自然言語処理　150－2
(2002. 7. 15)

TransAid
－英文書作成支援システム－

高倉 佐和† 古郡 廷治‡

†十文字学園女子大学 〒352-8510 埼玉県新座市菅沢 2-1-28
‡電気通信大学 〒182-8585 調布市調布ヶ丘 1-5-1
E-mail: †takaku-s@jumonji-u.ac.jp, ‡furugori@cs.uec.ac.jp

あらまし 本稿は，日本人が適切な英語を書くための文書作成支援システムについて述べる．このシステムは，日本語の文章と，市販の機械翻訳システムによるその翻訳例を入力とする．次いで，翻訳システムの出力を訂正したり，洗練させたりして，特定の目的に合った英語にするため，有用な英語の例文をインターネット・コーパスから抽出する．会議の通信文やその他の文章の実験では，当システムは，労力，スタイル，内容の点で，人間または機械によって書かれた文章を著しく改善することがわかった．

キーワード 文書作成支援システム，インターネットコーパス，和英翻訳，データベース，情報抽出

TransAid
－A Writer's Aid System for Translating Japanese into English－

Sawa TAKAKURA† and Teiji FURUGORI‡

† Faculty of Jumonji University 2-1-28 Kanzawa, Niza-shi, Saitama, 352-8510
‡ Dept. of Computer Science, the Univ. of Electro-Communications 1-5-1 Chofugaoka, Chofu-shi, Tokyo, 182-8585
E-mail: †takaku-s@jumonji-u.ac.jp, ‡furugori@cs.uec.ac.jp

Abstract This paper describes a writer's aid system that helps the Japanese write sound English sentences. Our system takes as input a Japanese writing and its English translation produced by a commercial machine translation system. It then extracts, from the Internet Corpus, English sentences and expressions usable for correcting and/or polishing the translation to suit our specific purpose. An experiment for writing correspondences in conferences and others shows that the system is easy to use and improves the writings by machines or humans considerably in effort, style, and content.

Keyword Writer's Aid System, Internet Corpus, Japanese to English Translation, Database, Retrieval

1. はじめに

近年，英語を母国語としない人が英語を書いたり読んだりする必要性がますます高くなってきている．その必要性を反映し，商用日英機械翻訳システムや翻訳メモリが数多く開発されている[3, 4, 13, 14]．英文の文章作成支援では，誤綴の訂正[5, 7, 8]，主語と動詞の一致や名詞の一致などを含む文法のチェック[11]，用語や文の長さ，構文に関するスタイルのチェックなどのツールをはじめとして多くの研究がある[6, 16, 17, 18, 19]．しかしながら，大方のシステムやツールでは作成できる文章の分野が限定されていたり，英文の質がまだ低かったりして，ユーザに満足のいくソフトウェアとなっていない．

本論文では，より適切な英文作成を支援するためのシステムを提案する．本システムは，日本語の原文と商用機械翻訳システムにより訳された英訳文を入力とする．それらを利用して，インターネットコーパスから例文を抽出し，英訳文を，より自然な表現に改善するために用いる．ユーザには例文を効果的に用いることによって，労力，スタイル，内容の点で，よりよい英文の作成が可能となる。

2. 人間の翻訳と機械による翻訳

日本人は多年にわたって英語を学習してきているにもかかわらず，英文を書くとき多くの誤りを犯す．相当の知識を持っている人でも，適切な英語表現を書くのは容易なことではない．機械翻訳システムや翻訳メモリを利用しても、実用のために，その出力を修正したり加工したりするのは大変な作業である．

2.1. 英文の質の比較

図 1(b)は日本人が図 1(a)の日本語を英訳した例である．ここに，日本人が犯しやすい典型的な誤りがある．たとえば，*manuscript* (本稿), *test* (実験), *labor* (労力)などの単語選択，*in order to make up for proper English* (目的に合った英語にするため)の句表現は不適切である．また，*in terms of the labor, style, and contents* (労力，スタイル，内容の点で)は用法的に誤りがある．

