

Google を活用した英作文支援システムの構築

大鹿 広憲[†] 佐藤 学[†] 安藤 進^{††} 山名 早人^{†††}

[†] 早稲田大学理工学研究科 ^{†††} 早稲田大学理工学術院

^{†,†††} 〒 169-8555 東京都新宿区大久保 3-4-1

^{††} 翻訳者

E-mail: [†]{sika,manavy}@yama.info.waseda.ac.jp, ^{††}sando@inter.net, ^{†††}yamana@waseda.jp

あらまし 近年、英語の必要性はますます高くなってきており、それに伴い Web 上の翻訳システムも多く開発されてきている。しかし、機械翻訳では直訳や単語の適合性および文脈の関係などから適切な英作文が行えないという問題がある。これに対し、より適切な英作文を行う方法として、検索エンジンを使った方法がある。フレーズ検索をはじめとした様々なタイプの検索を行いながら、Web 上の様々な情報を参照することで、作成した英作文の文型に対して汎用性を調べることができる。そこで、本稿では以上の技術を自動化し、誰でもスムーズに英作文の作業が行えるための英作文支援システムを提案する。検索エンジンには Google を使用する。

キーワード 翻訳, Web サービス, 検索エンジン

An English Composition Support System using Google

Hironori OSHIKA[†], Manabu SATOU[†], Susumu ANDO^{††}, and Hayato YAMANA^{†††}

[†] Graduate School of Science and Engineering, Waseda University

^{†††} Science and Engineering, Waseda University

^{†,†††} 3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555 Japan

^{††} Translator

E-mail: [†]{sika,manavy}@yama.info.waseda.ac.jp, ^{††}sando@inter.net, ^{†††}yamana@waseda.jp

Abstract In recent years, the necessity for English becomes high and many translation systems on the Web have been developed. However, machine translation has some problems that appropriate English composition can be performed from neither literal translation, the conformity of words, nor the relation of the context. There is a way using a search engine as a way of performing more suitable English composition. We can examine the flexibility of the created English composition by referring to the various sentence pattern on the Web, performing searches of various types including phrase search. This paper proposes an English composition support system of English for anyone to keep English composition smoothly. In this paper, the above technique is automated by using Google on a search engine.

Key words translation, Web Service, Search Engine

1. はじめに

近年、企業の海外事業展開が活発になってきたことや英語教育の推進などから、英語に触れる機会が増えてきている。それに伴い、多くの Web 上の和英辞典や翻訳ソフトが多く開発されてきた。しかし、機械翻訳では直訳で訳してしまうという問題点があったり、ノンネイティブな人でない限り、和英辞典だけで完全な英作文をすることは難しい。

以上の問題点を解決する方法として、検索エンジンを使った

検討がある。文献 [1] において、Google [2] を用いた英作文支援システムを構築し、ワイルドカードを使った検討や多義語の検討の自動化を行った。本稿では、文献 [1] で構築した英作文支援システムにおいて、辞書データベースの改良や活用形の対応、ドメインや用例の参照などについての機能を追加し、有用性の評価を行う。

以下、2 節では関連研究、3 節では検索エンジンを使った英作文の検討について述べる。4 節ではその検索テクニックの自動化について述べ、5 節で評価を行い、6 節でまとめを行う。

2. 関連研究と研究目的

翻訳に関する関連研究は、機械翻訳と英作文支援に大別できる。本節では、翻訳に関する関連研究について述べる。

2.1 機械翻訳の現状

機械翻訳は、定められた法則に基づいてデータベースを使って翻訳を行う。しかし、日本語は複雑で、特定の単語や文について幾通りもの翻訳・解釈の仕方があるために、完全な翻訳が難しい。近年では、幾通りもの意味解析に対応した対訳コーパスの作成の研究が多くなされているが、特定の品詞だけにおいてのものが多く、網羅性に欠けるのが欠点である。[3]

2.2 英文書作成支援

電気通信大学の高倉らは、機械翻訳の質が低いことを指摘し、文書作成支援システム TransAid[4] を提案している。インターネット・コーパスに検索エンジンを用いているところは本研究と似ているが、学会に関するページのみを集め、動詞と名詞に限定した訂正を行っているため、前置詞を使った熟語やその他の品詞に対して、汎用性がないのが欠点である。

2.3 本研究の位置づけ

以上で紹介した手法は、翻訳の法則を自動化するには難しいものであったり、実際に Web で提供されているものは少ないのが現状である。そこで、本研究では、Web 上での翻訳のサービスを実現することを目的に、検索エンジンを用いて、訳語候補を自動的に提示できる英作文支援システムを提案する。Web ページは、人手で作成されたものが多い。従って、検索エンジンを用例データベースにすることによって、多量の Web ページを用例として参照できるほか、汎用性の高い文型を用例と共に検索結果件数で比較ができるという利点がある。

3. 検索エンジンを使った英作文の検討

英作した文章に対して、フレーズ検索を行うことによって汎用性を調べることができ、また、ワイルドカードを使うことによって、よりの確な英訳が見つけることができる。

本節では、文献[5]を参考にし、検索エンジン Google を使った英作文の検討の方法について述べる。

3.1 フレーズ検索による検討

フレーズ検索を行うことによって、入力した文型がよく使われている表現かどうかを調査することができる。

例として、以下のような文章の英訳の検討を行うとする。

その選手は汗でびしょ濡れだった。

まず、この文章を Excite[6] で英作文してみると、以下のようになる。

The player was dripping wet in the sweat.

