

B2IM2026

修士論文

英作文支援のための用例検索に関する研究

高松 優

2014年 2月 10日

東北大学 大学院
情報科学研究科 システム情報科学専攻

本論文は東北大学 大学院情報科学研究科 システム情報科学専攻に
修士(工学) 授与の要件として提出した修士論文である。

高松 優

審査委員：

乾 健太郎 教授 (主指導教員)

篠原 歩 教授

徳山 豪 教授

岡崎 直観 准教授 (副指導教員)

英作文支援のための用例検索に関する研究*

高松 優

内容梗概

日本人にとって、英語で文章を書くことは容易ではない。そのため、英作文を支援するための研究は多くなされているが、本研究では、学習者に用例を提示することにより英作文支援を行う、用例検索に焦点をあてる。用例検索とは、ある表現を含む英文をコーパスから検索し、検索文や検索文数を閲覧することによって、その表現の適切性や用法を確認するものである。用例検索に関する先行研究はいくつか存在するが、どのような検索機能がどう有用であるかは明らかになっておらず、詳細な評価実験もなされていない。そこで、本研究ではまず、用例検索の観点から英作文における問題点を分類・整理し、その結果に基づいて用例検索システムを設計・実装した。次に、本システムを利用して英作文の問題を解く形式の評価実験を設計・実施した。その結果、各機能がどのように利用されているかを詳細に分析することができた。今後の課題として、評価実験の結果を踏まえて各機能の改善や必要な機能の追加などを行うことにより、システムを洗練していくことがあげられる。

キーワード

自然言語処理、英作文支援、学習支援、用例検索、第二言語学習、検索エンジン

*東北大学 大学院情報科学研究科 システム情報科学専攻 修士論文, B2IM2026, 2014年2月10日.

目次

1	はじめに	1
2	関連研究	4
3	英作文における用例検索の有用性	14
3.1	表現の適切性を確認したい場合	14
3.2	表現の組み合わせが分からない場合	14
3.2.1	表現を言い換えたい場合	16
3.2.2	組み合わせる表現が思いつかない場合	16
3.3	表現の用法が分からない場合	16
4	システムの設計	20
4.1	検索機能	20
4.1.1	拡張ワイルドカード検索	20
4.1.2	品詞検索	20
4.1.3	類義語検索	21
4.2	検索結果の提示方法	21
4.2.1	統計表示	21
4.2.2	用例の提示方法	21
4.2.3	用例の仕分けの方法	21
5	システムの実装	23
5.1	検索機能の実装	26
5.1.1	拡張ワイルドカード検索	26
5.1.2	品詞検索	27
5.1.3	類義語検索	27
5.2	検索結果の提示方法の実装	28
5.2.1	KWIC 表示	28
5.2.2	用例の仕分けと統計表示	29

5.2.3	目次の表示	29
5.2.4	提示された用例のフィルタリング	29
5.3	システムの利用例	30
6	予備実験	33
6.1	第一回 予備実験	33
6.1.1	実験設定	33
6.1.2	評価方法	36
6.1.3	実験結果	37
6.1.4	考察	40
6.2	第二回 予備実験	41
6.2.1	実験設定	41
6.2.2	評価方法	42
6.2.3	実験結果	43
6.2.4	考察	46
7	評価実験	47
7.1	実験設定	47
7.2	評価方法	51
7.3	実験結果	52
7.4	考察	62
8	おわりに	64
	謝辞	65

目 次

1	用例検索の仕組み	2
2	StringNet Navigator : 用例の集約	5
3	StringNet Navigator : グループ “[verb] as a result of” に属する用例	5
4	ESCORT : 依存構造による用例の分類	6
5	EReK : KWIC 表示と用例のソート	7
6	SCOPE : 用例の集約	8
7	SCOPE : フレーズ “we will show <NP>” を含む用例	8
8	exemplar : “we show” を含む用例とジャーナルの情報	9
9	NativeChecker : 誤りの検出	11
10	NativeChesker : 誤り訂正後の用例提示	11
11	ESL Assistant : 誤り訂正後の用例提示	12
12	用例の仕分け	22
13	用例検索システム PoEC	23
14	用例検索の流れ	24
15	“make * model” の検索結果	30
16	“build * model” の検索結果	31
17	“[J] coverage” の検索結果	31
18	TOEIC の穴埋め問題	34
19	[第一回予備実験] 英文穴埋め問題	34
20	本システムのクエリログ	36
21	[通常採点] 正解率	37
22	[厳密採点] 正解率	38
23	[第二回予備実験] 1 文を英作文する問題	41
24	[第二回予備実験] 実験結果	44
25	ExcelVBA を利用した解答フォーム	48
26	評価実験の手順	50

表 目 次

1	“approach for” で検索した際の用例検索の例	15
2	“result [V] that” で検索した際の用例検索の例	17
3	“[J] idea” で検索した際の用例検索の例	18
4	“precision * %” で検索した際の用例検索の例	19
5	英文の indexing	24
6	類義語を指定する synonyms.txt の例	25
7	本システムで indexing の際に用いる synonyms.txt	26
8	被験者の分類	42
9	[A グループ/前半セット (本システム)] チェックポイントの分類	52
10	[A グループ/後半セット (Google)] チェックポイントの分類	53
11	[B グループ/前半セット (Google)] チェックポイントの分類	53
12	[B グループ/後半セット (本システム)] チェックポイントの分類	53
13	[A グループ/前半セット] 1 周目の解答の正解率	54
14	[A グループ/後半セット] 1 周目の解答の正解率	54
15	[B グループ/前半セット] 1 周目の解答の正解率	54
16	[B グループ/後半セット] 1 周目の解答の正解率	55
17	[Google] チェックポイントの改善理由	56
18	[本システム] チェックポイントの改善理由	57
19	[A グループ/前半セット] 本システムを利用した際の検索回数	59
20	[B グループ/後半セット] 本システムを利用した際の検索回数	59
21	[A グループ] 2 周目の平均解答速度と自信の上昇度	60
22	[B グループ] 2 周目の平均解答速度と自信の上昇度	60

1 はじめに

近年、英語の重要性がますます高まっているが、英語が母語ではない日本人にとって、英語で文章を書くことは容易ではない。英語で文章を書く際に、分からない単語や表現がある場合、辞書や表現集を活用することが考えられる。しかし、辞書や表現集は1つの表現に対する用例が少なく、その表現を実際にどのように英作文に取り込めば良いのかが分からない場合がある。また、表現の用法を確認したい場合、既存の検索エンジンを用いて用例を検索することが考えられる。この場合、既存の検索エンジンは用例を検索するために設計されていないため、多くの用例を集約したり、複雑な検索には不向きであり、必要な用例を容易に閲覧することは難しい。このように、辞書や表現集、既存の検索エンジンだけでは十分に英作文を支援することは難しいため、英作文を支援するための研究が多くなされている。英語学習者の英作文を支援する方法の1つに、学習者自身がパソコンと検索ソフトを利用して実際の用例を閲覧することで語法を発見し理解していく、帰納的な学習方法がある [1, 2]。こうした学習方法はデータ駆動型学習 (Data-Driven Learning) と呼ばれる。DDL では帰納的学習によって、学習者の「気づき」が導かれるため、発見した語法に関する情報が学習者の記憶に残りやすい [3]。このように、英語学習者にとって DDL が有用であることは既に明らかになりつつある [2, 3, 4, 5]。自然言語処理の分野では、機械翻訳、スペル訂正、文章校正、用例検索などの研究が英作文支援と関連が深い。

本研究では、基本的な英文法は理解している大学院生程度の学生を支援対象として想定し、論文執筆などのテクニカルライティングを支援するため、用例検索に焦点をあてる。用例検索とは、多数の英文を収録したコーパスからある表現を含む文を検索し、検索文や検索文数を閲覧することによって、その表現の適切性や用法を確認するものであり、DDL を行うための手法の一つである。このように、実際の用例を効率良く検索し閲覧することによって、辞書や表現集、既存の検索エンジンの問題に対応できると考えられる。例えば、図1では、“approach for”を検索クエリとして入力して、大量の英文が含まれるコーパスから “approach for” を含む文を検索し、検索結果として “approach for” を含む文と、その検索文数を提示している。このとき、検索文や検索文数を閲覧することによって、“approach

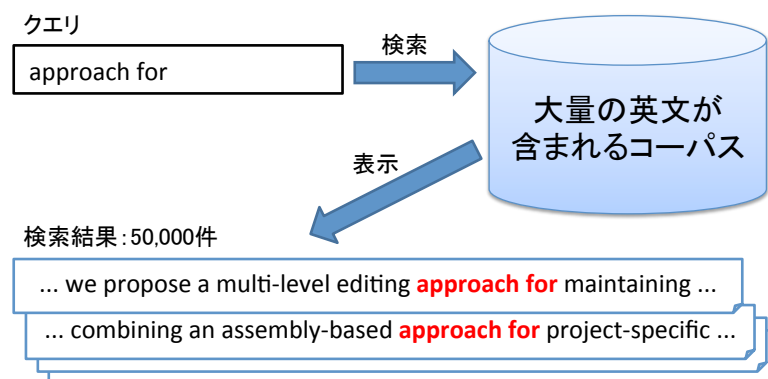


図 1: 用例検索の仕組み

for” という表現が実際に頻繁に用いられていることが確認できる。近年、用例検索を中心とした英作文指南本がいくつか出版されるなど、注目されている作文支援技術である [6, 7, 8]。

しかし、既存の用例検索システムは英作文における問題を十分に考慮して設計されていない。そのため、英作文における用例検索システムの有用性について、詳細な評価を十分に行える段階にない。

本研究では、英作文における用例検索システムの有用性を詳細に評価するため、まず、英作文に関する文献に基づいて、日本人が英語を執筆する際に直面する問題に対する、用例検索システムによる支援の可能性を調査した。次に、調査結果に基づいて、英作文支援に必要な機能を検討し、それらの機能を備えた用例検索システムを設計し、実装した。最後に、英作文における用例検索システムの有用性を評価できるような評価実験を設計し、実際に実験を行い、実験結果について評価を行った。テクニカルライティングを対象とするため、英文コーパスには ACL Anthology 上の主要会議の論文データ¹を用いた。

以降、第 2 章では、関連研究を取り上げ、英作文のための用例検索の現状について記述する。第 3 章では、第 2 章で取り上げた現状に対応するため、英作文における問題点を用例検索の観点から分類・整理し、用例検索によって英作文をどのように支援できるかについて考察する。第 4 章では、第 3 章で行った考察に基づいて、用例検索システムの設計を行う。第 5 章では、用例検索システムの実装

¹英文数は約 1,340,000 文である。

について記述する。第6章では、予備実験について、第7章では、評価実験について記述する。第8章では、本論文のまとめと、今後の課題について記述する。

2 関連研究

英作文を行う際、用例を閲覧することによってある表現の用法を確認したい場合がある。このとき、用例検索の方法として、代表的なものに、Googleなどの検索エンジンを活用することがあげられる。調べたい表現を検索クエリとした場合の検索文書数やスニペットを閲覧することで、英作文に役立てることができる[6]。しかし、検索エンジンは用例検索を目的として設計されていないため、多くの用例を集約したり、品詞の指定、類義語拡張といった複雑な検索には不向きであり、検索クエリを作るのが難しい、検索結果に用例とは関係のない部分が多い、といった問題がある。そのため、用例検索を容易に行うために、用例検索に特化したツールである用例検索システムが研究されてきた。

既存の用例検索システムには以下のような様々なものがある。

Wibleら[9]によるStringNet Navigator²は、図2のように、用例を品詞別・語形別に集約することで抽象化した形で表示する。[verb]や[noun]といった特定の品詞を含めて検索することが可能であり、例えば、動詞に対応する[verb]という表記法を用いて、“[verb] result”というクエリを入力して検索すると、図2の結果が表示される。ここで、図2のNo.1の “[verb] as a result of” というグループの Examples のアイコンをクリックすることにより、図3のような画面が表示され、“[verb] as a result of” という表現を含む用例を閲覧することができる。また、図2に示される “[verb] as a result of” の Parents をクリックすると、“[verb] as a result [prep]” や “[verb] as a [noun] of” といったより抽象化された表現が含まれる用例を閲覧することができ、Children をクリックすると、“occur as a result of” や “suffer as a result of” といったより具体化された表現が含まれる用例を閲覧することができる。

松原ら[10]によるESCORT³は、入力されたクエリ中の全ての単語が含まれており、かつその単語間に直接的な関係のある用例のみを提示する。その関係は依存構造によって捉えられる。図4のように、用例を依存構造ごとに分類して表示することにより、学習者が閲覧したい用例を容易に探すことができるように設計

²<http://nav.stringnet.org/>

³<http://escort.itc.nagoya-u.ac.jp/>

[Next page](#)
















No.	Chunks	Frequency ?	Examples	Parents ?	Children ?
1	[verb] as a result of	1123			
2	[noun] [verb] as a result of	271			
3	[verb] [adv prt] as a result of	92			
4	[verb] [adv] as a result of	181			
5	[verb] as a result of the	344			

図 2: StringNet Navigator : 用例の集約

Example sentences of "[verb] as a result of " are from [British National Corpus](#)

[Next page](#)

No.	Sentences
1	Over the next twelve months we will be providing support to a growing number of new and existing local programmes, identified as a result of an extensive survey we have just completed.
2	Along with the thousands of appeal letters sent as a result of Amnesty 's campaign, perhaps the clearest direct pressure is being exerted by the Sri Lanka Aid Consortium, made up of donor governments and multi-lateral institutions.
3	Statistically these deaths largely occur as a result of the inhalation of vomit and alcoholic poisoning, and it therefore seems somewhat bizarre for the police to be in charge of those suffering in such a way.
4	So perhaps my overkill misses the target: perhaps the behaviourist analysis of perception is sound, even though a general behaviourism is not, and what BS lacks is not knowledge of the nature of certain mental states, but only the ability to respond spontaneously to visual stimuli, that is, to respond as a result of actually seeing them.
5	The functionalist view I advocate is that mental phenomena emerge as a result of the way that the neurons etcetera are functionally organized in the brain, not as a result of the physical properties of neurons per se.

図 3: StringNet Navigator : グループ “[verb] as a result of” に属する用例

ESCORT

cost 0

検索キーワード: [-v result] 再検索

コスト: 0 + -

ヒット数: 10000

パターン1

```

  graph TD
    dobj --> v[-v]
    dobj --> result[result]
  
```

Fig.6 shows the result .
 Fig.7 shows the result .
 Fig.9 shows the result .

[6692 例すべて見る](#)

パターン2

```

  graph TD
    ncm --> v[-v]
    ncm --> result[result]
  
```

This is quite a promising result .
 We now have the following result .
 Figure 3 shows the achieved results .