図 1(c)は市販の翻訳ソフトの 1 つを使って，図 1(a)を入力したときの出力である．ここでもまた，語彙レベル，構造，意味的に誤りを指摘することができる．

2.2. 利点と欠点

機械翻訳システムの利点は英文作成の時間を節約できることにある．10 人の日本人学生による予備実験を行った結果では，図 1 の英文を作成するのに，約 2 時間かかった．一方，機械翻訳システムは，英文の質はともかく，数秒のうちに翻訳結果を出力する．しかし，機械翻訳システムでは，その訳の意味不明さと同時に、辞書に未登録の単語が入力されると，原文のままの単語を出力してしまうという欠点もある．

3. インターネットコーパスの例文を利用した翻訳メモリ

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．手紙文を書くときでも，論文を書くときでも，われわれは適切な例文を見て，それに倣って英文を書くことが多い．本英文書作成支援システムは，英文を作成するときに必要な，生きた英語をインターネットコーパスから抽出した例文により提示する．

3.1. 処理の流れ

図 2 はシステムの流れ図である．本稿では点線で囲まれた部分を中心的に述べる．TransAid は入力された文書(*JS*)に含まれる各文に対し，前処理として，複合語の自動分割を行う．さらに，前処理後の日本文(*JS'*)と英文(*ES'*)の双方から抽出したキーワードをもとに，データベース(*DB*)から適切な例文を抽出する．ここで，*JS'* と *ES'* は機械翻訳システム(*MT*, 本稿では TOSHIBA, The Honyaku Professional V7.0 を用いる)への入力と出力である．ユーザは，TransAid から提示される例文をもとに，*ES'*を洗練し，修正する．図中の *E1* は *ES'*を改めた後の出力，*E2* は *MT* を使って人が作成した英文，*E3* は *MT* や TransAid を用いずに作成した英文を意味する．

(a) 本稿は，日本人が適切な英語を書くための文書作成支援システムについて述べる．このシステムは，日本語の文章と，市販の機械翻訳システムによるその翻訳例を入力とする．次いで，翻訳システムの出力を訂正したり，洗練させたりして，特定の目的に合った英語にするため，有用な英語の例文をインターネット・コーパスから抽出する．会議の通信文やその他の文章の実験では，当システムは，労力，スタイル，内容の点で，人間または機械によって書かれた文章を著しく改善することがわかった．

(b) This manuscript states the Writing Support System to let Japanese write a proper English. You input original Japanese sentences and their English sentences translated by system on the market into this system. And in order to make up for proper English sentence to match the specific purpose by the correction and the refinement of the output, they look up appropriate examples of English sentence through Internet corpus. As results of our test by sample of correspondence for conference and other sentences, it found that the translation of the sentence written by human or other system was remarkably improved by introduction of our system in terms of the labor, style and contents.

(c) This paper describes a document creation support system for Japanese people to write suitable English. This system considers that example of translation by the Japanese text and the commercial machine translation system as an input. Subsequently, correct the output of a translation system or it is made to refine, and in order to make it English suitable for the specific purpose, the example of useful English is extracted from the Internet corpus. In the experiment of the correspondence of a meeting, or other texts, it turned out that this system improves remarkably the text written by the human being or the machine in respect of a labor, a style, and the contents.

図 1: 日本人と機械翻訳システムにより作成された英文

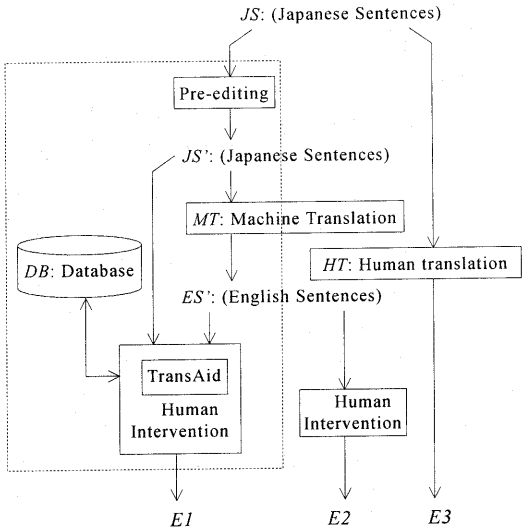


図 2: TransAid の処理流れ図

TransAid の性能は例文を蓄える *DB* の構造と適切な例文や表現を検索するためのキーワードの選択,また,どのように例文を抽出するかに左右される. *DB* のサイズが適切でなかったり,収集された例文の分野に偏りがあると,提示される例文の数が多すぎたり,少なすぎたり,また,無関係の例文を表示したりする.