以上の文章において、「汗で濡れる」という英訳に着目する。「sweat」に係る冠詞の有無についてフレーズ検索を行うことによって検討することができる。

表 1 より、「sweat」には冠詞をつけない方が圧倒的に検索結果数が多いことが分かる。従って、「汗で濡れる」という英文において、この場合に使う「sweat」には冠詞を付加しない方が適当であることが分かる。

表 1 冠詞の検討

検索文字列	検索結果件数
"wet in sweat"	655
"wet in the sweat"	7

3.2 ワイルドカードを使った検討

3.1 で例を挙げた文章において、さらに検証してみる。「汗で濡れる」という英訳に対して前置詞の使い方が気になったとする。そこで、前置詞の部分を

"wet * sweat"

と with の部分をワイルドカードに置き換えて検索すると、検索結果として表示された用例の中から、

with, from

を使っているものがあつた。Excite[6] で訳したときの前置詞「in」を含め、以上の 3 つを前置詞の候補とする。

次に、それぞれの前置詞を含めたフレーズ検索を行うと、結果は表 2 のようになった。

表 2 前置詞の検討

検索文字列	検索結果件数
"wet with sweat"	3280
"wet from sweat"	854
"wet in sweat"	274

表 2 から、「with」を使った場合が一番ヒット件数が多いことがわかり、この文型がよく使われているということがわかる。

3.3 多義語の検討

英作文の作業において、一つの日本語に対する英訳は複数存在するので、どのように使い分ければいいのか分からないことがある。例えば、以下のような日本語文において英訳を検討する。

その候補者は選挙の結果に失望した。

同じように、Excite[6] で試してみると、以下のようになる。

The candidate was disappointed at the result of the election.

「選挙の結果」という部分の英訳が気になったとする。「結果」という単語は複数あるが、この場合における「結果」はどの単語を使ったらいいか迷うことがある。

和英辞典で「結果」を調べると、「result」の他に「outcome」、「conclusion」がある。従って、「選挙の結果」の訳は 3 通りの候補が挙げられる。これらの英語をフレーズ指定して Google のヒット件数を調べると、表 3 のようになる。

表 3 「選挙の結果」の訳語

検索文字列	検索結果件数
"outcome of the election"	24,300
"result of the election"	11,300
"conclusion of the election"	1,100

表 3 から、「outcome of the election」として使用する方が検索結果数が多いことがわかる。従って、「選挙の結果」の訳は「outcome of the election」として使うのが適切であると考えられる。

4. 検索エンジンを使った英作文支援システムの構築

3. で Google を使った英作文の検討について述べたが、人手で入力する場合、以下のような問題点が存在する。

(1) フレーズ検索を行う際に「”」(ダブルコーテーション)を付加するのが面倒である。

(2) ワイルドカードを用いる際、品詞の特定ができないと、関係のないものが検索結果に紛れてしまう。

(3) それぞれの英語をフレーズ検索して検索結果ウィンドウを切り替えながら、検索ヒット件数を比較するのは効率が悪い。

以上の問題点を解決するために、文献[1]において、ワイルドカードを使った検討の自動化と多義語の検討の自動化を行った。本システムの構成は以下のとおりである。

- GoogleAPI [5] によって Google にアクセス
 - 辞書データベースとして、EDR [8] を使用
 - 品詞の特定には、Eric Brill の MontyLingua [9] を使用
- 以下、本節では本システムの機能と処理の流れについて述べる。

4.1 ワイルドカードを使った検討の自動化

3.2 では、Google を使った前置詞の検討について述べた。この一連の作業を自動化する方法について述べる。

再度、3.2 の例を挙げながら説明する。まず、ユーザに気になる部分の英語文 (wet in sweat) を入力してもらう。そして、その中で検討したい部分を選択する。3.2 の例の場合、前置詞を検討したいので「in」をドラッグで囲むことになる。

