

# 検索エンジンを利用した9ヶ国語作文支援の拡張

## Expansion of a Composition Support Tool for 9 Languages using a Search Engine

古川 陽平<sup>†</sup>  
Yohei Furukawa

綱島 祐一<sup>†</sup>  
Yuichi Tsunashima

安藤 一秋<sup>‡</sup>  
Kazuaki Ando

### 1. はじめに

近年、英語や中国語など第二言語の必要性が高まり、第二言語で読み書きする機会が増加している。通常、第二言語で作文する際は辞書を参照する。しかし、訳語や熟語に対する用例の数が少ないため、書きたい文の参考にできない場合が多い。また、複数の訳語候補から適切な語が選択しづらい、作成文が正しく書けているのか、自然な文なのかかわからないなどの問題もある。

これらの問題に対し、Web上に存在する膨大な多言語テキストを用例文と仮定し、検索エンジンによって得られるヒット数や用例文を作文に利用する方法が有効である[1-7]。多くの用例文を参照することで、書きたい文に対する表現や言い回しを参考にできる。また、ヒット数はWeb上での利用頻度とみなせるため、実際に利用される自然な訳語や表現の選択・判断基準となる。

本研究ではこれらの特徴を生かし、第二言語で作文するユーザに対し、検索エンジンによって得られるWeb上の多言語テキストを用いて作文時に有用となる情報を提示し、ユーザの試行錯誤を支援するツールの実現を目的とする。我々の先行研究[7]では、作文対象言語の基本文法を一通り学習した中級レベルのユーザを対象に、9カ国の言語(日本語・英語・ドイツ語・中国語・韓国語・フランス語・イタリア語・スペイン語・ポルトガル語)に対応した作文支援ツールを提案した。しかし、言語習得が初級レベルのユーザ(初級ユーザ)に対する支援も有用である。一般的に初級ユーザは語彙や文法知識などが未熟である。そこで、Web上のフリー翻訳ツールの利用が考えられるが、翻訳結果をそのまま利用するだけでは学習効果が期待できず、翻訳精度にも問題がある。

そこで本稿では、構文トレーニングを考慮した訳文作成支援機能を提案する。訳文作成支援は、訳語提示と編集支援の2つの機能で構成される。訳語提示機能では、母国語で作成した文を入力とし、その文に含まれる内容語の訳語をユーザに提示する。編集支援機能では、ユーザ自身が訳語をマウスで並び替えたり、修正・検証したりしながら、目的の訳文を作成することを支援する。これらの機能により、ユーザは、訳語の並び替え問題を解くスタイルで作文できるため、容易に作文できただけでなく、学習効果も期待できる。

### 2. ツールの主要機能

本ツールでは、文中の誤り候補を検出する機能、語・表現の妥当性検証を支援する機能、用例を提示する機能を備えている。具体的には以下の機能がある。

- n-gram 検索
- ワイルドカード検索

- ドメイン指定検索
  - 用例提示(日・英は語彙レベルでランキング可能)
- なお、Webアクセスには、Yahoo! ウェブ検索 Web API を利用している。

#### 2.1 n-gram 検索

n-gram 検索は、文中の誤り候補の検出を支援する。文の長さが長くなるとヒット数が得られにくいため、入力文を n-gram 系列に分割し、その系列のヒット数により、誤り候補の場所を指摘する。分割方法は作文の対象言語によって異なる。英語やドイツ語などの場合、空白までを1単語として分割、日本語の場合は Yahoo! 日本語形態素解析 Web API を利用して文節単位に分割、韓国語の場合は文節相当で分割、中国語の場合は1漢字単位に分割し、n-gram 系列を生成する。各 n-gram 系列を Yahoo! API で検索し、ヒット数を得た後、ヒット数に応じて色分けしたグラフでユーザに提示する。

例として、英語で “No matter how it is hot” と入力して、3-gram~4-gram で検証した結果を図1に示す。図1より、“how it is hot” のヒット数が少なく、赤色のグラフで表示されている。これは、“how it is hot” が、一般的な表現ではない可能性を示している。ユーザは、訳語の追加・削除、セルをドラッグして語順を変更するなど試行錯誤しながら作文できる。また、各系列のヒット数をクリックすることで、用例文も参照できる。

検索結果	hitCount
No matter how it is hot	16100000
No matter how	1800000
No matter how it is	5980000
No matter how it is hot	360000
No matter how	164000
No matter how it is hot	444

図1 英語の n-gram 検索結果

#### 2.2 ワイルドカード検索

ワイルドカード検索は、語・表現の妥当性検証を支援する。例えば、複数の訳語候補から妥当な語を選択したい場合、検索フレーズ中の検証したい箇所をワイルドカード化し、その候補として検証したい訳語を指定して検索する。ツールでは、ワイルドカードを指定された候補で置換したフレーズを生成し、フレーズ検索する。その結果は、ヒット数を基にランキング表示される。

例として、日本語で “機械が\*作動した” のように検証したい箇所をワイルドカード化し、その候補を “偽@誤@被@乱” のように@で区切って指定し、検索した結果

<sup>†</sup> 香川大学大学院工学研究科, <sup>‡</sup> 香川大学工学部

を図2に示す。検索結果はヒット数でランキング表示されるので、ユーザは最適な候補の選択が容易に行える。また、各フレーズをクリックすることで、用例を参照できる。

Rank	*の箇所該当する単語	抽出されたフレーズ	Hit count
1	誤	機械が誤作動した	157
2	被	機械が被作動した	0
3	乱	機械が乱作動した	0
4	偽	機械が偽作動した	0