3.2 インターネットコーパス

インターネットコーパスは動的にテキストを蓄積できる,世界最大のコーパスである[10, 12, 23]. 最もよく使われる検索エンジンの 1 つである Altavista を用いると,少なくとも 3 億 5 千万の英語のページにアクセスできることがわかる[1].

本システムに蓄える例文データベースは,有効な例文を抽出するために,十分に大きいサイズでなければならない. 同時に,無関係な例文の抽出を避けるためには,書かれる文章の分野にあったデータベースが必要である. そのため, TransAid は必要な例文をインターネットコーパスから抽出し,テキストデータベースを構築する. 本稿では,学会関係の英語テキストを例データベース構築の手続きを示す:

1. Search the Internet corpus with a name of academic societies and the word *conference* and get the pages, *P*. Repeat this process with a few different names and make *P* bigger.
2. Sort nouns in *P* using their frequencies and put them in the list, *M*.
3. Get the list for general texts, *G*, that is equivalent to *M* for all the nouns in the Internet.
4. Get the domain words, *D*, used in *M* as $D = M - (M \cap G)$
5. Get the clue words, *W*, from *D* as $W = D$ when $D \leq 10$, otherwise *W* = (the words of 0.5% in *D* starting from the one with the highest frequency).
6. Search the Internet Corpus with *W* and get the pages *P'*, that becomes the database, *DB*.

実験では,ステップ 1 に対して, 10 の学術学会の URL をランダムに抽出した. その結果, *M* と *G* において,最も頻繁に現れる語の上位 10 語を表 1 に示す. *M* と *G* から *D* を獲得し, *W* として次の 15 語 *W*={*abstract, author, conference, deadline, edu, form, paper, proceeding, science, sponsor, submission, symposia, symposium, university, workshop*} を得た. ただし,ステップ 2 とステップ 3 において,数詞,代名詞,固有名詞は対象外とする. 最終的に,上記 *W* をもとに, *P'* に含まれる文章から, 6948491 語のテキストデータベース *DB* を得た. 表 2 は *DB* に頻出する単語である.

表 1: 学術学会と一般のテキストに含まれる語

	Conference		General	
	Word	Frq	Word	Frq
1	Conference	719	document	195
2	Paper	423	site	182
3	Information	258	all	113
4	Submission	222	search	104
5	Contact	211	URL	80
6	Program	210	online	72
7	e-mail	187	home	71
8	international	178	top	70
9	Author	163	accessory	65
10	Workshop	161	estate	60

表 2: 学会データベースに頻出する語

	Word	Frq.
1	information	11686
2	year	11125
3	state	9665
4	student	8975
5	system	8878
6	paper	8102
7	work	7706
8	People	7672
9	service	7421
10	Report	6578

3.3 複合語の自動分割

現存の機械翻訳システムは,辞書に未登録の複合名詞を翻訳する際,個々の構成要素の名詞を翻訳した結

果を直結して出力する．例えば，図 1(a)中の *文書作成支援システムは document(文書) creation(作成) support(支援) system(システム)* と翻訳されている．しかし，そのままの英単語を用いるとしても，*support system for(/of) document creation* の方がよりよい．

表 3 は、学会に関する日本語 Web ページ 794 ページに含まれる名詞複合語の分析を行い、複合語の数と辞書登録比率をみたものである。名詞複合語は構成要素となる名詞の数が増えるほど，辞書登録率が低い．これらの名詞複合語を適切に翻訳するには，何らかの処理が必要なことがわかる．

表 3：名詞複合語の分析

n 名詞複合語	数	辞書登録率(%)
n=1	12034	54.1
n=2	21659	38.7
n=3	13293	5.7
n=4	7751	1.1
n=5 以上	10579	0.3

(注: n は構成要素となる名詞の数を示す．)

TransAid はまず，下記のアルゴリズムにより，複合名詞の要素間の依存関係を解析し[9, 22]，複合名詞を自動分割する．

1. Apply morphological analysis to *JS*.
2. Find the sequence of nouns and verbs, $W=(w_1,w_2,...,w_n)$, in the sentence $S \in JS$.
3. For any compound noun, $w_m(1 \le m \le n)$, being consisted of $c_1c_2...c_k$, do the following.