処理開始ボタンが押されたときのシステムの概要を図1に示す。

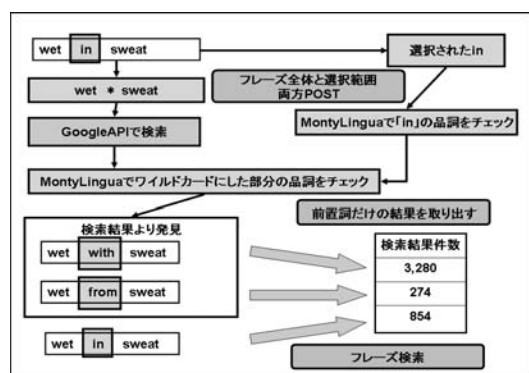


図1 ワイルドカードを使った検討の自動化

システムは、入力された英文と選択した部分の両方を POST し、選択した部分「in」をワイルドカードに置き換えて、Google でフレーズ検索を行う。このとき、「in」は MontyLingua から前置詞と判断する。そして、検索結果からワイルドカードに対応する部分に現われる単語を抽出する。

次にこれらの単語を品詞分解データベース MontyLingua により、「in」と同じ前置詞のものだけの用例を取り出す。前置詞のものを取り出した結果、「with」、「from」が使われているの

システムは、クエリを「wet in sweat」と「wet with sweat」「wet from sweat」として GoogleAPI を用いて並列処理で検索する。検索結果件数を比較する表にまとめ、ユーザにどちらの前置詞がよく使われているかの提示を行う。

4.2 多義語の検討の自動化

多義語の検討において、ユーザに入力してもらう情報は以下のとおりである。

- 検討したい英語の語句 (3.3 の例の場合、「選挙の結果」に対応する英訳)

- 調べたい英語の部分の日本語訳 (3.3 の場合、「結果」)

「結果」という日本語を入力してもらうのは、和英辞典で類語を探すためである。英単語をもとに類語を探すと、もとなる英単語が複数の異なる意味を持つ場合、的確な英単語表示をリストアップできない。そこで、最適な単語を探す場合には、日本語訳を入力してもらうようにした。また、英語で「選挙の結果」の部分まで入力してもらうのは、文型のパターンを読みとるため、この場合における「結果」の使い方について調べるためである。

本稿では、辞書データベースとして EDR の日英対訳辞書を用いた。収録されているレコード数は、364,430 個で、対訳情報は約 19 万語である。EDR のレコードの構成は、単語見出しとその対訳情報となっており、さらに日本語の見出しにおいての概要説明の情報も格納されている。ユーザは、概要説明の検索も行うことができ、訳語の完全一致だけでなく、概要説明からの訳語候補の選択を行うことができる。

「結果」に対する訳語を検索したときのシステムの処理の流れを7頁の図4に示す。同じ「結果」でも、「物事の最終的な結果」や「ある行為によって生じた結果」など複数の意味を持つ「結果」が存在するので、ユーザは状況に合わせて訳語候補を選択することができる。システムは選択された訳語候補をそれぞれ GoogleAPI で検索を行い、検索結果件数を表示する。

4.3 活用形や冠詞の対応

本項では、活用形や冠詞の対応におけるシステムの処理の流れについて説明する。

4.3.1 動詞の活用形の対応

本システムでは、入力された文章を MontyLingua で品詞を特定し、動詞が存在した場合は、以下のように活用形を変化させて OR 検索でフレーズ検索を行うようにした。活用形を変化させてフレーズ検索を行うことによって、より幅広い用例の検索を行うことができる。

EDR の英語単語辞書を使用し、動詞の活用形を調べる。調べた活用形の種類を「(現在形 OR 進行形)」のような形で指定することによって、検索結果件数に活用形を変化させたものを含むことができる。

例えば、「漫画を作る」の英訳として「make the cartoon」という表現は使われているかどうかのフレーズ検索を行うとする。するとシステムは、「make」の品詞を解析し、EDR の英語単語辞書で構成した品詞解析データベースから原形「make」、3人称単数「makes」、進行形「making」及び過去分詞形「made」を抽出し、それぞれ OR 検索で囲った形で検索を行う。(図2)

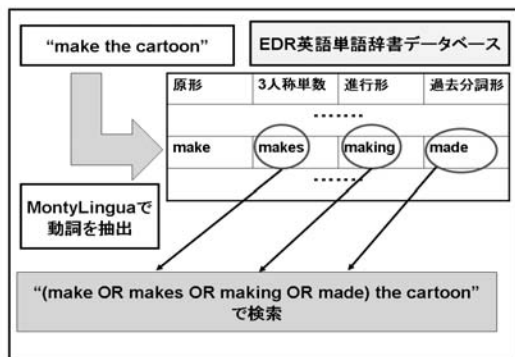


図 2 動詞の活用形の対応

4.3.2 名詞の複数形の対応

名詞は単数形だけでなく複数形も使われている。従って、単数形でフレーズ検索を行うと単数形の用例しか検索できない。以上の問題を解決するために、フレーズ検索の際、(複数形 OR 単数形) と指定することによって、単数形と複数形を検索対象に含めることができる。

本稿ではオプション機能として、検討したい語句の中に名詞が含まれていた際、自動的に(複数形 OR 単数形)に置き換える機能である「名詞の活用形の対応」を実装した。

4.3.3 冠詞の検討の自動化

3.1で冠詞の有無の検討について述べた。本システムでは、MontyLinguaの品詞解析で名詞と判定したものに対し、オプション機能で「the」をつけた場合と「the」をつけない場合のフレーズ検索を自動的に行うようにした。