[869 例すべて見る](#)

図 4: ESCORT : 依存構造による用例の分類

されている。特定の品詞を含む検索が可能であり、例えば動詞に対応する -v という表記法を用いて、“-v result”というクエリを入力して検索すると、図4のように、依存構造のパターンによって分類された用例を閲覧することができる。

大名 [11] は、和英辞書の機能を組み込むことにより、学習者が検索したい内容を表す単語が分からない場合に、適切な語を選んで検索できるように誘導している。和英辞書検索では、ひらがなで単語を入力することにより、対応する英単語の候補がいくつか表示され、その中から単語を選択してクリックすると選択した単語によって用例検索が行われる。例えば、「はし」を入力すると、候補として「橋:bridge」、「端:end(e.g. of street), edge, tip, margin, point」、「箸:chopsticks」という結果が表示され、この中から edge という単語を選択すると、edge という単語を含んだ用例が KWIC 表示により提示される。

英語例文検索 EReK⁴は、検索対象を Web 全体や.edu ドメイン、ニュースサイトに切り替えて用例を検索することができる。また、KWIC 表示されたフレーズの前後で、用例をアルファベット順にソートすることができる。例えば、“we show the result”というクエリを入力して.edu ドメインで検索し、フレーズの後

⁴<http://erek.ta2o.net/>

英語例文検索 EReK

we show the result 検索

All the Web Only .edu domain Only news sites

we show the result の使われ方 (用例)

total 3340 pages

並べ替える

we show the result using direct-mapped reuse buffer. In Figure 4 w...
we show the result of user editing where the goal was to trim everyt...
we show the result of this interpolation. Once k is known, FC can ...
we show the result of switching Elvis Presley and Richard Nixon...
we show the result of refinement, producing a smooth structure. Th...
we show the result of loading the file listed in (1). FIGURE 1. Init...
we show the result of four different scenarios, i.e. spatial scale...
we show the result of computing the area inside the marked square...
we show the result of automatically replacing the input faces (top) ...
we show the result of a connected-component based text-line extracti...
we show the result of 6 levels of iteration which produces 36= 729 ...
We showed the result in Fig.7. This result shows that the proposed a...
we show the result holds even if the full components are revealed ...
We show the result as a comparison to the time it takes to compute...
we show the result after 5 iterations using bisection refinement to ...

Next

並べ替える

図 5: EReK : KWIC 表示と用例のソート

ろ側でソートすると、図 5 のような結果が表示される。

松原ら [12] によるフレーズ検索システム SCOPE⁵は、検索対象が学術論文であり、「序論」や「関連研究」といった、論文中的特定の部分を選択して用例を検索することができる。例えば、“we show”というクエリを入力して「提案手法」の部分から検索すると、まず、図 6 のように、用例がフレーズごとに集約された状態で表示される。ここで、“we will show <NP>”というフレーズをクリックすると、図 7 のように、このフレーズを含む用例が提示される。

Springer による exemplar⁶は、様々な分野のジャーナルを検索対象としていて、用例とともに提示された出版年や分野などの情報によって検索を絞り込むことが可能である。例えば、“we show”というクエリを入力して検索すると、図 8 のように、“we show”という表現を含む用例だけでなく、検索されたジャーナルの出版年や分野の情報もあわせて提示される。この部分を選択してクリックすることにより、検索結果を絞り込むことができる。

竹内ら [13] は、語学教育における学習効率の向上を目指して、学習用例の提示方法を工夫した句動詞用例集を作成している。用例検索インタフェースでは、具

⁵<http://scope.itc.nagoya-u.ac.jp/>

⁶<http://www.springerexemplar.com/index.aspx>

'we show' の検索結果

ヒット数: 3

フレーズ	出現回数	出現論文数	長さ	スコア	章構成
we can show that <CL>	18	16	3	4.01	提案手法
we will show <NP>	16	16	3	3.84	提案手法
we will show how	16	14	3	3.84	提案手法

[Back](#)

図 6: SCOPE : 用例の集約

'we will show <NP>' の詳細情報

統計情報

出現頻度	文書頻度	長さ	スコア	章構成
16	16	3	3.84	提案手法

用例

- 1 **We will show that overgeneration** can be controlled by additional constraints.
- 2 In this section, **we will show the improvement**
- 3 Instead, **we will show examples 7** that our method could not handle properly.
- 4 **We will show later that** this seemingly trivial difference brings a significant improvement.

図 7: SCOPE : フレーズ “we will show <NP>” を含む用例

The screenshot shows the 'exemplar' search interface. At the top, it says 'words in context beta' and 'Powered by Springer'. The search bar contains 'we show' and has a 'Search' button. Below the search bar, there are filters for 'Subject' and 'Publication'. On the left, there is a 'year published' chart showing a peak around 2012 and a 'subject' bar chart with categories like Computer Science (36368), Life Sciences (18383), Mathematics (18242), Physics (16410), and Artificial Intelligence (incl. Robotics) (14630). The main area displays 'Showing 1 to 25 of 147089 matching articles' with a list of search results, each starting with '... to progression of breast and prostate cancer. We show that galectin-3 regulated functions like ...'.

図 8: exemplar : “we show” を含む用例とジャーナルの情報

体的な語句を入力して用例を検索することができるだけでなく、文型（動詞、名詞、受動態などの文法区分）やテーマ（不変化詞の機能をまとめた意味的な区分）から用例を検索することが可能である。複数の言語情報から用例へアクセスできるインタフェースが、ユーザーの気づきの促進や自発的な学習を進める原動力につながると考えている。

綱島ら [14] は、用例の語彙レベルと学習者の語彙レベルを比較してポイント付けし、ポイントの高い順に用例を並べ換えて提示することによって、用例を学習者の語彙レベルに合わせて表示する手法を採用している。また、英会話講師が大学程度の語彙レベルで並び替えた用例の順を正解とした場合に、この手法により正しく用例を並び替えられたかどうかについての評価を行なっている。更に、ツール全体についての評価も、主観的なアンケート調査により行なっている。小規模な評価にとどまっているが、実際にツールを利用した学習者は、ツールが英作文に有用であったと評価している。

三好ら [15] による用例検索ツール SOUP は、英作文時の語句選択を支援するために、用例に含まれる語の品詞や活用形ごとに用例を分類して提示する。用例を分類する際、クエリや検索条件の内容によって、分類方法を変更しており、こ

の分類を句構造分類と呼んでいる。また、16名の被験者を対象に英文アブストラクトの作文を行ってもらい、ツールの使用感に関するアンケート調査を行っている。主観評価にとどまっているが、ツールが既存の用例検索システムよりも有用であることが示されている。

坂本ら [16] は、入力文字列と用例文字列の類似度を N-gram に基いて計算し、類似度の高い用例を上位に出力する手法を採用している。類似度の計算は、多くの用例に含まれる出現頻度の高い要素の重要度を下げるために、idf を用いる。この検索手法は、医療分野における多言語用例対訳共有システム TackPad⁷ に導入されている。実際にこのシステムを利用したユーザーのログを利用してこの検索手法の再現率、適合率、F 値を求めて評価を行ったところ、通常のキーワードによる検索に比べ高い F 値を得ることができている。

難波ら [17] は、“regard ~ as” のような分離型イディオムを含む用例を検索することに焦点をあてている。分離箇所を節を含むような用例を検索するために、構文木上のノード間の距離を測る手法を提案している。提案手法の有効性を確認するため、提案手法、ベースライン手法（分離箇所について制限をしない単純マッチング手法と、分離箇所の単語数を 3 までに限定した手法）について検索の精度と再現率を調査する実験を行ったところ、分離箇所の距離が小さい場合はベースライン手法が、大きい場合は提案手法が優れていることが示された。

また、文章校正や機械翻訳の研究においても、用例検索の有効性が期待されている。

浜本による Native Checker⁸ は、入力した表現に誤りが含まれていた場合にその位置を検出し、訂正候補を出力する。さらに、訂正候補から表現を選択し、誤りを訂正して用例を検索することができる。例えば、“we show the resalt” というスペルミスを含んだクエリを入力して検索すると、図 9 のように、スペルミスが検出され訂正候補が表示される。この中から result を選択すると、図 10 のように、“we show the result” という表現を含んだ用例が提示される。

Chodorow ら [18] による ESL Assistant は、文章の自動訂正時に、訂正前の表現と訂正後の表現のそれぞれをクエリとして検索エンジンで検索した結果を提示す

⁷<http://med.tackpad.net/>

⁸<http://native-checker.com/>



図 9: NativeChecker : 誤りの検出



図 10: NativeChesker : 誤り訂正後の用例提示



図 11: ESL Assistant : 誤り訂正後の用例提示

る。その検索文書数やスニペットによって訂正の妥当性を示しており、これは用例検索のアプローチと類似している。例えば、“This is problem that I see every day. I am very interesting in solving this problem.”という英文を入力し Check のボタンをクリックすると、problem と interesting の部分がハイライトされ、誤りがあることが分かる。ここで、interesting の部分を選択すると、図 11 のように、訂正前の表現である “very interesting in solving” と訂正後の表現である “very interested in solving” のそれぞれが検索エンジンで検索され、それぞれの表現が含まれる英文の割合や、訂正後の表現が含まれる用例が確認できる。この結果から、学習者は interesting の部分を interested に訂正することが適切だと考えることができる。

大鹿ら [19] は、機械翻訳と検索エンジンを組み合わせた英作文支援に焦点をあてており、検索エンジンとして Google を用いた英作文支援システムを提案している。まず、機械翻訳により日本語の文を自動的に英文に翻訳したあと、英文の中

で気になる表現をシステムに入力し、その中でも特に検討したい部分を選択する。次に、選択された部分をワイルドカードに置き換えて、Google 検索を行う。システムははじめに選択された部分の品詞を、品詞分解データベース MontyLingua によって分析し、同じ品詞を含むものだけを Google 検索の結果から取り出してユーザーに提示する。ユーザーは提示された検索結果と検索結果件数を参考にすることで、よく使われている文型を知ることができる。この手法の有効性を確認するため、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ったあとこのシステムを使って修正を施す実験を1名の被験者を対象に行ったところ、冠詞や前置詞の検討、単純な SVO、SVC 構文における検討について高い精度が得られている。

このように、今日、英作文支援を行う様々なシステムが存在しており、用例検索に着目したシステムも数多い。しかし、綱島ら [14]、三好ら [15] によるアンケート調査や、坂本ら [16]、難波ら [17] による小規模な評価以外に、殆どの研究では、用例検索で英作文支援を行うためにはどのような検索機能が必要であるのかといった、教育学的な評価は行われてはおらず、主観的な評価も難しい。そのため、英作文における用例検索の有用性について、詳細な評価を十分に行える段階にならない。次章では、用例検索が英作文に有用であるかどうかについての評価を行えるシステムの設計のために、英作文における問題点について、用例検索の観点から分析を行う。

3 英作文における用例検索の有用性

日本人が英語で文章を書く際に生じる問題には様々なものがある。本章では、これらの問題に対する、用例検索による支援の可能性について述べる。

まず、日本人を対象とした英作文に関する文献 [6, 20, 21, 22, 23, 24] に基づいて、英作文の際の問題について、特に用例検索の観点から分析を行った。その結果、(1) 表現の適切性を確認したい場合、(2) 表現の組み合わせが分からない場合、(3) 表現の用法が分からない場合、の3つの場合に分類した。(1) は、クエリとして入力した表現がどの程度利用されているかに基づいて、その適切性を判断する場合であり、用例検索の最も基本的な利用方法である。(2) は、適切な表現の組み合わせを確認する場合である。(3) は、特定の表現を含む文を閲覧することにより、その表現の用法を確認する場合である。

次章以降では、それぞれの場合に、どのようなクエリが入力され、それに対してどのような用例を提示することが有効であるかについて述べる。

3.1 表現の適切性を確認したい場合

ある内容を表す表現として1つまたは複数の表現が想起できた場合、その表現が適切であるかを確認したいことがある。

例えば、「～をするための approach」という内容を表すために、“approach for ~”という表現を思いついたとする。この時、“approach for”をクエリとして入力する。それを文字列として含む表1のような用例を検索することができれば、その検索文数⁹や検索文の内容から、“approach for”という表現が頻繁に用いられているということが分かる。

3.2 表現の組み合わせが分からない場合

表現したい内容に対して、どのように表現を組み合わせたら良いか分からない場合がある。これは、組み合わせる表現の片方を言い換えたい場合と、ある表現

⁹ACL ARC[25] 内で検索した場合の検索文数である。

表 1: “approach for” で検索した際の用例検索の例

“approach for”で検索した結果：1,246 件
<ul style="list-style-type: none">•... we propose a multi-level editing <u>approach for</u> maintaining consistency•... an assembly-based <u>approach for</u> project-specific method construction• A kernel <u>approach for</u> learning from almost orthogonal patterns.• ... introduce the conventional <u>approaches for</u> automatic thesaurus construction in ...• This strategy is similar to our <u>approach for</u> answering definition questions.• Therefore we stress principled <u>approaches for</u> feature extraction and ...• In this paper we present a novel <u>approach for</u> building maximum entropy models.• ... we address the dearth of <u>approaches for</u> summarizing opinion information.

と組み合わせられる表現を俯瞰することにより選択したい場合に分類することができる。

3.2.1 表現を言い換えたい場合

英語のテクニカルライティングでは、同じ表現の繰り返しが嫌われる傾向にある [21, 22]。しかし、学習者の語彙が少ない場合は、繰り返しを避けるために他の表現を探すのに苦労することがある。

例えば、“results show that ...”という表現の“show”を別の語に言い換えたい場合、“results [V] that”と入力して、表2のような用例を検索することができれば、“show”の代わりに“indicate”や“demonstrate”を用いることが可能であると分かる。ここで[V]という記号は動詞を表し、あらゆる動詞が当てはまる。また、“indicate”と“demonstrate”の検索文数を比較すると、“show”と言い換える語としては“indicate”のほうがより一般的な表現であることが分かる。

3.2.2 組み合わせる表現が思いつかない場合

ある表現と組み合わせる表現として適切なものがまったく思いつかない場合について述べる。

例えば、“idea”という語と組み合わせる形容詞を俯瞰したい場合、“[J] idea”と入力して用例を検索することにより、表3のような用例を検索することができ、“main”や“key”、“novel”といった形容詞が用いられていることが分かる。

3.3 表現の用法が分からない場合

ある表現の用法が分からない場合、その表現を含む文を閲覧することにより、用法を確認できる。

例えば、“precision”について述べたい場合、“precision * %”と入力して、表4のような用例を検索することができれば、“precision of ~%”や“the precision is

表 2: “result [V] that” で検索した際の用例検索の例

“results [V] that”で検索した結果：2,543 件
<p>“results show that”を含む用例：1,156 件</p> <ul style="list-style-type: none">• Our <u>results show that</u> forward model adaptation alone ...• ... representation, however <u>results show that</u> system performance ...• The <u>results show that</u> the largest number of erroneous classifications occurred ...• These <u>results show that</u> the choice of classifier is extremely important in this task. <p>“results indicate that”を含む用例：429 件</p> <ul style="list-style-type: none">• Nevertheless, <u>results indicate that</u> our learning approach ...• ... and our own <u>results indicate that</u> the same description may ...• On the other hand, our <u>results indicate that</u> the proposed models are robust.• Thus , these <u>results indicate that</u> our corpus coding is adequately reliable and ... <p>“results demonstrate that”を含む用例：75 件</p> <ul style="list-style-type: none">• The <u>results demonstrate that</u> the sensor-system is a ...• Furthermore, our <u>results demonstrated that</u> recall plays a more important role ...• We do not believe, however, that the <u>results demonstrate that</u> the less ...• Experimental <u>results demonstrate that</u> the use of phrase-based translation ...