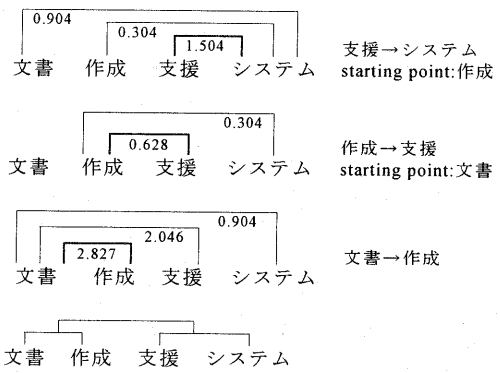
```
Main{
  For each c_i in c_1c_2...c_{k-1}, set {mark[c_i]=0}.
  call R_search(c_1c_2...c_i...c_k)
}
Sub R_search(c_1c_2...c_i...c_k){
  If (exist c_i|{mark[c]=0, i∈(1...k-1)}){
    Compute and find the smallest IMI(c_i,c_k)|{mark[c_i]=0∧i∈(1...k-1)} and let it be IMI(c_mua,c_k)
    Compute and find the largest IMI(c_mua,c_j)|{j∈(mua+1...k)} and let it be IMI(c_mua,c_mac)
    If IMI(c_mua,c_mac)<IMI(c_mua,c_k) set mac=k
    Acquire a dependency relation c_mua → c_mac
    set mark[c_mua]=1
    call R_search(c_mua+1...c_mac)
    call R_search(c_1...c_mua c_mac...c_k)
  }
}
```

アルゴリズム内の *IMI* はインターネットのテキストデータを利用した相互情報量で，次式により定義する．

$$IMI(c_1,c_2)=\log\left(\frac{N\times hit(c_i\cdot AND\cdot c_2)}{hit(c_1)\times hit(c_2)}\right)$$

ここで，*hit(c)*はインターネット検索エンジンを用いて，単語 c を検索したときのヒット数である．$hit(c_i\cdot AND\cdot c_2)$は同じページ内に c_1 と c_2 が出現しているヒット数である．N は 4200 万とした．これは，本実験で使用する検索エンジン Goo からアクセスできる日本語の URL のおよその数である[9].

図 1(a)中の *文書作成支援システム*を例に，複合語の自動分割の流れを示す．



複合語の自動分析の結果得られた，（（(文書)（作成)）（（支援）（システム)))の構造から，分割された各構成要素の名詞を“の”で接続することにより，英訳文の語順を改善するための前処理を行う．(文書作成)の(支援システム)を翻訳システムに入力すると，*The support system of document creation* を得る．

4. Translations with TransAid

TransAid は前処理後，*JS'*から選択されたキーワードをもとに，*DB* から適切な英文を検索し，提示する．

4.1. キーワード選択

良い例文を抽出するために，キーワード選択は，重大なポイントである．キーワード選択は以下の手続きによって行う：

1. Assign for each word in W the word meanings (senses), $M=(m_1,m_2,...,m_m)$, in English.

2. Reduce the senses of each word in W to two by choosing the two senses in M that have the higher frequencies in DB.
3. Find the Cartesian combination, $C = w_1(M) \times w_2(M) \times \cdots \times w_n(M)$ of the senses for W and make each of combination of words to be a set of clue words.