4.4 用例の参照

Googleの検索結果における snippet には、検索キーワードとその周辺の文章の情報が含まれているが、個々の検索結果では、用例は見にくい。そこで、本システムでは正規表現を使って訳語周辺の1文を抜き出し、それぞれの訳語候補に対して用例を見やすいようにし、訳語候補の検討を行えるようにした。

4.5 ドメインの参照

検索結果のURLのドメインを調べることによって、その訳語候補が英語圏で使われているところかを調べることができる。例えば、「激しい論争」という日本語文を英語にするときの訳語の候補として、「vehement controversy」があったとする。この「vehement controversy」をフレーズ検索を行い、検索結果のURLを見てみると、

home.wanadoo.nl/piet.fontaine/

www.nvvs.nl/

www.qantara.de/webcom/show_article.php/

となり、ドメイン名が「de」と「nl」のものが多かった。このことから、「vehement controversy」という語句は、非英語圏で良く使われているものが多いことが分かる。また、ドメイン名には「.edu」といった分野を特定するものも含まれている。従って、本システムでは検索結果のURLを自動的に収集し、ドメインを提示する。ユーザは、その情報によってその表現がどのような分野あるいは地域で使われているかが分かる。

4.6 品詞分解による構文解析の検討

機械翻訳の構文解析に関する問題点で、関係代名詞を使って翻訳する傾向が多いことや、「挿入」「省略」「倒置」に関するルールには弱いことが挙げられていた。

従って、本節では、MontyLinguaによって品詞分解を行った結果を解析し、関係代名詞を使った構文と副詞の挿入の位置に関する検討を自動的に行う機能をオプションとして実装した。

以下に、システムの流れについて説明する。

4.6.1 関係代名詞を使った構文の検討

機械翻訳において、特に精度が低いのが関係代名詞を使った文の翻訳である。「～な人」や「～なもの」などの名詞節を機械翻訳で英訳すると、that や which を使った関係代名詞を使った文にしてしまうことが多い。しかし、関係代名詞を使った英訳に対しフレーズ検索を行うと、件数が少なく汎用性が低いことが分かる。従って、以上のような名詞節を英訳する場合、関係代名詞を省略し名詞を修飾する動詞の「～ing」や「～ed」を使って、名詞に係る英訳の方法がある。

例えば、「Web上に存在するデータ」を英訳するとき、Exciteの翻訳では、「data that exists on the Web」となる。「data that exists on the Web」の語句をフレーズ検索すると、0件になる。従って、一般では使われない表現だということが分かる。そこで、「倒置」などを行った様々な文型に置き換えてフレーズ検索を行ってみた。表4に結果を示す。

表 4 文型の汎用性

検索語句	検索結果件数
"data that exists on the Web"	0
"data existing on the Web"	6
"existing data on the web"	88
"data on the web"	37,900

表4からthatを使った修飾よりも、「ing」を使った名詞の修飾、あるいは「ing」を前に置いた名詞の修飾の表現の方が使われていることが分かる。更に「exist」は省略可能な動詞なので、「data on the Web」としても可能な表現であることが分かる。

以上の検討を本システムで自動的に行えるように、MontyLinguaで品詞解析を行って「SVO」の関係抽出し、それぞれのパターンについてフレーズ検索を行う機能を実装した。

構文解析機能の概要を図3に示す。

例えば、「data that exists on the Web」に対しての「SVO」の関係抽出すると、以下のようになる。

(Verb-Subj-Obj-Obj) = ("exist" "that" "on Web")

本来なら主語に当たる部分は「data」が抽出されるべきだが、MontyLinguaは関係代名詞の「that」を抽出してしまう。Sに「that」や「which」が抽出された場合、一つ前の単語に着目することに。「that」が「あの、その」の指示語として使われている場合、検討したい語句として入力するときには「that」が一番最初に来るので、指示語と関係代名詞の判断は前に名詞が存在しているかどうかで判断する。

次いで、EDRを使って、Vの活用形を調べる。Vに当たる

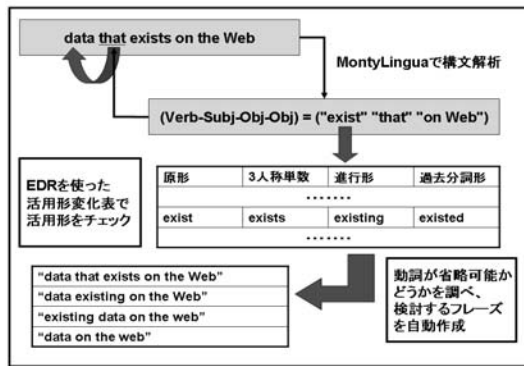


図 3 構文解析機能の概要

部分が「exist」や「get」の場合、省略可能動詞と判断し、省略した形も検討する。MontyLingua では、C に当たる部分は O として抽出する。例えば、「data which get interested」のような「get + C」は「～になる」という文法なので、「data interested」とすることも可能である。本システムでは、「exist」「get」を省略可能な動詞としてデータベース化を行った。