表 3: “[J] idea” で検索した際の用例検索の例

“[J] idea”で検索した結果：7,426 件
“main idea”を含む用例：274 件 <ul style="list-style-type: none">• The <u>main idea</u> behind this heuristic is to find ...• ... these components supported the <u>main idea</u> of the proof.• The <u>main idea</u> of the proposed approach is to compute the composition ...• ... make annotations freely and summaries which reflect the <u>main ideas</u> of the text.
“key idea”を含む用例：165 件 <ul style="list-style-type: none">• Our first <u>key idea</u> is to maximize the target signal to ...• The <u>key idea</u> is that a solution of the ...• The <u>key idea</u> in our method is to avoid the complexity hierarchical tree sturcture.• Our <u>key idea</u> is to detect the intervals and then compare the consecutive items ...
“novel idea”を含む用例：26 件 <ul style="list-style-type: none">• ... an ability to produce many <u>novel ideas</u> or solutions, a flexible approach to ...• This is not, of course, a <u>novel idea</u>, but we think that we have refined it in ...• It is not a <u>novel idea</u> to use machine learning in connection with ...• We propose a <u>novel idea</u> of forming solder microbumps on the ...

表 4: “precision * %” で検索した際の用例検索の例

<p>“precision * %” で検索した結果：3,353 件</p> <ul style="list-style-type: none"> • ... lexicon was found to have a <u>precision of 77%</u>. • Evaluation shows a <u>precision of about 30%</u>. • This conflating has a linguistic <u>precision of 99%</u>. • ... the performance of our overlapping disambiguation with <u>precision of 84.1%</u>. • ... shows the results, in which the <u>precision was improved to 50%</u> by ... • The <u>precision is improved to 62%</u> by refining ...
--

improved to ~%”といった言い回しが用いられていることが分かる¹⁰。この結果から、「precisionが~%である」と表現したい場合は“a precision of ~%”という表現が可能であり、「precisionが改善され~%となる」と表現したい場合は“the precision is improved to ~%”という表現が可能であると分かる。

¹⁰*はワイルドカードを表し、任意の単語が入る。

4 システムの設計

前章では、英作文における問題点を、用例検索の観点から分類・整理した。この結果に基づき、本章では、これらの問題点に対応可能な用例検索システムを構築するために、検索機能と検索結果の提示方法の2つに分けて、その設計方針を述べる。

4.1 検索機能

まず、英作文における問題に対して、用例検索が提供すべき検索機能について述べる。

4.1.1 拡張ワイルドカード検索

拡張ワイルドカード検索とは、ワイルドカード部に任意の単語列が入ることを許して、用例を検索する機能である。拡張ワイルドカードとは、1つから複数の任意の単語に対応する記号であり、これをアスタリスク (*) を用いて表記する。一般的なワイルドカードでは対応する語数を選択できないのに対して、拡張ワイルドカードは対応する語の上限数を学習者が自由に設定できる。これにより、ワイルドカード部分に必要以上に多くの語が対応してしまうことを防ぐ。

4.1.2 品詞検索

品詞検索とは、特定の品詞の語を含む用例を検索する機能である。拡張ワイルドカード検索では検索文数が多すぎる場合に、目的の表現を発見するのが困難になる。品詞検索を活用して、検索文数を絞り込むことによって、この問題に対応する。3.2.1で示した例のように、クエリに [V] といった品詞を含めることにより品詞検索を行う。

4.1.3 類義語検索

類義語検索とは、特定の語を含む用例だけでなく、その類義語が含まれている用例も検索する機能である。品詞検索より更に検索結果を絞り込みたい場合には、この機能が有効である。類義語検索で用例を検索する場合は、言い換えたい語（例えば“results show that”の“show”）に類義語検索を行う記号として“+”を付与（例えば“results show+ that”）して検索するとする。

4.2 検索結果の提示方法

用例検索を利用するという観点では、検索結果の提示方法を検討することも重要である。

4.2.1 統計表示

統計表示とは、ある表現を含む用例に対するコーパス中での出現数を表示する機能である。前章で用いた例における検索文数がこれに相当する。この機能によって、出現数の多い表現をより一般的な表現であると判断することができる。

4.2.2 用例の提示方法

用例を一覧表示する際、図??に示すように、クエリと一致した部分を中心に配置し、前後に文脈を表示する。これを KWIC (Key Word in Context) 方式と呼ぶ[2]。KWIC 方式の表示は、コーパス言語学や言語教育の分野など、幅広い場面で頻繁に用いられている。

4.2.3 用例の仕分けの方法

拡張ワイルドカード検索、品詞検索、類義語検索などで語を指定せずに用例を検索した際、拡張ワイルドカードや品詞に埋めた語で仕分けし、それぞれの統計情報を表示して欲しいことがある。例えば “[J] idea” というクエリで検索を行っ

た場合、図 12 に示すように、“[J]”の部分に当たる “basic”や “main”といった語で用例を仕分けし、それぞれの統計情報を表示する¹¹。

The screenshot displays a search interface with two main panels. The top panel is titled "[J] idea" : 1000 hits, 125 groups. Below it, the "basic idea" panel shows 217 hits. The "main idea" panel shows 138 hits. Both panels contain text snippets with highlighted terms and search bars.

**"[J] idea" : 1000 hits
125 groups**

basic idea : 217 hits

Length normalization represents a deviation from this **basic idea** .

To accomplish this , we relied on two **basic ideas** .

4.1 The **Basic Idea** As stated above , vectors in the same cluster must have a small distance between each other in the ideal geometry .

There are two **basic ideas** Brussels does not get , freedom and liberty .

We detail each below.1 The **basic idea** The complementarity between the pattern-based relation-oriented approach and the anchor-based argument-oriented approach as discussed above naturally leads us to consider combining them .

main idea : 138 hits

Our approach leverages on two **main ideas** .

For our purposes it also allows the straightforward graphical presentation of the **main ideas** .

Before giving a formal specification of our model , we will provide a high-level overview of the **main ideas** .

Yet , many of these approaches have a large number of parameters that can not be easily estimated.Our **main ideas** .

Additionally , these topics provide an informative picture of user search history.Terms with highest **Main Idea** A user's search history usually covers multiple topics

図 12: 用例の仕分け

¹¹[J] は、形容詞の検索を行うための記号である。

5 システムの実装

前章で述べた設計方針に従って、図 13 に示した、英作文支援のための用例検索システム PoEC (Partner of English Composition) を実装した。実装の際、Web アプリケーションフレームワークとして Django、全文検索エンジンとして Apache Solr を用いた。Django は、Python によって容易に Web アプリケーションを開発することができる MTV (Model Template View) フレームワークである。Apache Solr は、OSS の全文検索のエンジンとして有名な Apache Lucene をベースに、HTTP での入出力 (サーバアプリケーション化)、管理 Web アプリケーション、キャッシュ機構などの機能拡張を行った Java の Web サーバアプリケーションである。検索するフィールドを指定して高速な検索を行うことが可能であることから、今回は全文検索エンジンとして Solr を用いた。また、コーパスには、ACL Anthology 上の主要会議の論文データを用いた。英文数は約 1,340,000 文である。

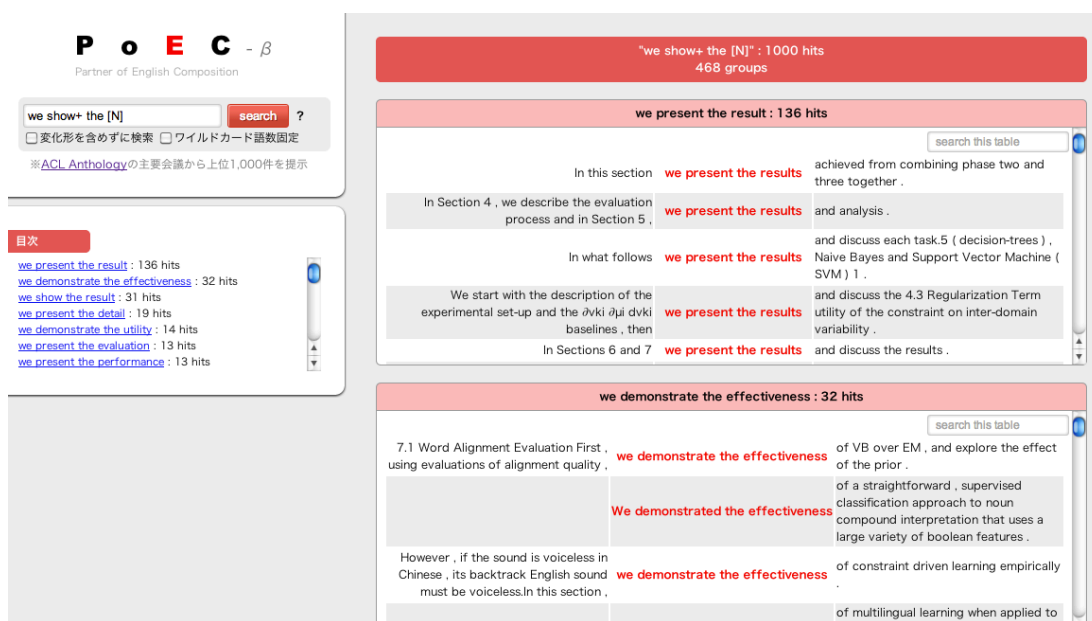


図 13: 用例検索システム PoEC

本システムで用例検索を行う大まかな流れを図 14 に示す。まず、入力されたクエリを Django へ送信し、Solr で全文検索を行うためのクエリに変換する。例えば、図 14 のように “[J] idea” というクエリが入力された場合は、base という名

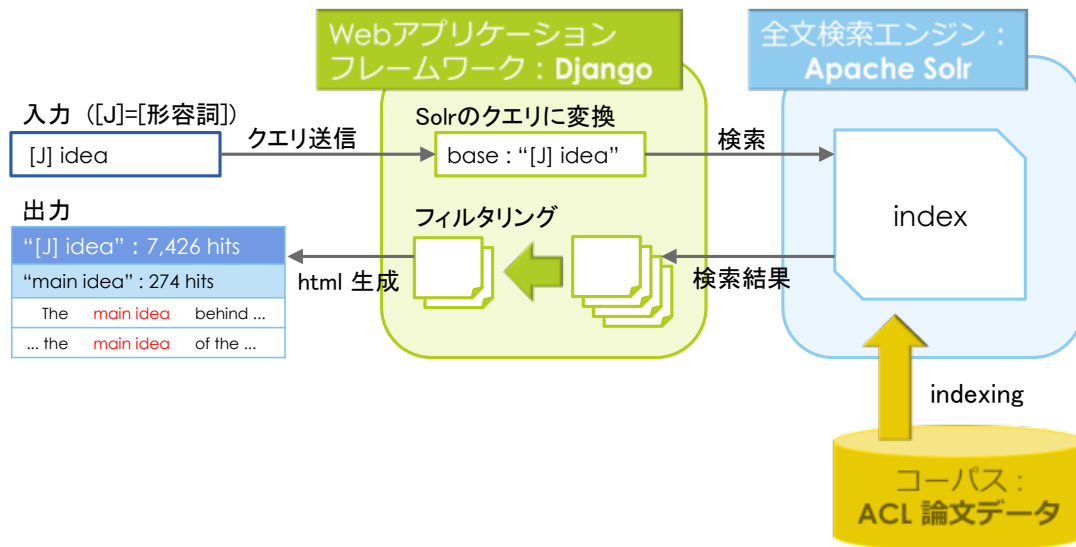


図 14: 用例検索の流れ

前のフィールドから “[J] idea” という文字列を全文検索するという Solr のクエリ 「base: “[J] idea”」 に変換する。この際、クエリに入力された単語は全て NLTK の WordNetLemmatizer により原形に変換する。Solr には変換されたクエリが送信され、全文検索が行われる。次に、検索結果を Django で取得し、はじめに入力されたクエリに合致しない余分な検索結果をフィルタリングによって取り除く。最終的に、残った検索結果から html を生成し、ユーザーに提示する。

コーパスに含まれる英文の indexing は、表 5 のように、1 つの英文に対して合計 3 つのフィールドを設けることにより行う。

表 5: 英文の indexing

id	P08-1062.tok.gtag-28
sent	The main ideas are as follows .
base	The main idea be as follow .

id というフィールドには、英文が収録されている論文ファイルのファイル名と、その英文が論文内の何番目の文かを示す番号を組み合わせ、英文固有の id を登録する。例えば、表 5 における id は、ファイル名 P08-1062.tok.gtag と英文番号 28 を組み合わせたものである。sent というフィールドには、英文そのものを

登録する。base というフィールドには、sent フィールドに登録された英文に含まれる全ての単語を原形に変換したものを登録する。原形への変換には、GENIA Tagger の解析結果を用いた。

このように、それぞれのフィールドに対応するデータを格納しておくことにより、フィールドを指定した場合に、そのフィールドに格納されているデータから検索が可能となる。

本システムでは、時制や数が異なる表現をまとめて検索するため、通常は、クエリに含まれる各単語を原形に変換して base フィールドから検索を行い、sent フィールドの英文を用例として提示する。例えば、“research is” というクエリを入力した場合は、クエリを “research be” に変換し、さらに base フィールドから検索を行う検索式「base:“research be”」に変換する。この検索式により、base フィールドに “research be” を含む英文を検索することができ、検索結果として、“research is” や “research was”、“research are” などを含む用例が提示される。特定の時制や数について検索を行いたい場合は、図 13 に示した「変化形を含めずに検索」のチェックボックスにチェックを入れて検索することで、クエリを原形に変換せずに sent フィールドから直接検索できる。

また、indexing を行う際に、Solr の類義語検索機能を応用することで、sent フィールド、base フィールドに登録された各英単語に、品詞を紐付ける。Solr の類義語検索機能とは、表 6 に示したような内容の synonyms.txt というファイルを用意することで、全文検索を行う際に、類義語指定した単語を同時に検索できる機能である。

表 6: 類義語を指定する synonyms.txt の例

Microsoft, マイクロソフト
Google, グーグル
マクドナルド, マック, マクド

synonyms.txt に表 6 のように記述して indexing を行うことで、“マイクロソフト” というクエリで “Microsoft” という表現が含まれるデータを検索したり、“マック” というクエリで “マクドナルド” という表現が含まれるデータと “マクド” という表現が含まれるデータを同時に検索したりすることが可能となる。

この機能を利用して、本システムでは、表7のような synonyms.txt を用意する。このような synonyms.txt により、英単語と品詞を結びつけることで、特定

表 7: 本システムで indexing の際に用いる synonyms.txt

attempt, [N]
attempt, [V]
appendix, [N]
anything, [N]
...