ここに，W として，3つの名詞 x, y, z と動詞 d がこの順序で出現し，和英辞書を検索すると，x に1つ，y に2つ，z に3つ，d に4つの語義が登録されていると仮定する．W に含まれる各語の解釈の候補は，上限を2とし，選択する．ここで，1x2x2x2 の組み合わせで，C に対して最大，8つの解釈が成り立つ．C の解釈は x, y, z, d の4語の各語義からなっていて，例文を抽出するためのキーワード候補の組となる．

4.2. 例文抽出

キーワード候補の組を用いて，大量の例文から適切な例文を選ぶ過程も、例文抽出の性能を大きく左右する．キーワードの候補の組が与えられると，TransAid は DB から例文を抽出するための表現を次の手続きにより公式化する：

1. Formulate the expressions and extract the sentences from DB that contain all the words in each reading in C.
2. Select the words in the reading, $K = (w_1(m_i), w_2(m_j), ..., w_n(m_r))$, where $1 \le i, j, ..., r \le m$, that extracted the greatest number of sentences.
3. Find, if possible, the nouns and verbs, $W' = (w'_1, w'_2, ..., w'_m)$, in the sentence $S' \in ES$ that would correspond to S.
4. Get the keywords , K', as $K' = K \cup W'$.
5. Calculate the similarity rate (SR) and the matching rate (MR) for each sentences extracted using K' and select a few sentences with the highest combined rates to be used for writing and/or correcting English sentence for S. If all the sentences thus selected are not usable, then show the sentences with the next highest rates, if necessary, and continue so on.

例えば，"describe"がキーワードとして選ばれたと仮定すると，例文検索式には，語幹 describ* を用いる．これは，(describe | describes | described | describing)を意味している．さらにすべてのキーワードを(&)で接続する．もし，実際のキーワードが paper, research, computer, science, describe とすると，例文検索式は(paper* & research* & computer* & science* & describ*)になる．

W'に対して，paper, manuscript, describe の3語を得たと仮定すると、次の6語のキーワード群、paper, manuscript, research, computer, science, describe を用いて類似度と一致度を計算する．類似度(SR)と一致度(MR)は以下の式によって求める：

$$SR = \frac{k}{m}, \qquad MR = \frac{k}{n},$$

ここで，k は一致するキーワード群と例文に共通の語数，m はキーワード群に含まれる語数，n は例文に含まれる語数である．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．また，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事例であると評価する．

文抽出ステップ1とステップ5に失敗し，適切な例文が提示されない場合，ユーザは2つのオプションを選択することができる．自分でキーワードを入力するか，新規に W を獲得するために S に対して格構造分析を 行うことができる．格構造分析を行う場合は，S 中の subject マーカー(が/は), object マーカー(を/に)等，主要な格マーカーのみを用いる．いま，S 中で，格マーカー付きの名詞と動詞が現れるとすると，W はそれら2語とし，キーワード検出ステップ3を繰り返す．これは，格マーカーに直続される名詞と動詞は文中で，主要な役割を担うという仮定に基づく．4.3 節では入力文の具体例を用いて，処理の流れを示す．

4.3. 処理の流れの例

図3は会議開催に関する手紙文である．この日本文をもとに，システムが英文書作成の支援をする過程は次のとおりである．

> ***JS***: 第5回自然言語処理会議を2000年7月20-23日に東京で開きます．ついては，参考文献を含め，5000字以内の論文を募集いたします．提出論文は，新規性と有意性をもとに査読を受けます．3月15日までに論文のコピー3部を当会事務局宛てにお送りください．論文には，必ず氏名，所属先，手紙のあて先，電話番号，Eメールアドレスを明記してください．5月30日までに採否をお知らせいたします．

図3： 会議に関する手紙文

図4は前処理前の日本文 JS を入力した際の機械翻訳システムの出力である．翻訳システムの用いる辞書に対応する英語が未登録の日本語(査読)は，出力中に日本語のまま残ってしまう．

ES: The 5th natural-language-processing meetings are held in Tokyo on 20-July 23, 2000. Therefore, I invite the papers of less than 5000 characters including bibliography. A presentation paper receives 査読 based on freshness and significant nature. Please send copy 3 part of a paper in this society secretariat by March 15. Please be sure to specify a name, an affiliation place, the address of a letter, a telephone number, and E mail address in a paper. I will announce you adoption or rejection by May 30

図4： 図3の手紙文に対する機械翻訳システムの出力

JS の第 1 文 S_1 と前処理後の S_1'と ES_1'は

S_1 ：第 5 回自然言語処理会議を 2000 年 7 月 20-23 日に東京で開きます.