4.6.2 副詞による修飾の位置の検討

英訳する際、副詞の挿入位置に疑問を感じることもある。例えば、「影響を強く受ける」という文を英訳する際、翻訳ソフトで英訳すると「The influence was received strongly」となる。副詞が受身形の後ろで修飾されているが、副詞を受身形の前に持ってきて、修飾することもできる。従って、どちらの置き方がよく使われているかを検討することができる。実際に Google でフレーズ検索をすると表 5 のようになる。

表 5 副詞の位置の検討

検索文字列	検索結果件数
(be OR is OR was OR were OR been OR being) strongly received	145
(be OR is OR was OR were OR been OR being) received strongly	74

以上の検討を本システムで自動的に行う機能を実装とした。処理の流れは、以下のとおりになる。まず、MontyLingua の品詞解析で、「be 動詞 + 動詞の受身形 + 副詞」と解析する。品詞解析は以下のとおりに行われる。

was/VBD/be received/VBN/receive strongly/RB/strongly

原形の形も解析されるので、以上から「be 動詞 + 動詞の受身形 + 副詞」と判断することができる。従って、副詞の位置を置き換え、それぞれにおいてフレーズ検索を自動的に行うようにした。

4.7 ワイルドカードの応用的使用

ワイルドカードは、「*」一つで任意の単語 1 語に相当するので、1 語のものしか検索できない。従って、複数指定することによって熟語が検索できたり、また、ワイルドカードの置き方によっては新たな文型の単語の使い方が発見できることがある。

文献[1]において、ワイルドカードの複数指定による検討について述べた。文献[1]のように前置詞を含んだ動詞句や熟語を検索するには最適の機能である。しかし、「多義語の検討」と違い、日本語訳を指定して検索することはできないので、単語と単語の間に入りそうなものを探す、いわゆる共起を調べる程度に過ぎない。

本システムでは、オプション機能として、ワイルドカードを使った検討を行う際に、複数調べることを自動的に行うかどうかの選択をユーザが選ばせる機能を実装した。検討したい語句について、ワイルドカードの部分を「A * B」と置き換えるのに加え、「A * * B」や「A * * * B」と自動的にワイルドカードの個数を増やしてフレーズ検索を行う。また、A が S、B が V の場合にはよく使われる副詞や助動詞も発見することができる。

更に、MontyLingua で品詞特定を行うことによって、検討したい品詞だけを抽出機能もオプションとして選択できるようにした。以上の機能を実装することにより、ワイルドカードの複数指定による検討を効率良く行うことができる。

4.8 英作文支援システムの機能のまとめ

本稿では、第 4 節で紹介した機能の一覧を 7 頁の表 10 に示す。英作文の検討の際に生じた問題点と問題点を解決するための本システムの機能に対応させると表 10 のようになる。

5. システムの評価

本節では構築したシステムの有用性を調査するための評価を行う。本稿では、英訳が頻繁に行われる以下の分野において英作文の評価を行った。

あらかじめ日本語文と英語文の対訳情報のある正解データを用意し、日本語文に対し翻訳ソフトで英訳を行い、本システムを使って修正を施す。修正を施した場所において、正解データである英語文と使用している単語が一致しているならば「正解」とする。翻訳ソフトは、「The 翻訳プロフェッショナル V10」(製品版)を用いた。

本稿では、英訳が頻繁に行われる以下の分野において英作文の評価を行った。

- 一般の英作文 (日常生活で用いられる)
- 新聞記事の英文化 (時事問題として用いられる)
- 学術論文の英文化 (教育分野で用いられる)

一般の英作文の正解データには任意の英文集[10]、新聞記事の英文化には「読売新聞」に掲載されている「社説」[11]を、そして学術論文の英文化には、研究室で書かれた論文を使用し、人手で書かれた英語論文をネガティブスピーカーによって添削されたものを正解データとする。翻訳ソフトを使う際の日本語に関しては、一字一字正しく入力するものとする。

本節では、以上の分野でランダムに抽出した日本語訳に対する英作文の作業において、システムの評価と考察を行う。以下、評価項目、評価基準について述べ、評価結果と共に考察を行う。

