共起関係ならびに構文情報を考慮した英文汎化と英作文支援

An Approach to Simplify Retrieved English Sentences in Consideration of Their Co-occurrences and Structures to Help English Compositions

天野 翼[*1], 渡部 孝幸[*2], 田中 省作[*3], 宮崎 佳典[*4]
Tsubasa Amano[*1], Takayuki Watabe[*2], Shosaku Tanaka[*3], Yoshinori Miyazaki[*4]
[*1] 静岡大学情報学部
[*1]Faculty of Informatics, Shizuoka University
[*2] 静岡大学自然科学系教育部
[*2]Graduate School of Science and Technology, Shizuoka University
[*3] 立命館大学文学部
[*3]College of Letters, Ritsumeikan University
[*4] 静岡大学学術院情報学領域
[*4]College of Informatics, Shizuoka University
Email: wing5221@gmail.com

あらまし：英作文をする際，適切な語や構文を選択するために英文を参考にすることがしばしばある．しかし，参考にする英文が学習者にとって過度に複雑である場合，学習者がどこに注目していいのか判断することは一般に容易ではなく，結果として折角の英文が活用されない可能性がある．そこで，英文中の語のうち，学習者が英作文を行う上で参考にならないと思われる個所を品詞に置き換えることで簡略化し，構文情報を用いて語の削減をするといった汎化手法を提案する．
キーワード：英作文支援，汎化，itf-isf，共起，Web アプリケーション

1. はじめに

英作文をする際，学習者が適切な語や構文を選択するために英文を参考にするという方法がしばしば見受けられる．しかし，参考にする英文が学習者にとって過度に複雑である場合，学習者が英文のどこに注目して良いのか判断するのは一般に容易ではない．そこで我々は，何らかの尺度で収集された英文集合（例：キーワードで検索されたコーパス内の英文集合）に対し，その集合より得られる特徴量を利用することで，英文を簡略化して提示する手法（汎化）を提案してきた[(1)(2)]．英文を汎化することで，学習者の英作文作業に貢献する可能性が高くはないと考えられる語が品詞に置き換えられ，学習者は参考になる語を把握しやすくなると考えられる．

本研究では(1)で開発された英文汎化アプリケーションの有用性・有効性向上に向けた改善を目的とする．その英作文支援の効果を測定するための実験を行い，実験結果に考察・分析を与えるものとする．

2. 汎化の概要

英文の汎化は，コーパスから英文集合を取得した後に 2 段階のプロセスによって行われる．

2.1 頻度と共起関係を考慮した語の品詞化

品詞化する語は学習者にとって参考となる程度が低い語と推測されるべきである．そこで英文集合Sに含まれる全ての語wについて，次の値を計算する：

$$\text{itf–isf}(w;S) = \frac{1}{f(w)} \log\frac{|S|}{sf(w)}$$ ，ここに

$f(w)$はSに含まれるwの数，$sf(w)$はSのうちwを含む文の数，$|S|$は英文の総数を表す．この itf-isf は低頻度・低文数ほど大きな値を取る指標である．itf-isf の値が大きくなる語は，学習者が取得した英文集合Sにおいて共通して出現する語ではなく，文もしくはそれが含まれた文書のトピックに強く依存したものであると考えられる．すなわち，itf-isf の値が高い語は，英作文の参考にならないと考えられるため品詞ラベルに置き換えられる（品詞化）．

itf-isf が高い語の中には，“cache memory”のような隣接する共起語のいずれかが含まれてしまう可能性がある．その際，“cache <NN>”（<NN>は名詞を現す品詞ラベル）のように他方のみが品詞化されるが，“cache memory”は 2 語で 1 つの特定の意味を成している（単体では各々“隠し場”，“記憶”の意）ため，片方を残したところでユーザに資する情報となるとは考えにくい．そこで，品詞化されない語と強い共起性が認められた語については，たとえ itf-isf が高くとも品詞化せずに，双方が条件を満たしたときに併せて品詞化する．

2.2 語数の削減

2.1 節によって英文が以下のような文になったとする（<PR>，<JJ>，<DT>は各々，代名詞，形容詞，限定詞を表す品詞ラベル）：

<PR> propose a <JJ> solution for the <NN> of <DT> <NN> <NN>.

この文の末尾（下線部）は品詞と機能語の羅列であり，英作文をする上で参考になるような情報を有していないと考えられ，冗長である（ここに“solution for”は共起語とする）．そこで，構文情報を用いて語数の削減を行う．以下の図 1 が上記の文の構文木である（<VBP>，<P>は他動詞，前置詞を表す品詞ラベ