S_1': 第 5 回自然言語処理の会議を 2000 年 7 月 20-23 日に東京で開きます.

ES_1': The meeting of the 5th natural language processing is held in Tokyo on 20-July 23, 2000.

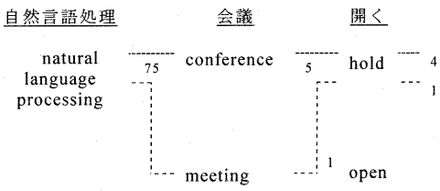
S_1'から *W*=(自然言語処理,会議,開く)を得る．日英辞書を引くことにより，*W* 中の各語に対して *M* を得る：自然言語処理 natural language processing; 会議 conference| congress| consultation| convention| council| meeting| table; and 開く hold| open| set up| throw| yield to. 表 4 はこれらの語の *DB* 内での出現回数である．

表 4 の頻度に基づいて，*W* に含まれる各語の語義を選択すると，自然言語処理は *natural language processing*,会議は *conference* と *meeting*,開くには *hold* と *open* となる．

表 4: 単語，語義，頻度

Word	Sense	Frq.
自然言語処理	natural language processing	75
会議	conference	5534
	meeting	3968
	table	2310
	council	1791
	congress	1245
	convention	638
	consultation	295
開く	hold	3319
	open	2677
	set up	1327
	yield to	284
	throw	254

次にこれらの語義の組み合わせすべてを含む文が *DB* 中に存在するかを検索する．結果，*natural language processing* を含む例 75 文，*natural language processing* と *meeting* を含む例 1 文，*natural language processing* と *conference* を含む例 5 文，*natural language processing* と *meeting* そして *hold* を含む例 1 文，*natural language processing* と *conference* そして *hold* を含む例 4 文を検索した．



このように

K={*conference, hold, natural language processing*}

W'={*hold, natural language processing, meeting*}

K'=*K*∪*W'*={*conference, hold, meeting, natural language processing* }

となる．*K'*を用いて，*MR* 値と *SR* 値を計算して，値の高い例文から，以下のように表示をする．

1. The other meeting is the Conference on Empirical Methods in Natural Language Processing, that will be held at the University of Pennsylvania, May 17-18, 1996, in conjunction with the University's 50th Anniversary celebration of the Eniac Computer. (MR=4/4, SR=3/12)
2. CALL FOR PAPERS The first International Conference on Recent Advances in Natural Languages Processing (RANLP) was held in 1995. (MR=3/4, SR=3/8)
3. PACIFIC ASSOCIATION FOR COMPUTATIONAL LINGUISTICS PACLING '93, the 1st Pacific Association for Computational Linguistics Conference, formerly JAJSNLP, the Japan- Australia Joint Symposia on Natural Language Processing, will be held on April 21-24. (MR=3/4, SR=3/10)

これらの例文のうち，ユーザは第 1 文と第 3 文を参照し，ES_1' を以下のように修正した：

The 5th Conference on Natural Language Processing will be held in Tokyo, on July 20-23, 2000.

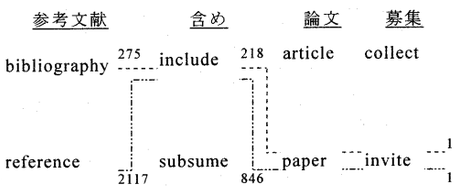
JS の第 2 文 S_2 と前処理後の S_2' と ES_2' は:

S_2=S_2': ついては，参考文献を含め，5000 字以内の論文を募集いたします.

ES_2': Therefore, I invite the papers of less than 5000 characters including bibliography.