5.1 評価を行う検討項目

評価を行う検討項目を 6 頁の表 6 にまとめた。精度に関しては、表 6 の 7 項目に対して算出するものとする。

5.2 評価基準

本項では、評価基準について実例を用いながら説明する。

5.2.1 検索結果件数が一番多いのフレーズが他のものより極端に多い場合

次の文において英訳を検討する。

賃料が手頃ならどんなマンションでもいいです。

表 6 評価を行う検討項目

評価の対象となる検討項目	検討内容
ワイルドカードの検討	前置詞の検討に関しては、4.1 のような方法でワイルドカードを使って評価を行う。その他の品詞に関しては、「ワイルドカード複数指定の検討」の方法で、評価を行う。
動詞における多義語の検討	動詞において 4.2 の和英辞書を用いた多義語の検討の方法で、評価を行う。
名詞における多義語の検討	名詞において 4.2 の和英辞書を用いた多義語の検討の方法で、評価を行う。
形容詞における多義語の検討	形容詞において 4.2 の和英辞書を用いた多義語の検討の方法で、評価を行う。
冠詞の検討	気になった冠詞の有無について 4.3.3 のような方法で検討し、評価を行う。
構文解析の検討	気になった構文について、4.6 の構文解析の検討の方法で、評価を行う。
ワイルドカード複数指定の検討	気になった動詞の使い方や <i>SVO</i> の関係について、4.7 のワイルドカードを複数置いた検討の方法で、評価を行う。検討を行う品詞は限定しない。

以上の文を「The 翻訳プロフェッショナル V10」で訳すと以下のようなになる。

If a rent is handy, what kind of apartment house is sufficient.

「手頃な」の英訳に着目する。「賃料」に対する「手頃」なので、「手頃」において本システムの 4.2 の機能を使って検討すると、表 7 のような結果が得られる。

表 7 「賃料」に対する「手頃」の多義語の検討

検索文字列	検索結果件数
"rent is handy"	3
"rent is suitable"	40
"rent is reasonable"	925

表 7 から「reasonable」を使ったものが圧倒的に検索結果件数が多い。以上の場合、「rent is handy」はおかきな表現と判断する。また表 4.3.3 の機能から「a」より「the」の方が適切だと判断する。従って、以下のように書き換える。

If the rent is reasonable, what kind of apartment house is sufficient.

正解データの文献[10]に掲載されている英文は以下のとおりである。

Any apartment will do as long as the rent is reasonable.

文は倒置になっているものの、「家賃が手頃な」という部分においては、修正した通りになっている。従って、「形容詞」の「多義語の検討」および「冠詞の検討」において成功したことになる。

5.2.2 検索結果件数が一番多いフレーズが他のものと大差がない場合

次の文において英訳を検討する。

個人は地域社会の基本的な構成要素である。

以上の文を「The 翻訳プロフェッショナル V10」で訳すと以下のようなになる。

An individual is a fundamental component of a community.

「構成要素」の英訳に当たる「fundamental component」について検討する。「要素」において、本システムの 4.2 の機能を使って検討すると、表 8 のような結果が得られる。

表 8 「構成要素」の多義語の検討

検索文字列	検索結果件数
"fundamental element"	25,400
"fundamental factor"	6,780
"fundamental component"	23,900
"fundamental constituent"	1,100

表 7 から、「fundamental component」と「fundamental element」は同じくらいの件数であることがわかる。以上から「fundamental element」と書き換えても良いということがわかり、以下のような 2 パターンの書き方があることを示す。

An individual is a fundamental [element/component] of a community.

正解データの文献[10]に掲載されている英文は以下のとおりである。

An individual is a fundamental element of a community.

全く同じ文になり、本システムを使って検討することができたことになる。以上の場合、「名詞」において「多義語の検討」が成功したことになる。

書き換えを判断する検索結果件数は、一番検索結果件数が多いものとするが、2 番目に多いものとの差が 2 倍以下の場合、どちらの表現でもおかしくないと判断する。

5.2.3 評価基準のまとめ

評価基準を表 9 にまとめた。

表 9 評価基準

評価結果	評価内容
正解	検索結果件数の一番多いフレーズと正解データで使用しているフレーズが一致した場合
正解	検索結果件数の一番多いフレーズと 2 番目に多いフレーズの検索結果件数の差が 2 倍以下で、正解データが上記の 2 つのフレーズのうちのどちらかを使用していた場合
不正解	多義語の検討において、どのパターンのフレーズ検索も検索結果件数が 0 件であり、修正することができなかった場合
不正解	検索結果件数の一番多いフレーズ (2 番目との差が 2 倍以下の場合) は検索結果件数が多かったフレーズと 2 番目に多かったフレーズの 2 パターン) に直した結果、正解データで使われている単語が違った場合
評価の対象にしない	機械翻訳の単語が検索結果件数 1 位だった場合 (修正する必要がなかった)