の品詞の単語を検索することが可能となる。例えば、表7のように synonyms.txt を記述すると、“[N]”というクエリで“attempt”を含む文、“appendix”を含む文、“anything”を含む文が検索できる。また、attempt のように、複数の品詞になり得る単語は、表7のように、なり得る品詞を全て記述する。この synonyms.txt は、検索対象の論文データに含まれる全ての英文の各単語について GENIA Tagger によって品詞タグ付けを行った結果から自動で生成する。品詞タグ付けを行う際に、Tagger が誤った品詞タグを付けることがあるため、ある単語に対応する品詞の中で、タグ付けされた回数が極端に少ない品詞は synonyms.txt には記述しないことにする。

なお、表7における attempt に対する2行の記述は、“attempt, [N], [V]”という1行の記述と同様の意味をもつ。

5.1 検索機能の実装

前章で述べた設計方針に基づいて実装した主な検索機能について述べる。

5.1.1 拡張ワイルドカード検索

入力されたクエリ内にアスタリスク (*) が含まれている場合、クエリを Solr の近傍検索を行う検索式に変換して全文検索を行う。Solr の近傍検索では、「フィールド:“単語 単語” 距離」のような検索式により、指定したフィールドから、複数の単語間の距離を指定した検索ができる。

例えば、“we *2 that”というクエリが入力された場合は、「base:“we that” 2」という検索式に変換する。これにより、base フィールドから、we と that の間に 2 語以内の単語を含む文を検索でき、検索結果として以下のような用例が提示される。

- Like previous researchers, we found that CRF models outperform ME models.
- We can see that the extension is very straightforward.

この際、上記のように、we と that の間が 1 語のものも検索されてしまう。間に 2 語含む文だけを検索結果として表示したい場合は、図 13 に示した、「ワイルドカード語数指定」のチェックボックスにチェックを入れて検索する。これにより、Solr の全文検索の結果に対して、正規表現によるフィルタリングが行われ、ワイルドカードに対応する語数は、アスタリスクの後ろに指定した数字と等しい数になるため、we と that の間が 2 語の用例だけが検索される。

5.1.2 品詞検索

英文の indexing の際に、7 のような synonyms.txt を用意することで、Solr の類義語検索機能により、単語に品詞を結びつけることができる。このため、クエリに [V] などの品詞を含めることで、特定の品詞について検索することができるようになる。例えば、“we [V] that”というクエリが入力された場合は、検索結果として以下のような用例が提示される。

- We found that this worked surprisingly well.
- Similarly, our scoring function is simplistic and we believe that it can be improved.

5.1.3 類義語検索

入力されたクエリ内の単語に “+” が付与されている場合、その単語の類義語を WordNet から取得し、OR 検索を行う。例えば、“we show+ that”というクエリが

入力された場合は、WordNet から show の類義語として indicate、demonstrate といった語が取得できるため、クエリを「base:“we show that” OR base:“we indicate that” OR base:“we demonstrate that” OR ...」という検索式に変換する。これにより、base フィールドから “we show that”、“we indicate that”、“we demonstrate that”…のいずれかを含む文を検索することができ、検索結果として以下のような用例が提示される。

- We show that the proposed method significantly outperforms existing Japanese text processing tools.
- While mixture components are difficult to interpret, we demonstrate that the patterns learned are better than random splits .

“+”が付与されている単語の類義語が WordNet 内に存在しない場合は、類義語検索を行わず、クエリから “+”を除去した形で検索する。例えば、“we+ show that”というクエリが入力された場合は、we の類義語は存在しないため、“we show that”というクエリに変換され検索される。

5.2 検索結果の提示方法の実装

前章で述べた設計方針に基づいて実装した、用例の主な提示方法について述べる。また、本システムを試作している段階で、ユーザーから用例の一覧性を高める機能の提案を受けたため、合わせて述べる。

5.2.1 KWIC 表示

用例内の入力されたクエリに対応する部分の前後に html タグを挿入することによって、クエリに対応する部分を中心に配置し、ハイライトする。これにより、図 13 に示したように、クエリに対応する部分を中央に配置し、赤くハイライトして用例を提示する。

5.2.2 用例の仕分けと統計表示

Solr から Django に送信された検索結果を、クエリに対応する文字列を key、検索文を value にした辞書構造に変換する。クエリにワイルドカードや品詞が含まれる場合はその部分を具体的な単語で埋めたそれぞれの表現を key とし、類義語検索する語が含まれる場合は OR 検索されたそれぞれの表現を key とする。例えば、“we *2 that” というクエリが入力された場合は、“we can see that” や “we will see that” といった表現が key となり、“we [V] that” というクエリが入力された場合は、“we find that” や “we believe that” といった表現が key となり、“we show+ that” というクエリが入力された場合は、“we show that” や “we indicate that” といった表現が key となる。これにより、検索された用例を、ワイルドカードや品詞に含まれる語、類義語ごとに仕分けすることができる。html を生成する際には、辞書の key になっている部分を KWIC 表示によりハイライトする部分とみなす。

また、1つの key に対応する value の数を数えることにより、ある表現が含まれる用例の数が分かるため、その用例数も合わせて表示する。

5.2.3 目次の表示

検索結果として多くの用例が提示された場合、ページ上部に表示された用例から下部に表示された用例までを見渡し、目的の用例を見つけやすくするため、図 13 に示したような目次を表示する。目次は、用例の仕分けを行う際の key の内容と value の数から自動的に生成する。

5.2.4 提示された用例のフィルタリング

FilterTable という jQuery を利用することで、提示された用例を、指定した表現でさらにフィルタリングする機能を実装する。例えば、“we show” というクエリで検索を行った後に、検索結果のテーブルの右上に表示されているテキストボックスに “result” と入力すると、“we show” と “result” の両方を含んでいる用例が表示される。ここで、“result” が含まれているセルは FilterTable の機能によりハイライトされる。

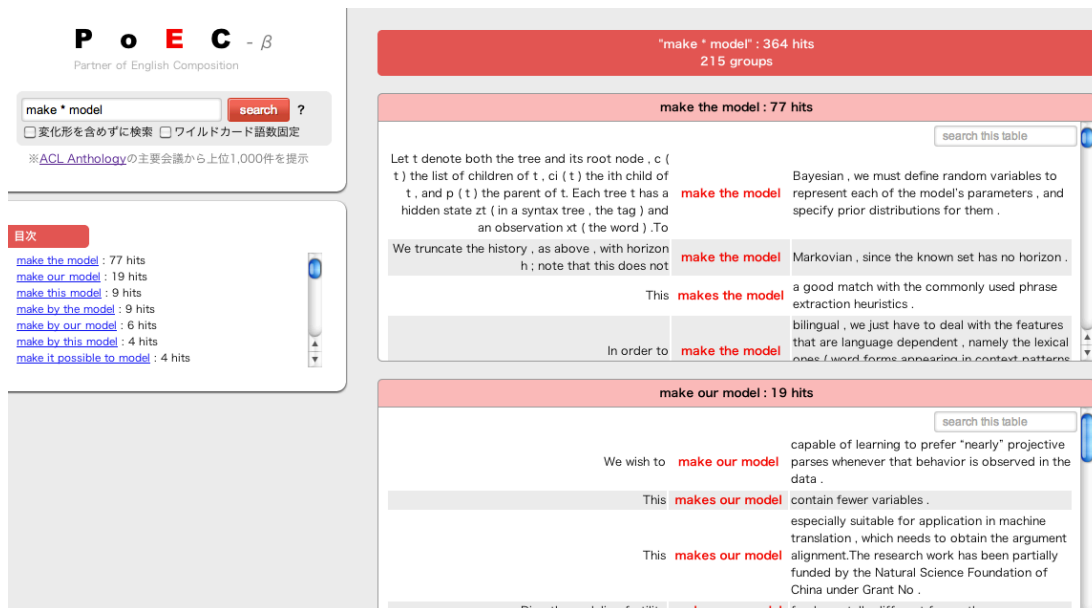


図 15: “make * model” の検索結果

5.3 システムの利用例

実際に英作文を行う状況を想定し、本システムの利用例について述べる。

「モデルを作る」と書きたい場合 英語で「モデルを作る」と書きたい場合に、make ~ model や、build ~ model という表現を思いついたとする。このとき、この2つの表現がそれぞれ含まれる用例を閲覧することによって、実際にこれらの表現がどのように利用されているかを確認する。“make * model”、“build * model”というクエリを入力してそれぞれ検索すると、図15、図16のような画面が提示される。検索文の内容や、“build * model”の検索文数が“make * model”の検索文数よりも多いことから、build ~ model という表現が適していることが分かる。

「カバレッジが広い」と書きたい場合 英語で「カバレッジが広い」と書きたい場合に、large coverage や high coverage という表現を思いついたとする。このとき、これらの表現が実際に利用されているかどうかを確認するために、“large coverage”、“high coverage”というクエリを入力してそれぞれ検索す

P o E C - β
Partner of English Composition

build * model search ?
 変化形を含めずに検索 ワイルドカード語数固定
 ※ACL Anthologyの主要会議から上位1,000件を提示

目次

- [build a model](#) : 51 hits
- [build the model](#) : 28 hits
- [build language model](#) : 18 hits
- [build a language model](#) : 16 hits
- [build statistical model](#) : 10 hits
- [build a translation model](#) : 9 hits
- [build the language model](#) : 8 hits

"build * model" : 696 hits
422 groups

build a model : 51 hits

We also **built a model** (11) based on one empirically optimized discount $D = 0.55$ and an empirically optimized value of $\delta = 0.9$.

To **build a model** and score the same essay sample would take up to an hour .

) Here , for simplicity , we restrict the ROSE model to use only the "official" ROUGE metrics to **build a model** based on a given year and then evaluate that model on a subsequent year .

We do not just **build a model** based on latent factors , but we use the latent factors to determine which of the features in the original word vector are the salient ones given a particular context .

build the model : 28 hits

This is used as ground truth for **building the model** .

The toolkit supports German and English ; and it uses the openNLP MAXENT package1 to **build the model** .

86 % F1 for the "off-event" class when only 50 % of the IBC data is used to **build the models** .

The algorithm locally maximizes the likelihood function via an iterative procedure.We used the following TMLM parameters to **build the model** .

Although the usage of the two types of model is **building the models** .

図 16: “build * model”の検索結果

P o E C - β
Partner of English Composition

[J] coverage search ?
 変化形を含めずに検索 ワイルドカード語数固定
 ※ACL Anthologyの主要会議から上位1,000件を提示

目次

- [broad coverage](#) : 138 hits
- [good coverage](#) : 98 hits
- [high coverage](#) : 86 hits
- [low coverage](#) : 77 hits
- [lexical coverage](#) : 66 hits
- [wide coverage](#) : 53 hits
- [full coverage](#) : 33 hits

"[J] coverage" : 1000 hits
146 groups

broad coverage : 138 hits

The more the theory uses rich representations the more the system may achieve **broad coverage** .

It is currently unknown how to produce a deep understanding program with **broad coverage** .

This has aided in the Japanese grammar rapidly achieving **broad coverage** .

Therefore , we merged entries from E-to-J translation dictionary in order to get much **broad coverage** .

Compared with the multinomial mixture , one advantage of our method is that it has **broad coverage** .

Therefore , we do not need to use any backoff **broad coverage** .

good coverage : 98 hits

have a **better coverage**) .

Redundancy is allowed to achieve **better coverage** .

(2010) obtains superior results and has **better coverage** .

For larger n-grams , however , the FSMLM model has the **best coverage** .

We want to know whether a very simple interlingua like IF can have **good coverage** .

Interestingly , although Lin is smaller than WASPS it has **better coverage** .

図 17: “[J] coverage”の検索結果

ると、どちらも検索文数が少なく、表現として適切かどうかを判断することが難しい。そこで、クエリを一段階曖昧にして、形容詞を表す [J] という表記を用いて、“[J] coverage”というクエリを入力して検索すると、図 17 のように、broad coverage や wide coverage という表現が見つかる。これらの表現は、検索文数や検索文の内容を閲覧しても、適切だと判断できる。

6 予備実験

前章で構築した用例検索システムを利用して、用例検索が英作文に有用かどうかを評価するための予備実験を実施する。システムのユーザビリティ評価に関する書籍 [26, 27] によると、大人数で一度の評価実験を行うよりも、少人数で繰り返し評価実験を行ったほうが、システムに関する重要な問題を多く発見することができる。そのため、本研究では、まず、少人数での予備実験を二回行うことで、システムや実験方法の問題点を発見する。その後、発見した問題点を改善した上で、評価実験を行う。

6.1 節では第一回予備実験について、6.2 節では第二回予備実験について述べる。

なお、本章以降では、本システムの拡張ワイルドカード検索を、単にワイルドカード検索と呼ぶ。

6.1 第一回 予備実験

6.1.1 実験設定

第2章で言及した通り、今日、英作文支援のために用例検索に着目した様々なシステムが存在しているが、殆どの研究では、用例検索で英作文支援を行うためにはどのような検索機能が必要であるのかといった教育学的な評価は行われていない。そのため、まずは、教育学的な評価を行うことができるような実験を設計する必要がある。

英作文支援のための用例検索の有用性を評価するための実験の形式は、以下の3つが考えられる。

- TOEIC の穴埋め問題を試験問題とする実験
- 自作の穴埋め問題を試験問題とする実験
- 1文を英作文する問題を試験問題とする実験

TOEIC の穴埋め問題とは、図 18 のような問題である。これは、前後の英文から問題となっている文の内容を予測し、解答を4つの選択肢から選択する問題で

I am ----- to confirm our offer of part-time employment at Western Enterprises.
 141. (A) pleased
 (B) pleasing
 (C) pleasant
 (D) pleasure

図 18: TOEIC の穴埋め問題

問題例	
1	
	How much does the described algorithm improve performance of a state-of-the-art QA 3 . We use all factoid questions in TREC' s QA test sets from 2002 to 2006 for evaluation for which a known answer exists in the AQUAINT corpus . Additionally , the data in (Lin and Katz , 2005) is used . In this paper the authors attempt to identify a much more complete set of relevant documents for a subset of TREC 2002 questions than TREC itself .
(何もなし)	We adopt a cross validation [] our evaluation .
	評価には、交差検定法を採用する。

図 19: [第一回予備実験] 英文穴埋め問題

ある。試験問題として TOEIC の穴埋め問題を利用する場合、選択肢があるため、選択肢を 1 つずつ用例検索システムで調査すると容易に正解に辿り着ける可能性がある。また、本研究で支援する対象としている、書きたい内容は思いついている上で、その内容に対応した英作文を行う状況とは大きく異なっていると考えられる。一方、試験問題として 1 文を英作文する問題を利用する場合、被験者・採点者の負担が非常に大きくなってしまふことが想定できる。以上より、第一回予備実験では、試験問題として自作の穴埋め問題を利用することとする。

試験問題として、Huang ら [28] の実験手法を参考に、図 19 に示す英文穴埋め問題を制作する。図 19 において、青い枠で囲まれたセルが穴埋め問題であり、穴埋め問題の上のセルは文脈を表す英文、穴埋め問題の下のセルは穴埋め問題に対応した日本語訳である。