第 1 文と同様の処理を行い，検索式 (bibliography*| reference*) & (include*| subsume*) & (within*) & (paper*| article*) & (collect*| invite*) を得た．しかしながら，この検索式を満たす例文は検出されなかった．

ここで，S_2 に対して格構造分析の結果を適用し，新しい検索式 (bibliography*| reference*) & (include*| subsume*) & (paper*| article*) & (collect*| invite*) を得る．各語の 2 つの語義で，16 の解釈がある．



下のように，*K*, *W*', *K*' を得る．

K = (*include*, *invite*, *paper*, *reference*)

W'= (*bibliography*, *character*, *include*, *invite*, *paper*)

K' = (*K*∪*W*')=(*bibliography*, *character*, *include*, *invite*, *paper*, *reference*)

それぞれの例文に対する *MR* と *SR* 値は下に示すとおりである．

1. Paper format: Original unpublished papers of up to 12 pages (single-spaced, 12 points, including figures, tables, and references) are invited. (MR=4/6, SR=4/10)
2. Authors are invited to submit an extended abstract or full draft paper of at most 10 pages, including glossary, bibliography, and illustrated examples, to arrive by January 17, 1999. (MR=4/6, SR=4/17)

2 つの例文を用いて，ユーザは ES_2 を次のように修正した．

Therefore, authors are invited to submit papers (not exceeding 5000 words including bibliography).

第 3 文以降にも同様の処理を行い，当システムを用いた場合の最終的出力は図 5 中の *E1* のようになった．

E1: The 5th Conference on Natural Language Processing will be held in Tokyo, on July 20-23, 2000. Therefore, authors are invited to submit papers (not exceeding 5000 words including bibliography). The submitted paper will be judged on originality and significance. Please send 3 copies of the paper to conference organizer by March 15. Please be sure the paper include title, author(s) full name, affiliations, complete address(es), telephone number(s), and electronic mail address(es). Authors will be notified by May 30 whether full review indicates acceptance, rejection, or a need for revision.

図 5： TransAid を使用した場合の出力

図 5 中の英語は完全ではなく，修正が必要な点もあるが，*JS* の訳として *ES* よりかなり改善されたといえる．

5. TransAid の評価

システムの評価法には，多くの基準がある[2, 22]．機能，コスト，使いやすさ，応答速度などである．

言語翻訳は難しく，時間のかかるタスクである．2 章で示した実験は機械翻訳システムを補う翻訳メモリ TransAid のようなシステムが必要であることを示している．

実験中，TransAid を用いて応対文を書いたすべての学生は TransAid が使いやすく，興味深いシステムであると応えている．しかし，表示される例文から，真に有効な文を決定する部分は，改善の余地があるとの意見もあった．

より具体的に TransAid を評価するため，5 人の大学院生に TransAid を使って書かせた英文書を 3 人のネイティブスピーカーに評価してもらった．評価の基準は，機械翻訳システムの出力と比較して，文構造の質の改善度と，意味の捉えやすさの改善度を，100～-100%で表してもらった．表 5 はその結果を示している．表中の数字は，平均のパーセンテージを示している．また，A, B, C は 3 人のネイティブスピーカーを表す．

すべてのネイティブスピーカーが，TransAid を用いた方が英文が改善されるという評価をしている．しかし，ここで，改善度はテキストのタイプによって，ばらつきがあることが分かる．paper abstract が他のタイプの文書より改善度が低いのは，我々が用いた会議データベースのテーマに合わなかったからだと考えられる．

表 5: 翻訳システムの出力に対する
TransAid の改善率

Type of Correspondences	Improvement(%)					
	Structure			Meaning		
	A	B	C	A	B	C
Call for paper	80	50	70	80	70	70
Acceptance letter	80	95	70	70	95	70
Deadline postponement	70	40	90	70	40	90
Inquiry for admission	80	60	80	80	60	80
Paper abstract	60	50	50	60	60	50

6. 結論

日本人の英語ユーザにとって，効果的な機械翻訳システムの実現は依然として困難である．そこで，日英翻訳作成支援システム TransAid を構築した．翻訳時の英作を支援し，労力，スタイル，内容の点で，人間または機械によって書かれた文章の質を向上させる．本システムを用いることにより，現存の機械翻訳システムに残された問題を改善することができる．図 4 に評価結果が示されているように TransAid を用いることにより英文は内容と読みやすさの点での改善がみられる．

システムの性能はキーワードの選択と例文の抽出法に大きく依存する．例文の抽出源としてインターネットコーパスを利用することは，テーマによって，質的，量的に大変効果的である．