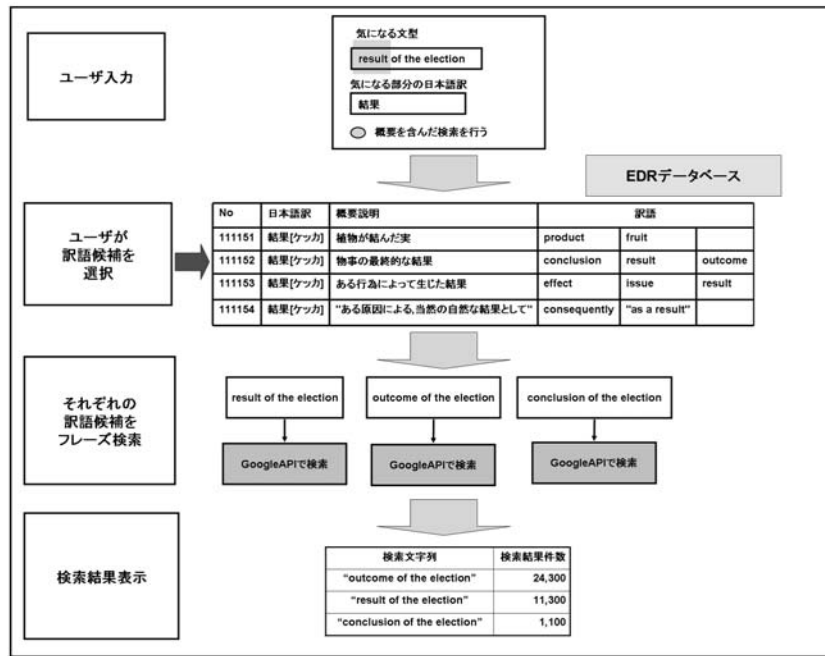


図 4 多義語の検討の自動化

表 10 検索エンジンを使った英作文支援システムの機能一覧

従来の提案手法における問題点及び英作文の作業における問題点	問題点を解決するための本システムの機能	機能の概要
従来の英日に関する翻訳の提案手法では、検討する品詞が限定されている。	ワイルドカードを使った検討	気になる部分をワイルドカードで指定することによってどのような単語が使われているかを調べる。全ての品詞に対応できる。
用いられる文によって単語の英訳が変わってしまうので、使い分けが分からない。	多義語の検討	気になる部分を EDR を用いて、同じ意味を別の単語に置き換えて、フレーズ検索を行うことにより汎用性を調べる。
文の状況に応じて、冠詞をつけたいのかわからない。また、フレーズ検索では 1 種類の活用形しか検討できない。	活用形の対応 (名詞の冠詞の検討及び複数形の対応)	それぞれのパターンにおいて、フレーズ検索を行う。また、OR 検索を行うことによって、活用形の対応を自動的に行う。
英作した文型が英語圏で使われているかどうかを調べたい。	ドメインの参照	検索結果 URL のドメインを抽出して統計を表示することにより、英語圏で使われている表現かどうかを提示する。
従来の英日に関する翻訳の提案手法では、コーパスの分野が限定されてしまっている。	用例の参照	検索結果の snippet を抽出することにより、検討したい語句がどのように使われているかを見やすく提示することにより、用例を参照できる。
従来の英日に関する翻訳の提案手法では、関係代名詞を使った翻訳や副詞の位置の置き方などが良くない。	品詞分解による構文解析の検討	それぞれの文型のパターンでフレーズ検索を行い、検索結果件数を表にまとめて提示
検討したい単語がどのように使われているかを調べたい。2 つの語句の関係を調べたい。	ワイルドカード複数指定による熟語発見	検討したい語句についての新しい使い方を発見できる。

表 11 評価結果のまとめ

	任意の単語集			読売新聞社説			海外論文			合計		
	検討数	正解数	精度	検討数	正解数	精度	検討数	正解数	精度	検討数	正解数	精度
前置詞の検討	9	7	77.78%	5	5	100.00%	4	3	75.00%	18	15	83.33%
多義語の検討 (動詞)	12	8	66.67%	12	8	66.67%	5	4	80.00%	29	20	68.97%
多義語の検討 (名詞)	25	20	80.00%	35	24	68.57%	9	5	55.56%	69	49	71.01%
冠詞の検討	5	4	80.00%	5	5	100.00%	3	3	100.00%	13	12	92.31%
多義語の検討 (形容詞)	31	20	64.52%	9	5	55.56%	8	6	75.00%	48	31	64.58%
構文解析の検討	5	3	60.00%	8	5	62.50%	7	6	85.71%	20	14	70.00%
ワイルドカード複数指定	6	5	83.33%	3	2	66.67%	4	3	75.00%	13	10	76.92%

5.3 評価結果

以上の評価データと評価基準を元に、3分野の評価対象データから任意の単語集の対訳集から91文、読売新聞の社説の日英対訳集から57文、そして海外論文の対訳集から33文の計181文に対して、本システムの評価を行った。評価結果は7頁の表11のようになった。

5.4 考察

本節では、評価に対する考察を行う。検索エンジンを使った検討で修正が上手くいった場合と、修正が上手くいかなかった場合においてそれぞれ考察を行う。

5.4.1 修正が上手くいった場合

「～な問題」や「～なもの」などの名詞節、冠詞の検討およびワイルドカードを使った前置詞に関する検討は精度が高かった。構文が一律に定まっている場合や、単純な「SVO」、「SVC」構文における検討は汎用性が調べやすかったと考えられる。従って、任意の単語集からの英訳の検討は全体的に精度が高かった。