テクニカルライティングを支援する状況を想定して、この問題を以下の手順で制作する。

1. 英語論文でよく使われる表現を集めた表現集 [29] に記載されている表現の

中から、特に重要だと考えられる表現を抜き出し、重要表現リストを作る。
この重要表現リストは、英語論文執筆上級者2名が選択した表現を統合したものである。

2. 重要表現リストの表現を含む英文とその文以前の5文を、ACL Anthology上の2012年の論文データから抽出する。
3. 重要表現リストの表現を含む英文は、重要表現部分を穴埋め部分とし、穴埋め問題とする。また、その文以前の5文は、文脈を表す英文として利用する。
4. 穴埋め問題の英文を日本語訳する。
5. Excelシート上で、文脈を表す英文、穴埋め問題、穴埋め問題に対応した日本語訳の順に表示し、解答シートを作成する。

なお、手順4.における英文の日本語訳は、筆者が行う。

上記の方法で、合計40問を含む解答シートを作成し、英語を母語としない日本人学生2名に対して実験を行う。実験では、以下に示すように、4通りの問題の解き方を設定する。

辞書のみ 本システムやGoogleを利用せずに、Mac OSXに標準搭載されている辞書を用いて解答する。ここで用いる辞書は、プログレッシブ英和・和英中辞典に限定する。

PoEC(simple) 辞書に加え、本システムを利用して解答する。ただし、ワイルドカード検索、品詞検索、類義語検索を利用しないこととする。

PoEC(full) 辞書に加え、本システムを利用して解答する。ワイルドカード検索、品詞検索、類義語検索を利用してもよい。

Google 辞書に加え、Googleを利用して解答する。ただし、Googleで検索を行う場合は、日本語のクエリは入力しないこととする。

2013-12-17 10:16:14.614369	seeds of [n]
2013-12-17 10:16:29.045269	seed of [n]
2013-12-17 10:16:45.000112	seed [i]
2013-12-17 10:19:40.938726	seed for [n]
2013-12-17 10:20:16.346122	seed of [n]
2013-12-17 10:20:26.126662	seed of *
2013-12-17 10:22:47.763144	[i] this study
2013-12-17 10:23:11.825330	in this [n]

図 20: 本システムのクエリログ

被験者は、全 40 問の穴埋め問題を、以上の 4 種類の解き方を 1 問ずつ順番に用いることで解答する。具体的には、問題 1 は辞書のみでの解き方で、問題 2 は PoEC(simple) の解き方で、問題 3 は PoEC(full) の解き方で、問題 4 は Google の解き方で、問題 5 は辞書のみでの解き方で、というように 1 問ずつ解き方を変更して繰り返すことで解答を行う。これは、長時間同じシステムを利用することで 1 つのシステムに慣れてしまい、利用するシステムによるバイアスが生じることを防ぐためである。以上の手順で解答を行うと、それぞれの解き方で解答する問題数は各 10 問ずつとなる。また、40 問は全て異なる内容の問題であるため、それぞれの解き方で解答する問題は、別々の内容となる。

なお、実験を行う前に、問題形式と問題の解き方について口頭で説明を行う。

また、実験全体を通して、本システムのクエリログを収集する。クエリログでは、図 20 のように、クエリが入力された日時と、クエリの内容が記録されている。クエリログは、各被験者ごとに別々に収集する。

6.1.2 評価方法

2 名の被験者の穴埋め問題の解答について、以下の 2 種類の方法で点数を計算する。

通常採点 意味・構文が合っている場合は 2 点、意味・構文に大きな誤りはないが修正したほうが良い場合は 1 点、意味・構文が誤っている場合は 0 点として、点数を計算する。

厳密採点 穴埋め部分の解答が表現集 [29] の表現と完全一致の場合は 1 点、異なっ

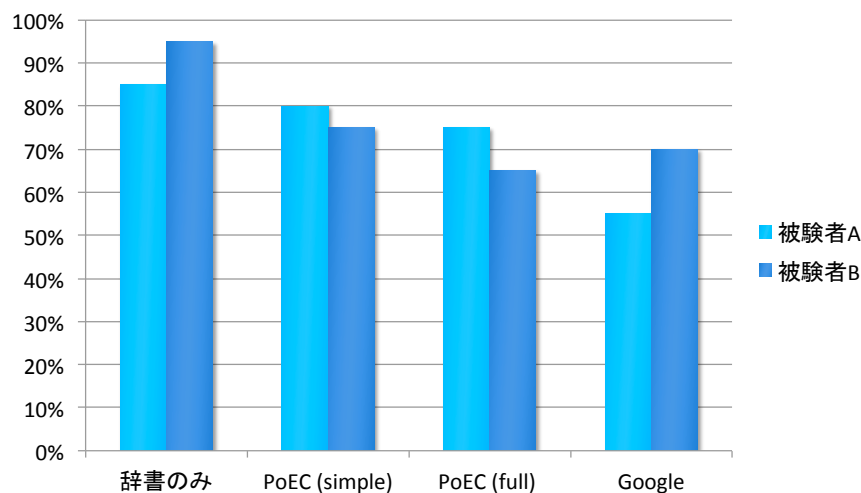


図 21: [通常採点] 正解率

ている場合は0点として、点数を計算する。

以上の方法で計算した点数を、点数の最大値で割ることによって、正解率を計算する。前節で示した4種類の解き方について、それぞれ正解率を計算して比較することにより、どの解き方が有用であるかを確認する。なお、通常採点は、英語を母国語とする採点者1名が行い、厳密採点は筆者が行う。

6.1.3 実験結果

2名の被験者の実験結果について、通常採点で採点した場合の正解率を図21に、厳密採点で採点した場合の正解率を図22に示す。

本システムやGoogleを利用することで、辞書のみで問題を解いた場合よりも正解率が上昇することを期待していたが、図21より、被験者A、被験者Bともに、辞書のみの場合の正解率が最も高くなってしまったことがわかる。また、本システムを利用して解く場合にも、PoEC(simple)よりPoEC(full)の場合のほうが正解率が上昇することを期待していたが、PoEC(full)の場合のほうが正解率が低下していることがわかる。

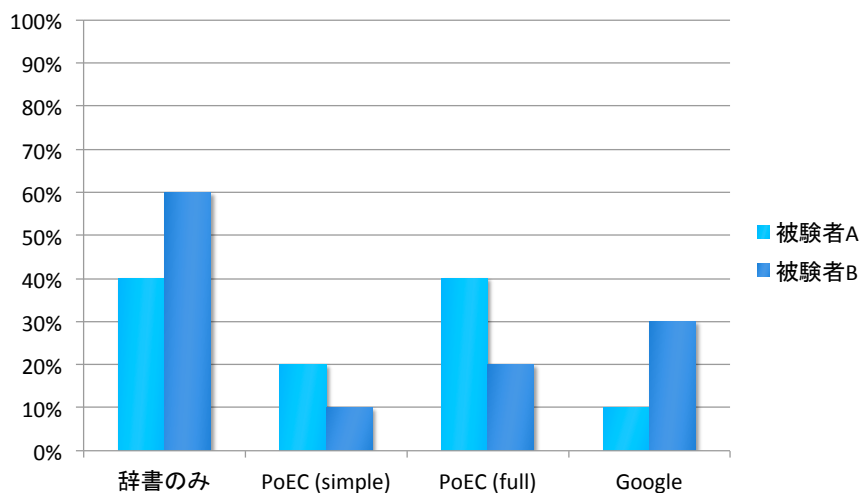


図 22: [厳密採点] 正解率

図 22 より、厳密採点をした場合でも辞書のみで正解率が最も高くなっていることがわかる。

以上の結果から、辞書のみで解答する問題が他の解き方で解答する問題と比較して平易な内容であったことが推測できる。問題数が少ない場合、問題の難易度差が大きく影響するため、問題の内容が異なると、正解率を比較することでどの解き方が有用であるかを確認することは難しいことがわかる。そのため、解き方の有用性を比較するためには、同じ内容の問題について解答する必要があるといえる。

本システムを利用して解答を行った問題について分析を行うためにクエリログを閲覧したところ、各被験者がどのようなクエリを利用して検索を行っているかを読み取ることはできるが、実際にどのようなクエリを入力した場合にどのような用例を参考にしているかを読み取ることはできず、どのような検索機能が役立っているかは判断できなかった。そのため、実験の記録は、クエリログのみでは不十分であるといえる。

実験終了後に、被験者から以下のようなコメントを得ることができた。

- 本システムや Google を利用して検索を行うことで表現を確認することがで

きるのは自信につながる

- 辞書のみで解答する場合は、用例を参考にできないので、不安である
- 本システムで用例検索を行う際に、品詞検索を行うと、品詞の部分に対応する単語ごとに用例を仕分けして表示されるため、どの単語がどのように利用されているのかを読み取りやすい
- 本システムで日本語のクエリを入力することで用例を提示してほしい
- 文頭を検索する機能がほしい
- 本システムで検索結果が0件だった場合でも、何かしらの情報を提示してほしい
- 1問ずつ解き方を変えて解答するのは負担が大きく、解き方を誤ることもある
- 穴埋め問題の形式だと、穴埋め部分の前後に解答の形式を合わせる必要があることがストレスである
- 平易な慣用句について問う問題が多く、辞書だけで足りてしまう
- 文脈を表す英文を読んでいる余裕がなく、参考にすることはほとんどなかった

以上のコメントより、本システムや Google を利用している際に、多くの用例を確認することができて自信につながっていることがわかる。解答の際に、解答に対する自信を記述してもらうことで、辞書のみで解答を行う場合と、本システムや Google を利用して解答を行う場合で、実際に自信が上がっているかどうかを確認できると考えられる。

また、本システムに関して、日本語のクエリを入力して検索を行う機能や、用例の文頭を検索する機能、検索結果が0件だった場合に他の情報を提示する機能が求められていることがわかる。これらの機能については、ワイルドカード検索

や品詞検索、類義語検索といった主要な機能の有用性を確認することができた後に実装を検討する。

実験手順については、解き方を1問ずつ変更する方法は被験者にとって負担が大きく、解き方を誤る可能性があることがわかる。今後問題数を増やして大規模な評価実験を行うことを想定すると、被験者にとって負担が大きい解答方法は避けるべきである。また、解き方を誤ると、それぞれの解き方について正しい正解率を計算することができなくなり、有用性を確認することができなくなるため、被験者が解き方を謝らない実験方法にする必要がある。

問題形式については、ある文型の解答を思いついても、穴埋め部分の前後に合わせるために文型を変更する必要がある、別の文型を思いつかない場合に解答を行うことができないことがあった。これは、実際の英作文では発生し得ない制約であり、穴埋め問題の形式は、英作文を行う状況とは大きく異なるといえる。そのため、問題の形式を変更する必要がある。

6.1.4 考察

自作の穴埋め問題を利用した評価実験を行った結果、穴埋め問題を解く形式は、実際に英作文を行う状況とは大きく異なり、試験問題の形式を変更する必要があることがわかった。実験手順について、解き方を1問ずつ変更する方法には問題があることがわかったため、実験手順の変更も行う必要がある。

また、クエリログを収集するのみでは、実際にどのような検索機能を利用してどのような用例を参考にした上で解答しているかを読み取ることができないことがわかった。この問題は、実験中に、パソコンの画面をキャプチャする動画を撮影することで対応できると考えられる。

以上の結果を踏まえ、第二回予備実験は、試験問題の形式や実験の手順を変更して実施する。また、画面をキャプチャする動画を撮影することで実験の記録を行う。

1		
和文	本研究はまだ進展中である。	自信
英文 take1		
英文 take2		

図 23: [第二回予備実験]1 文を英作文する問題

6.2 第二回 予備実験

6.2.1 実験設定

前節では、穴埋め問題の形式は評価実験に適していないことがわかったため、第二回予備実験は、1 文を英作文する問題形式を採用することとする。この実験では、1 文を英作文する問題形式で実験を実施することで、用例検索の各検索機能の有用性を評価できるかどうか、被験者・採点者の負担が実際にどの程度であるのかを調査する。

試験問題として、図 23 に示す英作文問題を制作する。図 23 において、問題文は和文と表記してある部分の日本語の文で、この文に対応する英文を、英文 take1 という部分と英文 take2 という部分に解答する形式となっている。また、解答に対する自信を、解答の右のセルに、◎、○、△、×の4種類から選択して入力する。最も自信が高い場合は◎、まったく自信がない場合は×を入力する。

問題文の元となる英文は、前節と同様に、重要表現リスト内の重要表現を含んでいる英文を ACL Anthology 上の 2012 年の論文データから抽出することで得る。この英文を筆者が和訳したものを、問題文として提示する。

第一回予備実験のコメントより、被験者に文脈を表す英文を提示する必要がないと判断したため、第二回予備実験では、文脈を表す英文は提示しない。

上記の方法で、合計 10 問を含む解答シートを作成し、英語を母語としない日本人学生 4 名に対して実験を行う。4 名の被験者のうち、2 名は英語論文執筆経験・読解経験があり、2 名は英語論文読解経験がある。この 4 名を、表 8 のように分類する。これは、英語論文執筆経験の有無により、解答の出来が変化するかどうかを調査するための分類である。

実験は、以下の手順で行う。

表 8: 被験者の分類

利用するシステム	執筆経験 有	執筆経験 無
Google	被験者 A	被験者 B
本システム	被験者 C	被験者 D

1. 英作文問題 10 問を、Mac OSX に標準搭載されている辞書を用いて解答する。ここで用いる辞書は、プログレッシブ英和・和英中辞典に限定する。
2. 被験者 A、B は、10 問を、辞書と Google を利用して解き直す。被験者 C、D は、10 問を、辞書と本システムを利用して解き直す。

以降、上記の手順 1 を 1 周目と呼び、手順 2 を 2 周目と呼ぶこととする。1 周目の解答は、解答シートの英文 take1 という部分に入力し、2 周目の解答は、英文 take2 という部分に入力する。2 周目では、1 周目の解答を修正する必要があると感じられた場合に、1 周目の解答を修正したものを解答とする。修正の必要性が感じられなかった場合には、1 周目と全く同様の解答となってもよい。1 周目の解答と 2 周目の解答を比較することにより、ある問題について、本システムや Google を利用することでどのように誤りを改善できているかを読み取ることができると想定される。

実験の記録は、前節で述べた通り、パソコンの画面をキャプチャした動画を撮影することで行う。動画の撮影には、Mac OSX に標準搭載されているマルチメディアソフト QuickTime を利用する。

なお、実験を開始する前に、各被験者に本システムと Google の検索機能について口頭で説明を行う。また、実験の際には、1 周目と 2 周目の間に、約 10 分の休憩を挟む。

6.2.2 評価方法

4 名の被験者の英作文問題の解答について、以下の 2 種類の 방법으로評価を行う。なお、以下の評価は、英語を母語とする採点者 1 名が 4 名の被験者の全解答について誤り部分を発見・修正した結果を閲覧することで、筆者が行う。