図2 候補指定ありのランキング表示

ワイルドカードに対する候補が不明な場合は、候補を指定せずに検索することで、ツールがワイルドカード位置に存在する候補を用例から自動で収集する。その後、収集した候補を基に検索フレーズを生成・検索し、ヒット数でランキング表示する。なお、ワイルドカードに対する候補は、検索結果に含まれる任意数 (default で 50) の用例から収集する。

例えば、“機械が\*作動した”に対して候補を指定せずに検索すると、ツールが候補を自動収集し、図3に示すように、ヒット数でランキング表示する。

Rank	*の箇所該当する単語	抽出されたフレーズ	Hit count
1		機械が作動した	450
2	誤	機械が誤作動した	157
3	正しく	機械が正しく作動した	40
4	再	機械が再作動した	2
5	実際	機械が実際作動した	1

図3 候補指定なしのランキング表示

### 3. 日英訳文作成支援機能

一般的に初級ユーザは語彙や文法知識などが未熟である。そこで、構文トレーニングを考慮した訳文作成支援機能を提案する。訳文作成支援は、訳語提示と編集支援の2つの機能で構成される。訳語提示機能では、母国語で作成した文を入力とし、その文に含まれる内容語の訳語をユーザに提示する。編集支援機能では、ユーザ自身が訳語をマウスで並び替えたり、修正・検証したりしながら、目的の訳文を作成することを支援する。これらの機能により、ユーザは、訳語の並び替え問題を解くスタイルで作文できるため、容易に作文できるだけでなく、学習効果も期待できる。

以下に、訳語提示の手順を示す。

- ①日本語入力文を文節に分割し内容語を抽出する。和英辞書から内容語に対する訳語を検索し、日本語文と同じ並びで訳語を配置する。なお、訳語が存在しない場合、Web翻訳を用いて訳語を取得する。
- ②ある内容語に複数の訳語が存在する場合、訳語一覧をサジェスト表示する。
- ③日本語文に主語が存在しない場合、先頭に主語を自動挿入する。
- ④訳語列をフレーズ検索し、ヒット数を表示する。
- ⑤訳語列を n-gram 検索し、誤り位置の特定を支援する。
- ⑥訳語列に対する用例を検索し、提示する。

以下に、編集支援機能を示す。

- ①ドラッグ&ドロップによる訳語の並び替え
- ②任意単語(訳語)・文節の変更・追加・削除
- ③冠詞・前置詞・be動詞・助動詞・代名詞の挿入
- ④サジェスト表示された単語候補の選択と置換
- ⑤再検索によるヒット数・n-gram検索結果・用例の更新
- ⑥ワイルドカード検索による不明箇所の単語取得と候補一覧のサジェスト表示
- ⑦ログの保存と削除

例えば、“ときどき映画に行きます。”と入力し検索した結果を図4に示す。ツールは、入力文を文節ごとに区切り、各内容語に対する訳語を検索後、入力文の語順“I sometimes movie go to”で表示する。また、入力文には主語が存在しないため、自動で“I”を挿入する。ユーザは、用例や n-gram 検索結果を参考にしながら、“go to”などの訳語の並び替えを行う感覚で作文できる。

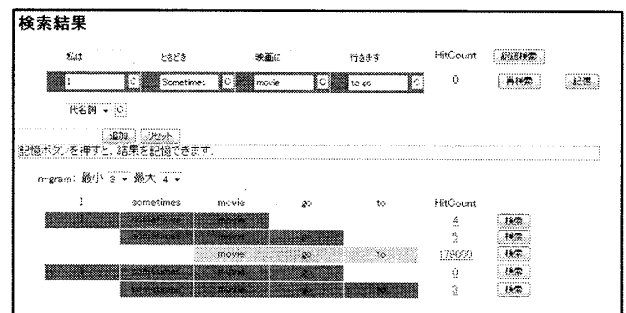


図4 “ときどき映画に行きます。”で検索した結果

### 4. おわりに

本稿では、多言語作文支援ツールの新しい機能として、初級ユーザに対する日英訳文作成支援について提案した。提案機能により、ユーザは、訳語の並び替え問題を解くスタイルで作文できるため、容易に作文できるだけでなく、学習効果も期待できる。今後は、日英訳文作成支援機能を英日にも対応させ、英日特有の機能の考案を行う予定である。またツールの有用性を評価し、使いやすいツールを実現する。

### 参考文献

- [1] 佐藤他, “検索エンジンを利用した英作文支援システムの構築”, NLP 第12回年次大会論文集, pp.664-667, 2006.
- [2] 有富他, “検索エンジンを用いた英文前置詞誤り修正支援”, DEIM 2010, F9-5, 2010.
- [3] 大鹿他, “Google を活用した英作文支援システムの構築”, DEWS2005 4-B-08, 2005.
- [4] 王他, “ウェブ検索を用いて中国語作文支援システムの構築”, FIT2008 論文集, pp.175-176, 2008.
- [5] 平野他, “検索エンジンを用いた英文冠詞誤りの検出”, 情報研報(DBS), Vol.2007, No.65, pp.139-144, 2007.
- [6] 山本他, “検索エンジンに基づく多言語用例指南ツール:KIWI”, NLP 第9回年次大会論文集, pp.654-657, 2003.
- [7] 古川他, “Web 検索を利用した 9ヶ国語作文支援の改良”, 信学技法 ET2009-125, pp.125-130, 2010.