5.4.2 修正が上手くいかなかった場合

修正がうまく行かなかった原因について以下のような原因が挙げられる。

- 意識による日本語の変化
- 構文解析による検討の限界
- ユーザの英語に対する知識

以下、それぞれについて説明を行う。

a) 意識による日本語の変化

新聞記事や海外論文では、意識が多かったため、十分な精度が出なかった品詞があったり、前置詞の検討を行っても正解では前置詞が使われていないことがあった。

翻訳ソフトは、前後に共起する単語を把握しながら、基本的に日本語を忠実に翻訳する。従って、検索エンジンを使った検討も構文解析の場合を除き、翻訳ソフトで英訳したものに基いて検討を行うので、意識が行われると精度が下がってしまう現象が起きた。

和英辞書を使った検討では、見出し語概念を用いることにより、意識の対応を行った。しかし、本来は「怪しい」と訳される「strange」という単語を、文脈上から「未知」という意味で翻訳しているなど、意識の対応が十分でなかった点があった。従って、見出し語概念の機能を拡張する機能が必要だと考えられる。

また、意識として省略を行った構文があった。「get」や「exist」など、どの構文においても省略しやすいものであれば、検討が行えるが、文章として成り立っている日本語を省略する「予測が難しい省略」の傾向があったため、的はずれな単語に修正してしまったケースもあった。

b) 構文解析による検討の限界

新聞記事や海外論文は、日本語文が長いために複雑な構文になっていることが多かった。単語の検討はできたが、構文解析による検討は十分に行えなかった。従って、構文解析の検討やワイルドカード複数指定の検討数は少なかった。例えば、「不十分だ」という日本語文を英訳する際、翻訳ソフトで翻訳を行う

と、「inadequate」と英訳しているのに対し、正解では「hardly at an advantage」と訳してあった。肯定文を否定文で表現する場合において、文の構造が変わってしまうので、検索エンジンで検討するには難しいことが分かった。

c) ユーザの英語に対する知識

検索エンジンを使った英作文の検討は、1文そのものではなく、気になった部分に対して検討を行う方法であるので、ユーザによって「気になる文型」は異なってくる。従って、評価を行うユーザが異なると、精度も変わってくると考えられる。今回の実験では、ユーザ1人に対して行ったので、今後は多くの人に対して実験を行うべきだと考えられる。

6. おわりに

本稿では、日本語から英語に直す英作文の作業を支援するために、検索エンジンを使った英訳の検討の方法を提案し、検討のための作業を自動化する英作文支援システムの構築を行うことによって、有用性があることを明らかにした。

しかし、日本語のパターンは無数に存在し、数多くの英訳の方法があるので、検索エンジンを使った検討で以上のパターンを網羅するためには、本論文で紹介した以外にも様々な検討方法を発見する必要がある。そして、今回評価を行った英文以外にも多くの英訳に対し検討し、評価を行う必要がある。

また、本稿で構築したシステムは、ある程度の英語知識を持った人にとっては使いやすいシステムとなったが、英訳を行う際、どのように英訳したらいいかわからない人にとっては使いにくいシステムであることが分かった。

今後は、英作文支援の範囲を広げ、英語の知識を問わず誰でも使いやすい英作文支援システムを構築する必要があると考えられる。

本システムは、<http://ir.yama.info.waseda.ac.jp/trans/>にて公開中である。

参考文献

- [1] 大鹿広憲, 佐藤学, 安藤進, 山名 早人: "検索エンジンを使った翻訳サポートシステムの構築", 電子情報通信学会技術研究報告 Vol.104, No.107, pp.237-242 (2004.7)
- [2] Google
<http://www.google.com>
- [3] 鈴木健二, 梅村恭司: "情報検索システムを利用した日英対訳推定", 情処研報, NL-151-1, pp.1-6(2002)
- [4] 高倉佐和, 古郡廷治: "TransAid-英文書作成支援システム-", 情処研報, NL-150-2, pp.7-14(2002)
- [5] 安藤進著: "翻訳に役立つ Google 活用テクニック", 丸善, ISBN4-621-07294-3 (2003.10)
- [6] EXCITE 翻訳
<http://www.excite.co.jp/world/>
- [7] 山名早人監訳, 田中裕子訳, Tara Calishain & Rael Dornfest 著: "Google Hacks", オライリー・ジャパン, ISBN4-87311-136-6 (2003.8)
- [8] 通信総合研究所, EDR 電子化辞書仕様説明書 (2003)
- [9] MontyLingua Project
<http://web.media.mit.edu/hugo/montylingua/>
- [10] 鈴木陽一著: "現代英語の重要単語・熟語 2400", ICP, ISBN4-900790-00-1, 1995
- [11] DAILY YOMIURI
<http://info.yomiuri.co.jp/company/shimen/03.htm>