改善度合いの分類 1周目と2周目の解答を比較して、各問題を、

- a : 誤りを含まない表現が、更に改善された場合。また、表現が変化していても、自信が上がっている場合。
- b : 誤りを含まない表現が、誤りを含む表現に書き換えられた場合。
- c : 解答の内容が書き換えられ、誤りが改善された場合。
- d : 解答の内容が書き換えられているが、誤りが改善されていない場合。また、解答の内容が変化していない場合。

の4つに分類する。

三段階採点 各問題を、

- : 誤りを含まない文。
- △ : 問題文と意味は合致しているが、表現に誤りを含む文。
- × : 問題文と意味が合致しておらず、表現も誤っている文。

の三段階で採点する。

改善度合いの分類を行うことで、本システムや Google によって改善できている問題、改善できなかった問題がどの程度あるかを知ることができる。また、三段階採点を行い、○、△、×の割合を確認することで、正解率を計算することができる。特に、1周目の解答の正解率は、各被験者のもともとの英語能力と判断することができる。

6.2.3 実験結果

4名の被験者の実験結果を図24に示す。図24における解答時間は、撮影した動画の長さから算出した1問ごとの平均解答時間である。図24より、○と採点されている問題数に着目して各被験者の1周目の正解率を計算すると、被験者Aは30%、被験者Bは30%、被験者Cは20%、被験者Dは20%となる。各被験者の1周目の解答の正解率の差が小さいことから、英語論文執筆経験の有無による実

		1周目 (辞書のみ)		2周目 (辞書+システム)		改善度分類	
		正解率	解答時間	正解率	解答時間		
Google	有	被験者A	○:3 △:5 ×:2	3:45 / 問	○:3 △:4 ×:3	2:16 / 問	a:3 b:1 c:4 d:2
	無	被験者B	○:3 △:7 ×:0	2:32 / 問	○:4 △:5 ×:1	2:38 / 問	a:2 b:1 c:3 d:4
PoEC	有	被験者C	○:2 △:8 ×:0	3:28 / 問	○:1 △:9 ×:0	2:39 / 問	a:1 b:2 c:4 d:3
	無	被験者D	○:2 △:6 ×:2	2:59 / 問	○:3 △:5 ×:2	2:47 / 問	a:2 b:1 c:6 d:1

図 24: [第二回予備実験] 実験結果

力差はあまりないことがわかる。この結果より、評価実験の被験者として英語論文読解経験のある者を選択する場合は、英語論文執筆経験の有無は考慮しなくてよいと考えられる。

各被験者の解答時間から、英作文問題の解答は、1周目、2周目ともに、1問につき約2分から約4分で行われていることがわかる。

また、改善度合いの分類は、全ての被験者を通してaとcが多いことがわかる。そのため、本システムまたはGoogleを利用することで、解答を改善できる部分が多いといえる。bに分類されている問題は少なく、本システムやGoogleを利用することで解答が悪化することは少ないことがわかる。

以上より、1問につき約2分から約4分の時間があれば、1周目では辞書を利用して十分に解答を作成することができ、2周目では本システムやGoogleを利用して誤りを改善することができることがわかる。

なお、英語を母語とする採点者1名が被験者1名の全問題の解答に含まれる誤りを発見・修正するのに、約30分の時間がかかった。

以上の結果より、1文の英作文問題に解答する方法は、被験者・採点者の負担が大きくないことがわかる。

ここで、各被験者の解答について、画面をキャプチャした動画を閲覧することで、さらに詳細に分析を行った結果を以下に示す。

まず、2周目で解答を改善することができたもの（cに分類されたもの）の一つ

である、問4に対する被験者Dの解答について述べる。問4は、「改善の余地が残されている。」という和文に対応する英文を解答する問題である。この問題に対して、被験者Dは、1周目では“*There is a room for improvement.*”、2周目では“*There is room for improvement.*”という解答を行っており、1周目の解答は冠詞の誤りを含んでいるが、2周目では誤りが改善されている。この部分について動画を閲覧すると、被験者Dは、本システムに“*room for improvement*”というクエリを入力して検索することで、1周目の解答を確認しようとしていることがわかる。この検索により、被験者Dは、“*room*”に冠詞がついている用例がほとんど存在していないことを確認し、冠詞をつけた場合の表現が誤りであることに気づく。以上の手順により、被験者Dが解答を改善することができたことがわかる。

上記のように動画を閲覧することで各被験者の解答を分析すると、本システムの品詞検索を利用することで前置詞の誤りを改善できている問題や、Googleの通常の検索を利用することで検索結果の上位に提示された誤りを含む文を参考にしまい解答が悪化している問題があることがわかった。

また、2周目に本システムを利用して解答を行った被験者C、Dの動画を閲覧すると、本システムにおける通常の検索は、AND検索ではなくフレーズ検索であるにも関わらず、AND検索のつもりで検索している様子が伺えた。被験者Dの動画から、10問の英作文問題に解答する間の本システムの各検索機能を利用回数を調査したところ、検索回数の合計が45回であったのに対し、ワイルドカード検索は5回、品詞検索は3回、類義語検索は2回しか利用されていなかった。検索機能を利用することで容易に改善できる誤りが改善されていない問題が多いこともわかった。以上より、本システムの利用方法を誤っていること、検索機能を十分に利用できていないことがわかる。これは、本システムの利用方法についての説明が不十分であったためだと考えられる。図24より、2周目で本システムを利用した場合とGoogleを利用した場合で改善度合いに大きな差が見られないのは、このためだと考えられる。

このように、動画を閲覧することで、被験者がどのようなクエリを入力してどのような用例を参考にしているか、なぜ解答が書き換えられているのか、を読み取ることができるといえる。

しかし、動画を撮影することであらゆる情報を十分に読み取ることが可能だと考えていたが、解答時間や検索回数を動画から読み取るとは、非常に負担が大きいことがわかった。また、解答に対する自信を選択するのを忘れたまま次の問題に進んでしまい、後から前の問題に戻って自信の選択を行っている場合や、問題の順番通りに解答せずに、問題を飛ばしている場合があることがわかった。以上より、解答時間を自動で計測しながら、順番通りに問題の解答が行えるような解答シートが必要だと考えられる。検索回数は、クエリログを収集することで容易に集計できるため、動画の撮影と同時にクエリログの収集も必要だといえる。

6.2.4 考察

第二回予備実験では、穴埋め問題ではなく1文を英作文する問題に解答する方法を採用したが、想定よりも被験者・採点者の負担は小さいことがわかった。この方法で実験を行うことにより、実際の英作文の状況に近い状態で、どのような検索機能がどのように利用されているかを詳細に読み取ることができることがわかった。しかし、動画から解答速度を正確に読み取ることが難しいため、解答シートの改善が必要であることがわかった。

実験結果より、本システムを利用した場合と Google を利用した場合で改善度合いに大きな差は見られなかった。これは、被験者が本システムを日常的に利用しているわけではなく本システムに慣れていない状態であるのに、本システムの各機能の利用方法・利用場面について十分な説明が行えなかったためだと考えられる。

次章では、被験者に本システムの利用方法・利用場面について十分な説明を行った上で、英作文問題に解答する形式で行った評価実験について述べる。

7 評価実験

第一回予備実験、第二回予備実験の結果を踏まえて、評価実験を実施する。

7.1 実験設定

前章で述べた通り、第二回予備実験で重要表現を含む1文を英作文する問題を解く形式で評価可能であることが分かったため、問題形式と問題作成方法は6.2節と同様とする¹²。また、第二回予備実験の問題では、第一回予備実験の問題で用いたような文脈を表す英文を表示していなかったが、本実験では、文脈を表す英文を表示することとする。これは、第二回予備実験の際に、被験者から「問題文のみでは文脈が分からず、内容の解釈が難しい部分があった」というコメントをもらったためである。被験者には、問題文のみでは文脈が分からない場合に、必要に応じて文脈を表す英文を参考にして解答を行うように指示をする。

実験は、ExcelVBA を利用した図 25 の様な解答フォームを用いて行う。図に示したように、解答フォームには、上から順に、Start ボタン、問題番号、文脈を表す英文、問題文、解答用テキストエリア、解答に対する自信を選択するプルダウンリスト、1 周目の自分の解答、現在の時刻、開始時刻、Next ボタン、End ボタンを表示する。ただし、1 周目とは辞書を利用して問題に解答する手順であり、解答フォームの利用について述べた後で詳しく述べる。被験者は、事前に実験で用いるもの（辞書や本システム、Google）を準備した上で、以下の手順で解答フォームを利用する。

1. Start ボタンを押し、表示された問1に解答する。解答は、解答用テキストエリアに入力する。解答の入力後、解答に対する自信をプルダウンリストから選択する。現在の時刻と開始時刻を差し引きすることで解答時間を確認しながら解答を行い、全ての入力を終えたら Next ボタンを押す¹³。
2. Next ボタンを押すと、問2が表示されるので、問2以降も上記と同様に解答を繰り返す。

¹²問題に用いる英文も、前章と同様、ACL Anthology 上の 2012 年の論文データから抽出する。

¹³1 問の解答時間は 3 分を目安にする。

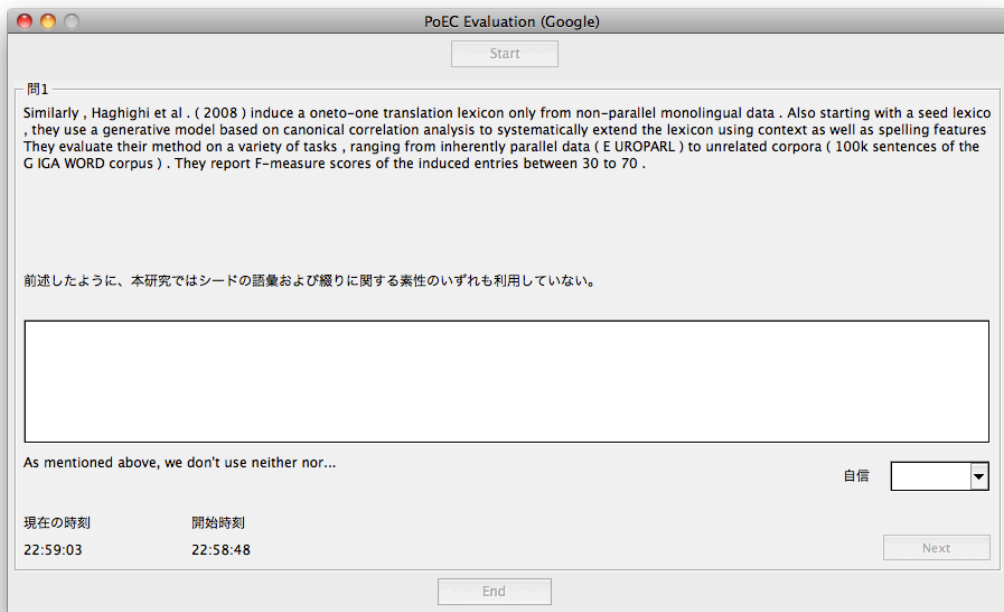


図 25: ExcelVBA を利用した解答フォーム

3. 最終問題の解答を終えると、問題の表示が消えるので、End ボタンを押して解答フォームを閉じる。

解答に対する自信は、プルダウンリストに登録されている◎、○、△、×のいずれかを選択することで表す。最も自信が高い場合は◎、まったく自信がない場合は×を選択することになる。この解答フォームを利用すると、Next ボタンを押したタイミングで解答と解答時間が自動で記録されるため、第二回予備実験で生じた解答時間を読み取りづらい問題は解決される。また、前の問題に戻ることはできない仕様になっている。

解答の際は、Start ボタンを押す前に QuickTime を起動し、画面をキャプチャする動画を撮影する。解答が終了したら動画撮影を終了する。このように、画面をキャプチャする動画を撮影することで、被験者が解答の際にどのようなクエリを入力して検索したか、どのような用例を参考にしたか、などの情報を記録することができる。また、本システムを利用して解答を行っている場合の各機能の使

用頻度の集計を容易にするため、クエリログも収集することとする。実験の記録は、ここで撮影した動画と、本システムのクエリログの2つとする。

本実験で用いる問題数は全40問とし、これを20問1セットとして、前半セット、後半セットに分けて実験を行う。

被験者は、英語論文の執筆経験・読解経験がある学生3名、英語論文の読解経験がある学生7名を合わせた10名とする。この10名を、5名ずつ、Aグループ、Bグループに分類して実験を行う。Aグループは執筆経験者1名、Bグループは執筆経験者2名を含む。ユーザビリティ評価に関する書籍[26]によると、5人のユーザーで実験を行うことで、システムの問題のうち約85%の問題を発見できる。そのため、本実験では、被験者を5人以上としている。

実験手順は以下の通りである。

1. 前半セットの20問を、Mac OSXに標準搭載されている辞書を用いて解答する。ここで用いる辞書は、プログレッシブ英和・和英中辞典に限定する。
2. 後半セットの20問を、上記と同様の辞書を用いて解答する。
3. Aグループは、前半セットの20問を、辞書と本システムを利用して解き直す。Bグループは、前半セットの20問を、辞書とGoogleを利用して解き直す。
4. Aグループは、後半セットの20問を、辞書とGoogleを利用して解き直す。Bグループは、後半セットの20問を、辞書と本システムを利用して解き直す。

以降、上記の手順1と手順2を合わせて1周目と呼び、手順3と手順4を合わせて2周目と呼ぶこととする。以上の実験手順をまとめると、図26のようになる。2周目の解答を行っている際は、図25の解答フォームにおける解答用テキストエリアの下に、1周目の自分の解答が表示されるため、1周目の解答を参考にしながら2周目の解答を行うことができる。

このような手順で実験を行うことで、全ての被験者が本システムとGoogleの両方を利用する機会ができ、被験者ごとの各システムに対する適性の差が軽減される。また、被験者は異なるが、全ての問題に対して2周目を本システムを利用

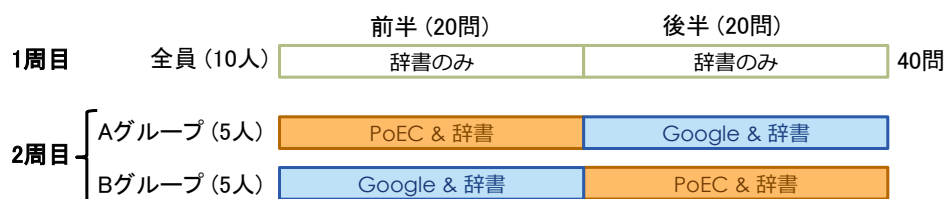


図 26: 評価実験の手順

して解き直した解答と Google を利用して解き直した解答を得られるため、1つの問題に対して、本システムを用いた場合はこのような検索方法で改善できるが Google を用いた場合は改善が難しい、といったシステム間の比較を行うことができると考えられる。実験の際には、上記の各手順の間に、約 10 分の休憩を挟む。

