようだ。

本システムでは、まず、ユーザに適切なキーワードを要求することでこの難点の回避をはかろうとしているが、これには限界がある。結束性や首尾一貫性といった単語間や文間にある言語現象の分析の成果^{18),19)}をとり入れると、システム側で例文検索のためのキーワード（ここでは、「大学」、「授業」、/生活/ といったもの）を自動的に抽出する方法も考え出せるはずで

ある。これができれば、文章全体の意味との関係によって、さらに適切な例文を引き出せるようになるだろうと思われる。

7. おわりに

本稿では、英文書作成支援システムの概要、例文検索方法、システムの作動例（文書作成例）を述べた。意味処理なしに日本語の文から等価の意味をもつ英文を作成するために、例文を使うことにし、対応表を使っての検索を行った。

このシステムでは、実験に二つのカテゴリの文書を使ったが、これはユーザの扱う文書の種類に依存して変えることができる。たとえば、カテゴリは種類別にした手紙文のようなものであってもよい。カテゴリの選択が適切に行われ、十分な例文をとることができれば、システムは有効に作動し、その実用性は高い。

謝辞 内容、表現、構成にいたるまで、査読者から多くの指摘と、詳細なコメント、提案を得ました。記して深く感謝いたします。

参考文献

- 1) MacDonald, N. H., Frase, L. T., Gingrich, P. and Keenan, S. A.: The Writer's Workbench: Computer Aides for Text Analysis, *IEEE COM*, Vol. 30, No. 1, pp. 22-48 (1982).
- 2) Heidorn, G. E., Jensen, K., Miller, L. A. and Byrd, R. J.: The EPISTLE Text-Critiquing System, *IBM Systems J.*, Vol. 21, No. 3, pp. 305-327 (1982).
- 3) 牛島: 日本語文書推敲支援ツール『推敲』, *Bit*, Vol. 23, No. 1, pp. 4-14 (1991).
- 4) 林ほか: 日本推敲支援システムにおける書換え支援機能の実現方式, *情報処理学会論文誌*, Vol. 32, No. 8, pp. 962-970 (1991).
- 5) 箱守ほか: 日本語の修飾構造を評価する添削支援システムを実現するための基礎研究, *情報処理学会論文誌*, Vol. 33, No. 2, pp. 153-161 (1992).
- 6) 奥村, 脇田, 金子: 日本語校正支援システム FleCS の新聞社における実用化, *情報処理学会自然言語処理研究会資料*, 91-5, pp. 33-40 (1992).
- 7) Takeda, A. and Furugori, T.: An Example-Based Systems of English Writing Aid for Japanese, *Proc. of 1st Pacific Association for Computational Linguistics Conference*, Vancouver, B. C. (1993).
- 8) Furugori, T.: Improving Spelling Checkers for Japanese Users of English, *IEEE PC*, Vol. 33, No. 3, pp. 138-142 (1990).
- 9) Bowyer, J. W.: A Comparative Study of Three Writing Analysis Programs, *Literary*

- and *Linguistic Computing*, Vol. 4, No. 2, pp. 90-98 (1989).
- 10) Hazen, M. et al.: *Report on Writer's Workbench and Other Writing Tools*, Univ. of North Carolina Microcomputing Support Center, Chapel Hill (1986).
- 11) Dobrin, D.N.: A New Grammar Checker, *Computers and the Humanities*, Vol. 24, pp. 67-80 (1990).
- 12) Lombardi, J.: Grammar Checkers: Keep Your Writing in Line, *INFOWORLD*, Oct. 28, pp. 64-77 (1991).
- 13) Furugori, T. and Takeda, A.: An Example-Based System of Writing English Sentences for Japanese English Users, *Literary and Linguistic Computing*, Vol. 8, No. 2, pp. 85-90 (1993).
- 14) 亀田, 藤崎: テーマ・キー概念・キーワード間の階層構造を利用する新聞記事情報の分類・検索システム, *情報処理学会論文誌*, Vol. 28, No. 11, pp. 1103-1111 (1987).
- 15) Driscoll, J.R. et al.: The Operation and Performance of an Artificially Intelligent Keywording System, *Information Processing & Management*, Vol. 27, No. 1, pp. 43-54 (1991).
- 16) 内山, 板橋: 共起関係を利用した日本語複合名詞の分割, *情報処理学会自然言語処理研究会資料*, 91-7, pp. 47-54 (1992).
- 17) 水野ほか: 日本語キーワードの自動抽出法, *情報処理学会自然言語処理研究会資料*, 91-6, pp. 41-46 (1992).
- 18) Morris, J. and Hirst, G.: Lexical Cohesion

- Computed by Thesaural Relations as an Indicator of the Structure of Text, *Computational Linguistics*, Vol. 17, No. 1, pp. 21-48 (1991).
- 19) Kozima, H. and Furugori, T.: Similarity between Words Computed by Spreading Activation on an English Dictionary, *Proceedings of 6th Conference of the European Chapter of the Association for Computational Linguistics*, Utrecht, pp. 232-239 (1993).

(平成4年11月2日受付)

(平成5年10月28日採録)



武田 明子

平成5年3月電気通信大学情報工学科大学院修士課程修了。横河システムエンジニアリング(株)に勤務。自然言語処理に興味をもつ。



古郡 廷治 (正会員)

昭和14年生。ロチェスター大学大学院修士課程(言語学), ニューヨーク州立大学大学院博士課程(計算機科学)修了。Ph. D. クリーブランド州立大学助教授を経て, 現在, 電気通信大学情報工学科教授。計算言語学, 言語情報処理, CAIの研究に従事。著訳書「文字列処理のための計算機科学概論」, 「プログラムのうちあけ話」など。ACM, 計量国語学会各会員。