ル，<NP>，<VP>，<PP>は名詞句，動詞句，前置詞句の句ラベルを表す)．

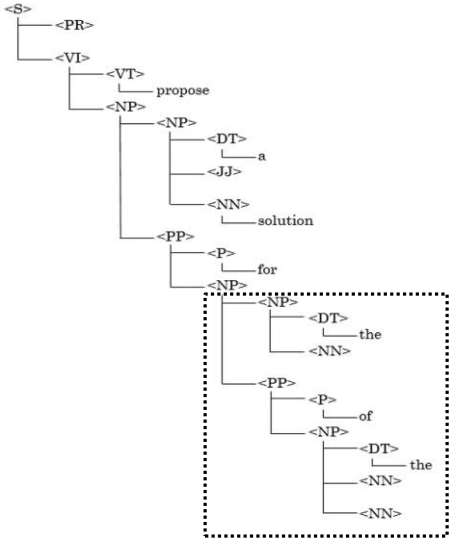

図 1 構文情報

節あるいは句の子すべてが「子を持たない節あるいは句」,「品詞ラベル」,「共起語に含まれない機能語」のみの場合は，それらの子の削除を再帰的に行うことで冗長な箇所を削除する．結果的に，図 1 の点線部が削除され，以下の簡略形が得られる：

<PR> propose a <JJ> solution for <NP>.

3. 現行汎化アプリケーションの改良

現行のアプリケーション(1)が実使用に耐えるには, 幾つかの改善点が必要と考え, 再実装を行った.

まず，(1)では学習者が汎化をする際に単一の英文に着目していると考え，汎化をその英文のみに対して行っている．しかし，英文集合の共通性を学習者が認識するには，汎化する対象を英文集合全体にすべきと考えた（「複数文の同時汎化」)．

さらに，共起語について検討した．共起語には，隣接したもの以外にも，“define A as B”のような離れた共起語が存在する（さらに A や B が 1 つの単語とは限らない)．そこで本研究では，離れた共起語についても隣接した共起語同様に，同一文中であればその距離にかかわらず品詞化の抑制を行うこととした（「離れた共起語の品詞化抑制」)．

加えて，(1)におけるコーパスには“where-ever”のような，本来一語であるべき語がハイフン区切りになっていたり，英文として成立していないものが多く含まれていた．これらをスクリーニングし，適切に訂正，削除の処理を行った．

4. 実験および結果

本研究で改良したアプリケーションでは，「複数文の同時汎化」により英文集合の共通性がより明瞭になることが期待される．そこで，某国立大学の学生 11 名に対して実験を行い, 汎化の有無によって英文集合内に存在する共起語に意識が働くのかどうか調査を行った．なお，今回は実験協力者間で汎化する度合いや試行回数を統一するため，敢えて実装済み Web アプリケーションではなく, それより生成される英文集合を印刷した紙媒体にて実験を行った.

問題は特定の動詞を用いた和文英訳問題 4 問で，課題 1 は和文英訳をする際に役に立ちそうな英文を英文集合から選んでくる作業，課題 2 は選定した英文を参考にし，和文英訳をしてもらった．英訳用和文として，特定の動詞と対になる共起語を含む英文の日本語訳を用いた．4 問中，2 問は隣接した共起語（具体的に“appear in”［問題 1］と“merge with”［問題 4］)，もう 2 問は離れた共起語（“define A as B”［問題 2］,“combine A as B”［問題 3］）を含む英文集合とした.

英文集合は特定の動詞を含む 30 文，そのうち 15 文はその共起語を含んだものとした. 問題の難易度，汎化なし／ありの順序による要因をなくすために被験者を 2 グループに分けた．グループ 1 は問題 1・2 の英文集合を汎化なし，問題 3・4 の英文集合を汎化あり，グループ 2 は問題 1・2 を汎化あり，問題 3・4 を汎化なしとした．汎化あり英文集合は，異なり語数 10 語まで汎化したものとした. 評価項目は，選定された英文中に共起語が含まれている英文の割合とした（適合率)．以下の表 1 が集計結果である.

表 1 実験結果

グループ	グループ1						グループ2					平均
被験者	A	B	C	D	E	F	G	H	I	J	K	
汎化なし	15	71	65	67	58	63	61	25	50	83	50	55
汎化あり	92	88	90	83	100	71	39	46	67	35	69	71

（単位：%，値は小数点第 1 位を四捨五入）

表 1 から共起語が含まれていた英文の割合が汎化ありで増加していることが分かる．これより，相対的に実験協力者は汎化によって共起語を認識できたのではないかと推測される．もう 1 つの課題である和文英訳の実験結果については紙面の都合上，発表当日に事例の紹介をはじめ，考察・分析を行う.

5. おわりに

本研究では，(1)に修正を加え，それによって生成される汎化された英文集合が英作文をする学習者にどのような影響を与えるかについて調査した．今後は，実際のアプリケーションを実験協力者に使ってもらう実験をデザインし，先行研究との比較を行うことを計画している.

参考文献

(1) 渡部孝幸,田中省作,宮崎佳典,構文構造と共起性を考慮した英文汎化手法，統計数理研究所共同研究リポート 338, pp. 59-66 (2015).

(2) 天野翼,渡部孝幸,田中省作,宮崎佳典,構文情報を考慮した検索英文集合に対する汎化手法,第 14 回情報科学技術フォーラム講演論文集, pp.(2)-211-214 (2015).