実験中用いる Google は、期間指定を行い、検索結果は 2011 年以前の情報に限定する。これは、問題に用いる英文を 2012 年の論文データから抽出していることで、元の論文が検索結果に表示されることを防ぐ必要があるためである。また、実験中 Google を用いて用例検索を行う場合には、日本語のクエリは入力しないこととする。これは、日本語のクエリで検索を行うことは、用例検索としてではなく通常の検索エンジンとしての利用方法であり、辞書的な利用方法だと考えられるためである。実験中に利用可能な機能は、Google 英文ライティング [6] に基づき、フレーズ検索、ワイルドカード検索とする。さらに、参考にする用例を発見しやすくするため、ファイルタイプ指定を許可する。ファイルタイプ指定は、クエリに filetype:pdf と付け加えることで、検索結果を pdf ファイルに限定することができる機能である。この機能を利用することで、動画サイトや書籍販売サイトなどのページが検索対象から除かれ、研究に関する文書や論文から検索することが容易になり、参考にできる用例が探しやすくなる。

なお、問題の解答をはじめめる前に、以下の手順で各機能について説明を行い、各機能の使用方法や使用場面を被験者に理解させた上で実験を行う。

本システム 通常の検索がフレーズ検索になっていることに加え、ワイルドカード検索、品詞検索、類義語検索について使用方法を説明し、実際にそれらを用いて練習問題を解く。検索時のオプションや、テーブル内フィルタリン

グ機能については、実際に本システムを提示しながら、口頭で説明を行う。

Google フレーズ検索、ワイルドカード検索、ファイルタイプ指定の方法について、実際に Google でそれらの機能を利用した場合の動きを提示しながら、口頭で説明を行う。

7.2 評価方法

まず、英語を母国語とする 2 名の採点者に、被験者の解答内の誤り箇所が発見・修正を依頼する。これは、文法的に合っているか、意味的に合っているか、論文の文章として適切であるか、という観点で行う。この 2 名の採点者の修正結果を統合し、評価を行う。ただし、食い違う部分は採点者と話し合うことで統合を行う。1 周目の解答の修正結果と 2 周目の解答の修正結果を比較することで、本システムや Google によって解答を修正できた箇所や修正できなかった箇所が発見できる。

評価は、問題内のチェックポイントに対して行う。問題内で、表現集 [29] に記載されていた重要表現部分に加え、重要表現ではないが重要だと考えられる部分を、チェックポイントとする。チェックポイントは、前半セットで 25 箇所、後半セットで 25 箇所あり、合計 50 箇所である。

それぞれのチェックポイントについて、1 周目の解答の修正結果と 2 周目の解答の修正結果を比較し、以下のように分類する。

改善 本システムや Google の影響により、1 周目の解答より 2 周目の解答が改善されているもの。

悪化 本システムや Google の影響により、1 周目の解答より 2 周目の解答が悪化しているもの。

変化なし 1 周目の解答と 2 周目の解答を比較して、解答の内容は変化しているが、改善も悪化もしていないもの。また、解答の内容が変化していないもの。本システムや Google の影響ではなく、辞書やその他の理由によって改善または悪化したものも含む。

この分類を行う際に、用例検索で対応する範囲ではないと考えられることから、スペルミスについては考慮しないこととする。例えば、1周目でスペルミスがあり、2周目でスペルミスが改善されたチェックポイントがあった場合は、上記の変化なしに分類し、解答が改善されたとはみなさない。

上記のように分類したそれぞれの問題について、本システムのクエリログや、画面をキャプチャした動画を閲覧することで、どのような検索クエリを入力し、どのような検索機能を利用することで1周目の解答から2周目の解答へ至ったのかを調査する。また、どの機能がよく使われていたのかを知るために、それぞれの機能が使われた回数を集計する。

7.3 実験結果

Aグループの被験者5人の解答について、前節で述べたチェックポイントごとの分類を行った結果を、表9、10に示す。同様に、Bグループの被験者5人の結果を、表11、12に示す。表9、10、11、12において、改善(システム非依存)は辞書やその他の理由によって改善したチェックポイント、悪化(システム非依存)は辞書やその他の理由によって悪化したチェックポイントである。

表9: [Aグループ/前半セット(本システム)] チェックポイントの分類

分類	被験者1	被験者2	被験者3	被験者4	被験者5	合計
改善	3	2	1	4	4	15
悪化	1	0	1	1	1	4
変化なし	21	23	23	20	20	106
改善(システム非依存)	2	0	2	0	1	5
悪化(システム非依存)	0	1	0	0	0	1

表9、10、11、12より、本システムやGoogleを利用することで解答が悪化することはほとんどないことが分かる。また、実験を撮影した動画を閲覧すると、本システムを利用して検索を行い、複数の用例を観察しているが解答には反映させない場合があり、被験者が用例の内容を理解して、取捨選択することができて

表 10: [A グループ/後半セット (Google)] チェックポイントの分類

分類	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	合計
改善	1	3	2	0	1	7
悪化	0	0	0	0	0	0
変化なし	24	22	23	25	24	118
改善 (システム非依存)	1	3	0	0	0	4
悪化 (システム非依存)	0	0	3	0	1	4

表 11: [B グループ/前半セット (Google)] チェックポイントの分類

分類	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	合計
改善	0	0	1	2	0	3
悪化	0	2	0	0	0	2
変化なし	25	23	24	23	25	120
改善 (システム非依存)	4	0	2	0	6	12
悪化 (システム非依存)	0	0	0	0	0	0

表 12: [B グループ/後半セット (本システム)] チェックポイントの分類

分類	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	合計
改善	2	2	2	1	2	9
悪化	0	0	0	0	0	0
変化なし	23	23	23	24	23	116
改善 (システム非依存)	0	0	0	0	0	0
悪化 (システム非依存)	1	0	0	0	0	1

表 13: [A グループ/前半セット] 1 周目の解答の正解率

	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	合計
正解数	12	8	9	9	13	51
不正解数	13	17	16	16	12	74
正解率	48.0%	32.0%	36.0%	36.0%	52.0%	40.8%

いると考えられる。これらの結果から、被験者は、検索結果の用例の中から解答を改善させるための用例を選択することができていると考えられる。

また、解答に誤りを含まないものを正解、誤りを含むものを不正解とした場合の、各被験者の 1 周目の解答における正解率を表 13、14、15、16 に示す。

表 14: [A グループ/後半セット] 1 周目の解答の正解率

	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	合計
正解数	16	1	13	10	13	53
不正解数	9	24	12	15	12	72
正解率	64.0%	4.0%	52.0%	40.0%	52.0%	42.4%

表 13、14、15、16 より、1 周目の時点で最低でも約 4 割の問題が正解となっていることが分かる。1 周目の解答が正解である場合、2 周目で更に改善することは不可能であるので、1 周目の解答が正解であるチェックポイントは改善の余地がないと判断する。つまり、改善の余地があるチェックポイントのうち実際に改善できた数が、表 9、10、11、12 における改善の分類となっている。

表 9、10、11、12 に示した変化なしの数から、表 13、14、15、16 に示した正解数、つまり改善の余地がないチェックポイントの数をひいたもので正解数を割ることで、本システムや Google によって解答を改善できた割合を求めることができる。この方法により各システムの改善率を求めたところ、本システムの改善率

表 15: [B グループ/前半セット] 1 周目の解答の正解率

	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	合計
正解数	16	17	10	13	10	66
不正解数	9	8	15	12	15	59
正解率	64.0%	68.0%	40.0%	52.0%	40.0%	52.8%

表 16: [B グループ/後半セット] 1 周目の解答の正解率

	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	合計
正解数	20	20	12	17	20	89
不正解数	5	5	13	8	5	36
正解率	80.0%	80.0%	48.0%	68.0%	80.0%	71.2%

は約 21.1%、Google の改善率は約 7.2% となった。Google を利用した場合に比べて、本システムを利用することでより多くのチェックポイントを改善することができたと言える。この結果から、通常の検索エンジンに比べて用例検索システムが有用であることが示唆される。

さらに、本システムや Google の影響により改善したチェックポイントに関して、以下の手順で詳細な分析を行う。

1. 改善した各チェックポイントを、第 3 章で示した、(1) 表現の適切性を確認したい場合、(2) 表現の組み合わせが分からない場合、(3) 表現の用法が分からない場合、の 3 つに分類する。ただし、1 つのチェックポイントが (1) と (2) の 2 つに該当する場合のように、重複することがある。
2. 分類したそれぞれのチェックポイントについて、本システムや Google のどの検索機能によって改善しているかを分析する。

本システムのクエリログや動画を閲覧することで各被験者がどのような狙いで検索を行ったかを読み取ることができるため、手順 1. において、チェックポイントの分類は、クエリログと動画を閲覧することにより行う。例えば、A グループの被験者 4 は、前半セットの問 9 の「以下の通り」という内容のチェックポイントに対して、1 周目では “The following, ” という書き出しで解答を行っているが、2 周目では “ in the following.” という解答に変化している。この部分について、本システムのクエリログを閲覧すると、“The following, ”、“The following”、“[i] the following” という順に検索を行っていることが分かる¹⁴。また、この部分の動画を閲覧すると、“The following,” というクエリで検索を行った際には検索結果が 0 件であったため、“The following” というクエリに修正して検索を行って

¹⁴[i] は、前置詞の検索を行うための記号である。

いることが分かる。この時、被験者4は提示された多くの用例を閲覧し、文頭に“The following”という表現がある用例は少なく、ほとんどの用例は文末に“the following.”という表現があることに気づく。ここで、“[i] the following”というクエリで品詞検索を行うことで、用例を更に絞り込んで、どのような前置詞が利用されているかを調査しようとしている。最終的には、品詞検索の結果を閲覧することで、“in the following.”という正しい表現に改善することができている。以上より、被験者4は、“The following,”、“The following”というクエリを入力した段階では1周目の解答の適切性を確認しようとしており、“[i] the following”というクエリを入力した段階では“the following”という表現と組み合わせる前置詞を調査しようとしていると判断できるため、被験者4により改善されたこのチェックポイントは、(1)、(2)の2つに該当するとみなす。被験者4の解答から、(1)の場合については本システムの通常の検索で対応することができ、(2)の場合については本システムの品詞検索で対応することができていると分かる。

以上のような手順で各チェックポイントを分析することで、第4章で、(1)表現の適切性を確認したい場合、(2)表現の組み合わせが分からない場合、(3)表現の用法が分からない場合、の3つに対して設計した検索機能が実際にどのように役立っているかを確認することができる。

以上の手順により各チェックポイントの改善理由を分析した結果を、2周目でGoogleを利用したチェックポイントについては表17、2周目で本システムを利用したチェックポイントについては表18に示す。

表 17: [Google] チェックポイントの改善理由

検索機能	(1) 適切性	(2) 組み合わせ	(3) 用法	合計
通常検索	9	1	1	11
ワイルドカード検索	0	0	0	0
合計	9	1	1	11

表17に示したように、Googleによって解答を改善する場合は、通常の検索が有用である。一方、本システムによって解答を改善する場合は、表18に示したように、(1)の場合については通常の検索が、(2)の場合については品詞検索が、(3)の場合については通常の検索やワイルドカード検索が有用である。

表 18: [本システム] チェックポイントの改善理由

検索機能	(1) 適切性	(2) 組み合わせ	(3) 用法	合計
通常検索	12	2	3	17
ワイルドカード検索	1	0	2	3
品詞検索	0	9	0	9
類義語検索	0	1	1	2
合計	13	12	6	31

また、表 17、18 を比較すると、Google によって改善できたのはほとんどが (1) 表現の適切性を確認したい場合であるのに対して、本システムでは (1)、(2)、(3) の全ての場合について改善できていることが分かる。これらの結果より、(2) の問題に対応するために設計した品詞検索や (3) の問題に対応するために設計したワイルドカード検索が実際に役立っていることが確認できる。

特に、(2) の場合について品詞検索で対応することができたのは、表 18 より 9 件であることがわかるが、このうち 6 件は前置詞の改善であった。以上より、ある表現に組み合わせる前置詞が分からない場合には、本システムの品詞検索が有用であると言える。

また、Google の通常検索によって改善できたチェックポイントの中には、検索結果だけではなく、クエリ入力時のクエリ候補を参考にしてしているものがあった。例えば、“compare” というクエリを入力すると、“compared with” や “compared to” というクエリ候補が表示される。このクエリ候補が表示された場合は、必ずしも正しい表現であるとは限らないが、何度も検索を行うことなく、頻繁に入力されているクエリを知ることができる。

ここで、本システムや Google の影響により悪化したチェックポイントについて分析する。表 9、10、11、12 に示した通り、本システムによって悪化したチェックポイントは 3 箇所、Google によって悪化したチェックポイントは 2 箇所ある。本システムによって悪化した 3 箇所のチェックポイントでは、以下のような現象が起きていた。

- 書こうとしている表現とは意味や文型が異なっている用例を参考にしてしまう。

- 検索結果としてわずかな量の用例しか提示されず、その用例が誤った表現を含んでいるにも関わらず、参考にしてしまう。

1点目については、動画を閲覧することで、本システムは入力されたクエリに対して合致している用例を提示できているが、被験者が用例を十分に観察していないため起こったことが分かる。これは、解答時間に目安を設定したために時間をかけて用例を閲覧することができなかったことや、被験者が本システムを利用して英作文をすること自体に慣れていなかったことが原因だと考えられる。2点目については、以下のような方法で対応する必要があることが分かる。

- 検索結果が少ない場合にはクエリを自動で抽象化・バックオフすることで、検索結果を増加させる。
- 検索対象として利用する英文を、英語を母国語とする著者の論文のものに限定することで、なるべく誤りを含む用例を減らす。

また、Googleによって悪化したチェックポイントでは、以下のような現象が起きていた。

- 検索結果として誤りを含む文が提示され、その文を参考にしてしまう。

Googleでは、検索対象のデータ量が膨大であるため、あるクエリに対して多くの検索結果が提示されても、検索結果の文が誤りを含んでいることがある。そのため、上記のような現象が発生してしまうと考えられる。

以上より、検索結果を増加させることが重要である一方で、検索結果に多くの誤りを含む場合は、多くの検索結果を提示したことによりユーザーが誤りを参考にしてしまう可能性があると考えられる。そのため、検索結果を増加させると同時に、誤っている用例を除去することが重要だといえる。

ここまでは本システムやGoogleによってどのように解答が改善または悪化したかを分析してきたが、以下では、本システムの各検索機能の利用状況についてさらに分析を行うために、本システムの検索回数についてまとめる。ただし、検索回数はクエリログを集計することで算出するため、チェックポイント以外の部分における検索も含んでいる。

Aグループの各被験者が2周目で本システムを利用した際の検索回数を表19に示す。同様に、Bグループの検索回数を表20に示す。なお、“[i] * result”のように、複数の検索機能を同時に利用するクエリが存在するため、表19、20におけるワイルドカード検索、品詞検索、類義語検索の回数を足しあわせた数値と、全体の検索回数の数値は異なる。

表 19: [Aグループ/前半セット] 本システムを利用した際の検索回数

機能	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	合計
ワイルドカード検索	31	47	23	26	58	185
品詞検索	79	35	15	9	17	155
類義語検索	15	19	23	32	58	147
検索回数	195	127	147	106	211	786

表 20: [Bグループ/後半セット] 本システムを利用した際の検索回数

機能	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	合計
ワイルドカード検索	17	16	51	14	10	108
品詞検索	29	12	3	7	26	77
類義語検索	22	9	11	0	30	72
検索回数	131	100	107	78	82	498