しかしながら， TransAid の欠点は受動的なシステムにある．この点の改善が必要である。たとえば、

① 複合語自動分割処理に加えて，辞書未登録語の処理を加える．

② 翻訳メモリシステム[15, 21]に用いられている翻訳履歴を蓄積できる機能を加える．

などが考えられる．さらに多くの機能を加えれば、TransAid はもっと利用価値のある能動的なシステムとなるはずである。

文　献

[1] AltaVista, http://www.altavista.com, 2001

[2] Bailey, J. and Pearson, S., Development of a Tool for Measuring and Analyzing Computer User Satisfaction, Management Science, pp.530-545, 1983

[3] BROTHER, TransLand/EJ-JE V4.0, http://www.brother.co.jp/transland/tldemo.html, 2001

[4] FUJITSU, ATLAS V8.0, http://www.fujitsu.co.jp/jp/soft/atlas/sample.html, 2002

[5] Furugori, T., Improving Spelling Checkers for Japanese Users of English, IEEE PC, Vol.33, No.3, pp.138-142, 1990

[6] Furugori, T. and Takeda, A., An Example-Based Systems of English Sentences for Japanese English Users, Literary and Linguistic Computing, Vol.8, No.2, pp.85-90, 1993

[7] Golding, R., A Bayesian Hybrid Method for Context-Sensitive Spelling Correction, Proc. of 3rd Workshop on Very Large Corpora, pp.39-53, 1995

[8] Golding, R. and Schabes, Y.,: Combining Trigram-based and Feature-based Methods for Context-Sensitive Spelling Correction, Proc. of 34th Annual Meeting of the Association for Computational Linguistics, pp.71-78, 1996

[9] Han, D. and Furugori, T., Structural Analysis of Compound Words in Japanese Using Semantic Dependency Relations. Journal of Quantitative Linguistics (to appear), 2002

[10] Mihalcea, R. and Moldovan, Dan I., A Method for Word Sense Disambiguation of Unrestricted Text, Proc. of 37th Annual Meeting of the Association for Computational Linguistics, pp.152-158, 1999

[11] Payette, J. and Hirst, G., An Intelligent Computer Assistant for Stylistic Instruction, Computer and the Humanities, Vol.26, pp.87-102, 1992

[12] Resnik, P., Mining the Web for Bilingual Text, Proc. of 37th Annual meeting of the Association for Computational Linguistics, pp.527-534, 1999

[13] TOSHIBA, The Honyaku Professional V7.0, 2002

[14] TRADOS, http://www.trados.com, 2002

[15] Webb, L. E., Advantages and Disadvantages of Translation Memory: A COST/BENEFIT ANALYSIS, Master's Thesis, Master of Arts in Translation of German, Monterey Institute of International Studies, California, http://www.webbsnet.com/translation/thesis.html, 1992

[16] 成田真澄,"日本人のための英文生成支援環境に関する研究", "COE形成基礎研究費研究成果報告(1),神田外語大学, pp.569-594, 1997.

[17] 成田真澄,"英文作成支援ツールWriter's Helperの試作", "COE形成基礎研究費研究成果報告(2),神田外語大学, pp.807-819, 1998.

[18] 成田真澄,"英文アブストラクト作成支援ツールの構築", "COE形成基礎研究費研究成果報告(3),神田外語大学, pp.547-566, 1999.

[19] 成田真澄,"英文アブストラクト作成支援ツールの言語資源構築", "COE形成基礎研究費研究成果報告(4),神田外語大学, pp.603-610, 2000.

[20] 成田真澄,"英文アブストラクト作成支援ツールのユーザ評価", "COE形成基礎研究費研究成果報告(5),神田外語大学, 2001.

[21] 西垣通,インターネットで日本語はどうなるか,岩波書店, 2001

[22] 宮崎正弘,"係り受け解析を用いた複合語の自動分割法",情報処理学会論文誌,Vol.25, No.6, pp.970-979,1984

[23] 宮平知博, 渡辺日出雄, 田添英一, 神山淑朗, 武田浩一, インターネット機械翻訳の世界, 毎日コミュニケーションズ, 2000