表19、20より、被験者ごとのばらつきは大きいですが、最も使用頻度の高い検索機能はワイルドカード検索であることが分かる。品詞検索や類義語検索も、20問の解答を行う間に平均で約14回以上利用されている。

また、各被験者の検索回数に着目すると、Aグループでは20問を解く間に平均で約157回、1問あたり約8回、Bグループでは20問を解く間に平均で約100回、1問あたり約5回の検索が行われていることが分かる。

以上より、英作文を行う際に用例検索を利用して表現を調査する場面は非常に多いと考えられる。

しかし、上述の通り、チェックポイントに関する分析では、本実験における本システムの改善率は約21.1%にとどまっているため、本システムを利用して検索を行ったが解答を改善できていない場面が多いことが示唆される。

最後に、各被験者の2周目の平均解答速度と自信の上昇度を表21、22に示す。自信の上昇度は、自信が◎の場合を3、○の場合を2、△の場合を1、×の場合を0とし、2周目の解答の自信の数値から1周目の解答の自信の数値を引くことで計算する。例えば、1周目の解答の自信が○、2周目の解答の自信が◎だった場合、自信の上昇度は、3-2の計算結果である1となる。表21、22で示した自信の上昇度は、20問の解答における自信の上昇度の合計値である。

表 21: [A グループ] 2周目の平均解答速度と自信の上昇度

	被験者 1	被験者 2	被験者 3	被験者 4	被験者 5	平均
平均解答速度 (本システム)	06:16	02:50	03:56	03:42	03:43	04:05
平均解答速度 (Google)	04:08	02:22	03:04	01:55	02:43	02:50
自信上昇度 (本システム)	19	21	10	10	-5	11.0
自信上昇度 (Google)	17	3	8	11	-3	7.2

表 22: [B グループ] 2周目の平均解答速度と自信の上昇度

	被験者 6	被験者 7	被験者 8	被験者 9	被験者 10	平均
平均解答速度 (本システム)	02:33	01:34	02:15	01:26	03:05	02:11
平均解答速度 (Google)	03:12	01:26	02:11	01:38	02:16	02:09
自信上昇度 (本システム)	29	-1	14	18	5	13.0
自信上昇度 (Google)	23	15	17	18	8	16.2

表21、22と表19、20の結果を比較すると、解答に時間がかかっている被験者は、検索回数が多いことが分かる。また、表9、10、11、12の結果と比較すると、長時間解答を行っていても改善数が多いとは限らないことが分かる。

自信上昇度に注目すると、本システムを利用した場合に自信が大きく上昇している被験者が多いことから、本システムの有用性が示唆される。

実験終了後に、各被験者から以下のようなコメントを得ることができた。

本システムに関して

- いくつか単語を思い出すことができて、構文がわからない場合に、用例検索を利用したくなる

- 用例検索システムがあると、一度に多くの用例を閲覧することができ、調べたい表現の情報が得やすい
- 文頭を検索する機能があるとより用例を探しやすくなる
- クエリ入力時にスペルミスをするだけでHIT件数が0件になってしまうので、スペルチェッカーを内蔵するか、柔軟な検索を行ってほしい
- HIT件数が極端に少ない場合や、0件だった場合に、自信が下がる
- HIT件数が少ないことで、誤りに気づくことができた
- 検索対象が自然言語処理分野の論文であるため、分野が絞られており、用例を信頼することができる
- 検索時のオプションのチェックボックスを一度チェックして検索すると、その後の検索でチェックを外すのを忘れることがあった
- ワイルドカード検索は、Googleとは異なり、対応する語数の最大値を指定できるので、大雑把な検索をしても参考にできる用例を探せることが多い
- 品詞検索を行うと、品詞部分に対応する単語ごとに用例が仕分けされるので、どの単語を用いる場合が多いのか、どのような内容を書く際にどの単語を使うのか、ということが読み取りやすい
- 類義語検索によって、自分で思いつくことができた単語を起点に実際に頻繁に使われる単語を発見することができた

Google に関して

- 検索結果にノイズ（動画や書籍の情報など）が多く、一度に10件のページしか見られないため、一覧性が低く、用例を探すのが容易でない
- クエリ内に誤りがある場合や、より適切なクエリがあると判断された場合に表示される、「もしかして」の検索機能が役立った
- 「もしかして」が表示されない場合、自分で思いつくことができた表現があるかどうかしか調べられない（別の表現への誘導が行われない）

- 検索対象のデータ数が非常に多いため、どのようなクエリを入力しても何かしらの検索結果が表示されるので、表そうとしている内容が検索結果に表示されても、正しいかどうかを判断できない
- 検索性が極端に少ない場合や、0件である場合はほとんどないため、自信は上がる

実験全体に関して

- 本システムにしても Google にしても、検索することで参考にできる情報が得られるため、辞書のみで解答する場合に比べて安心感がある
- 後半セットの問題は平易に感じた

ただし、複数の被験者から得られたコメントを羅列しており、被験者によって1つの現象に対する感じ方が異なるため、対立するコメントが存在し得る。上記のように、本システムでは一度に多くの用例を参考にすることができる一方、Googleでは一覧性が低く多くの用例を参考にすることは難しいため、単純に一覧性だけで比較しても、被験者は本システムと Google の間に大きな差を感じていることがわかる。また、本システムに関しては、各検索機能それぞれが別の場面で役立つと感じられていることがわかる。特に、本システムの品詞検索が有用だと感じたとコメントした被験者が多かったが、表 18 の結果から、実際に品詞検索によって改善された解答が多いことが確認できる。

7.4 考察

2回の予備実験の結果を踏まえて、画面のキャプチャ動画を撮影しクエリログを収集しながら1文英作文の問題を解く形式の実験を実施することで、実際に英作文を行う際にどのような機能が利用されどのように用例を参考にするのかを調査することができた。特に、画面を撮影した動画を閲覧することで、各機能の利用回数や、英作文問題の改善率といった数値情報だけではなく、品詞検索を利用して前置詞の誤りを直すことができている、といったリアルな情報を得ることが可能であることがわかった。改善されたチェックポイントについて、実際に動画

を閲覧すると、品詞検索で前置詞や動詞の誤りを改善していたり、ワイルドカード検索で正しい構文を発見していたり、類義語検索でより相応しい動詞を発見していたりと、各機能がそれぞれ別の目的で利用されていることが分かる。また、それらの機能を使わずに通常の用例検索を行った場合でも、冠詞や動詞など様々な部分を改善することができていることが分かる。検索回数も想定以上であり、英作文を行う際に用例検索システムの有用性や必要性が高いことが示唆される。

また、実験を行う前に本システムの各機能の使用方法について練習問題を設け、十分な説明を行うことができたため、第二回予備実験と比較して、本システムの各機能の利用回数が増加していると考えられる。被験者が用例検索システムを使い慣れていない場合は、実験を実施する前に、各機能の利用方法や利用場面を十分に説明することが重要である。

本評価実験により、英作文における問題に対応するために設計した検索機能を利用することで、実際に誤りを含む表現を改善することができ、各機能が有用であることが確認できた。さらに、改善されたチェックポイントや悪化したチェックポイントについて詳細に分析を行ったところ、検索結果を増加させると同時に、誤りを含む用例を減少させることが必要だと分かった。検索結果を増加させるには、クエリ内の単語を自動で品詞記号やワイルドカード記号に抽象化する方法や、クエリ内の単語を減らすバックオフを行う方法が考えられる。また、誤りを含む用例を減少させるには、用例として利用する英文を英語を母国語とする著者の論文の英文に限定する方法が考えられる。

8 おわりに

本研究では、用例検索が英作文支援に有用であるかどうかを評価できる十分な環境がないという現状を踏まえ、まず、英作文に関する文献に基づいて、英作文の際の問題を用例検索の観点から分類し、用例検索システムによる支援の可能性を調査した。次に、その結果に基づき、英作文支援に必要な機能を検討し、用例検索システムを設計・実装した。本システムの各検索機能の有用性を確認するために評価実験を実施したところ、品詞検索やワイルドカード検索によって、通常の検索エンジンでは対応できない問題を改善することが可能だと分かった。また、本システムの改善の方向性として、クエリの自動抽象化やバックオフ、検索対象の見直しが考えられた。

今後は、実験結果に基づいてシステムの実装を改善するとともに、英作文問題の難易度や被験者の英語能力を考慮したさらに大規模な評価実験を行う必要があると考えられる。

謝辞

本研究を進めるにあたり、終始適切なご助言をくださり、暖かく見守ってくださった乾健太郎教授に深く感謝いたします。また、プログラミングの面から研究内容まで様々な相談に乗っていただき助けてくださった岡崎直観准教授に心より感謝いたします。水野淳太氏には、研究テーマを決めるところから、システムの実装や論文のストーリーなど、様々なところでその都度丁寧にご指導頂き、本当にお世話になりました。ありがとうございます。Kruengkrai Canasai 研究員、Paul Reisert 氏には、試験問題の作成や実験結果の英文の採点の際に協力していただき、非常に助かりました。感謝いたします。

また、お忙しい中、審査委員をお引受けくださいました、篠原歩教授、徳山豪教授に深く感謝いたします。

渡邊陽太郎助教、研究員の福原裕一氏、秘書の八巻智子さんをはじめ、研究室の皆様方には、日ごろから深い議論をしていただいたり、貴重なご意見をいただくなど、多くの面で助けられました。深く感謝いたします。

最後に、本システムを公開してから、興味を持っていただき実際に使用していただいた皆様、特にシステムに関するご意見をくださった皆様に、心から感謝いたします。

参考文献

- [1] 斉藤俊雄, 赤野一郎, 中村純作 (編) . 英語コーパス言語学—基礎と実践—. 研究社出版, 1998.
- [2] 西垣知佳子, 峰岸識子, 中條清美. 中学・高校の英語教育におけるデータ駆動型学習に基づく帰納的学習の実践的研究. 千葉大学教育学部研究紀要, Vol. 60, pp. 419–426, 2012.
- [3] 西垣知佳子, 天野孝太郎, 吉森智大, 中條清美. 中・高生のためのコンコーダンス・ラインを利用したデータ駆動型英語学習教材の開発の試み. 千葉大学教育学部研究紀要, 第 59 巻, pp. 235–240, 2011.
- [4] 能登原祥之. 収斂型コンコーダンスの教育効果—日本人英語学習者 (大学生初級・中級) の場合—. 広島大学大学院教育学研究科紀要, 第 58 巻, pp. 165–174, 2009.
- [5] 中條清美, 西垣知佳子, 内堀朝子, キャサリン・オヒガン. データ駆動型学習による効果的な英語初心者向け文法指導の試み. 日本大学生産工学部研究報告 B, 第 41 巻, pp. 15–33, 2008.
- [6] 遠田和子. Google 英文ライティング. 講談社インターナショナル株式会社, 2009.
- [7] 杉本淳一. Google 英語勉強法. 日本実業出版社, 2011.
- [8] 衣笠忠司. Google 検索による英語語法学習・研究法. 開拓社, 2010.
- [9] David Wible and Nai-Lung Tsao. Stringnet as a computational resource for discovering and investigating linguistic constructions. In *The NAACL HLT Workshop on Extracting and Using Constructions in Computational Linguistics*, pp. 25–31, 2010.

- [10] 松原茂樹, 江川誠二, 加藤芳秀. 英文用例検索システム ESCORT:論文データベースを用いた図書館サービス. 第4回情報プロフェッショナルシンポジウム, pp. 125–129, 2007.
- [11] 大名力. 学習者用オンライン英文検索システムの構築. 群馬大学社会情報学部研究論集, 第7巻, pp. 125–142, 2000.
- [12] 松原茂樹, 酒井佑太, 小澤俊介, 杉木健二. 学術論文からの英語表現集の自動生成. 第7回情報プロフェッショナルシンポジウム, pp. 41–45, 2010.
- [13] 竹内豪, 竹内まち子, 佐野洋. Ajax 技術を用いた語学教育用例検索インタフェースの開発. 情報処理学会研究報告. コンピュータと教育研究会報告, Vol. 2008, No. 13, pp. 147–154, 2008.
- [14] 綱島祐一, 川崎優太, 安藤一秋. 検索エンジンを用いた英作文支援ツール. 信学技法, 第106巻, pp. 87–92, 2007.
- [15] 三好康夫, 越智洋司, 金西計英, 岡本竜, 矢野米雄. 英作文支援における句構造情報を利用した用例検索ツール. 日本教育工学雑誌, Vol. 27, No. 3, pp. 283–294, 2003.
- [16] 坂本廣, 北村泰彦, 福島拓, 吉野孝. N-gram に基づく多言語用例検索手法の評価 (言語グリッドと異文化コラボレーション). 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, Vol. 110, No. 428, pp. 51–56, 2011.
- [17] 難波英嗣, 森下智史, 相沢輝昭. 論文データベースからのイディオム用例検索. 情処学研報, NL-170, pp. 53–59, 2005.
- [18] Martin Chodorow, Michael Gamon, and Joel Tetreault. The utility of article and preposition error correction systems for English language learners: Feedback and assessment. In *Language Testing 2010*, 2010.
- [19] 大鹿広憲, 佐藤学, 安藤進, 山名早人. Google を活用した英作文支援システムの構築. *DEWS2005, 4B-i8*, 2005.

- [20] 中村哲三. 英文テクニカルライティング 70 の鉄則. 日経 BP 社, 2011.
- [21] Ann M. Korner. 一流の科学者が書く英語論文. 東京電機大学出版局, 2010.
- [22] 原田豊太郎. 理系のための英語論文執筆ガイド. 講談社, 2002.
- [23] 阿部一. 9割の日本人が使い方を間違える英単語 101. ジャパンタイムズ, 2011.
- [24] James H. M. Webb. 日本人に共通する英語のミス 151. ジャパンタイムズ, 2006.
- [25] Steven Bird, Robert Dale, Bonnie Dorr, Bryan Gibson, Mark Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir Radev, and Yee Fan Tan. The ACL Anthology Reference Corpus: A Reference Dataset for Bibliographic Research in Computational Linguistics. In *Language Resources and Evaluation Conference (LREC 08)*, 2008.
- [26] 樽本徹也. ユーザビリティエンジニアリング—ユーザー調査とユーザビリティ評価実践テクニク—. オーム社, 2005.
- [27] Steve Krug. ウェブユーザビリティの法則 ストレスを感じさせないナビゲーション作法とは. ソフトバンク パブリッシング株式会社, 2001.
- [28] CC Huang, MH Chen, ST Huang, HC Liou, and J S.Chang. Grasp: Grammar- and syntax-based pattern-finder in call. In *the Sixth Workshop on Innovative Use of NLP for Building Education Applications*, pp. 96–104, 2011.
- [29] 安原和也. 英語論文基礎表現 717. 株式会社三修社, 2011.

発表文献一覧

国内会議・研究会論文

- 高松優, 水野淳太, 岡崎直観, 乾健太郎. 英作文支援のための用例検索システムの構築. 言語処理学会第18回年次大会論文集, March 2012.