様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成　２７　年　　６　月　　８　日現在

機関番号：　１７１０２
研究種目：　挑戦的萌芽研究
研究期間：　2012～2014
課題番号：　２４６５２１２２

研究課題名（和文）言語処理技術を用いた任意の英文書の内容に関する問題と解答の自動生成

研究課題名（英文）Automatic Generation of Comprehension Tests for Arbitrary English Texts

研究代表者
冨浦　洋一（Tomiura, Yoichi）

九州大学・システム情報科学研究科（研究院・教授

研究者番号：１０２１７５２３

交付決定額（研究期間全体）：（直接経費）　　2,700,000円

研究成果の概要（和文）：英文書を，日本語に訳すことなく内容を整理しながら読むトレーニングとして，読んだ後に内容に関する質問に答える多読トレーニングが効果的と考えられる．トレーニングで読む文書は，学習者のレベルに合った興味ある内容のものが良い．そこで，このようなトレーニングを支援することを目的として，学習者が選んだ任意の英文書に対して，選択肢の中から読んだ文書の内容に合う文を１つ選択する形式の問題と解答を自動生成する手法を研究した．中心技術は大規模な学習データを必要としない重要文の抽出手法と言い換え手法であり，これらの開発を行った．研究期間内に実用レベルには達しなかったが問題の自動生成の目処は立ったと考えている．

研究成果の概要（英文）：Reading extensive English documents with comprehension test after reading is a good training for Japanese learners to comprehend English text without translating it into Japanese. They should read texts that are interesting and with an appropriate level for them. We have studied the method to generate automatically Multiple-Choice Test to support such training. The main techniques are the method for extracting the important sentences in the text and the method for paraphrasing them, which do not need a huge training data. We have developed these two methods. Although their performances are not practical level, we have got prospects of development.

研究分野：　自然言語処理

キーワード：　テストの自動生成　パラフレーズ　重要文　多読支援

１．研究開始当初の背景

英文書の多読は，英語を英語の語順で理解するためのトレーニングの一つである．しかし，読んでいるそのときは１文１文の意味は理解できているようでも，読み終えた後に読んだ文書の内容について問われると答えられないということが初学者には良くある．実践的な英語能力を身に付けるためには，単に読み流すのではなく，内容を理解し整理しながら読むトレーニングが必要である．これには，読んだ文書の内容に関する問題とその解答が付与された教材を利用するのが効果的と考えられる．一方，多読の対象の文書としては，多くの研究者・教育者が指摘するように，学習者にとって興味ある内容で，かつ，学習者のレベルに合ったものを教材とするのが良い．Web 上には，様々な内容，様々な難易度の英文書があり，そのような教材を作成するための素材の宝庫である．しかし，当然のことながら，Web 上の文書で，その内容に関する質問が付いているものは非常にまれである．そこで，自然言語処理技術を活用し，学習者が選んだ任意の文書に対して，その内容に関する問題と解答を自動生成し，多読学習の教材とすることが期待される．

これまでにも，学習者が英文書を読んだ後に取り組む問題の自動生成に関する研究が行われている（國近他：「英語長文読解学習のための質問の複雑さの定義とその評価」，人工知能学会論文誌，Vol．17，pp.521-529 (2002 年)）(國近他：「英語物語に関する質問応答のための意味比較による正誤判定」，電子情報通信学会論文誌 D-Ⅰ,Vol．J88-D-Ⅰ，pp.25-35 (2005 年))．これらは，中学生等の初期の英語学習者を対象として，文法や語彙の知識の定着を目的とした学習支援システムである．一方，英語をある程度学んだ者が実践的な能力を身につけるために行う多読トレーニングは長期に渡る．これを教師が常に指導するということは不可能であるため，このようなトレーニングこそ計算機システムによる支援がより重要であると考えている．

２．研究の目的

本課題では，自動要約，言い換え，質問応答，照応解析などの言語処理技術を利用して，任意の英文書に対する内容に関する問題と解答を自動生成する手法を開発し，多読のトレーニングを支援するシステムを構築することを目的とする．

３．研究の方法

当初の計画では，以下の３段階で研究を進める予定であった．

(1) 調査・検討
自動要約，言い換え，質問応答システム等の言語処理研究の最新の動向調査を行う．並行して，言語テストの分野の研究成果や実際の TOEIC 等のテスト問題の調査を基に，内容を理解・整理しながら読めているかを確認するための問題として，どのような形式の問題が適切で，かつ，問題とその解答を自動生成できるかを検討する．

(2) システム作成
上記(1)の検討結果を基に，文書の内容に関する問題と解答を自動生成するシステムを作成する．

(3) システム評価
システムが生成する問題と解答の質（システムが生成した問題が適切であり，当該の文書を読み直すことなく，答えられるかどうか，また，システムが生成した解答が正しいかどうか）を評価する．また，被験者実験により，システムを利用した場合としなかった場合の多読トレーニングの効果の観点でのシステム評価を行う．

【平成２４年度】

上記の計画の(1)を実施した．

清水の報告（清水真紀：「リーディングテストにおける質問タイプ」，STEP BULLETIN，17，48-62 (2005)）では，問題を以下のように分類している．

(a) パラフレーズ質問
文章中の局所的なある一部分を言い換えると，質問とそれに対する正解が得られる質問．

(b) 推論質問
文章に基づいて適切に推論されることについて問う質問．

(c) テーマ質問
パラグラフまたは文章全体の主題について問う質問

(d) 指示質問
代名詞または指示表現の先行詞の理解について問う質問

(e) 語彙質問
語彙の意味について問う質問（文脈の情報がなくても正解することができるもの）．

(f) 文章構造質問
比較・対照や時間順など文章構造について問う質問やある内容が文章中のどの部分で述べられていたかを問う質問．

内容を理解・整理しながら読んでいるかを確認する問題としては，パラフレーズ質問，推論質問，指示質問が適切と考えられる．推論質問に答えるには，常識的な推論を必要とするが，このような問題と解答を自動生成するには，非単調な推論（矛盾を含む知識からの推論）の枠組と，常識という知識の収集が必要となり，現状では困難である．指示質問の解答作成のためには，照応解析を行う必要があるが，現状ではまだ先行詞の推定精度は十分ではない．一方，文を別の表現で言い換え

る（パラフレーズする）ことは，現在の自然言語処理技術でも可能であるため，パラフレーズ質問が最も実現可能性が高い．そこで，読んだ文書の文の中から，比較的重要な文を複数抽出し，その内の１つに意味を変えない変換を施し，残りの文には読んだ文書の内容と異なるような変換を施し（たとえば，肯定・否定の反転，目的語の名詞を同じ属性を持つ文書中の別の名詞に置換するなど），これらを選択肢とする『読んだ文書に合う文を１つ選択する』選択式の問題を自動生成する問題とすることにした．正解は意味を変えない変換を施して生成した文であるから解答も同時に生成できる．比較的重要な文を元にして選択肢の文を作成するのは，学習者が重要部分を同定し，その意味を理解しているかを問うためである．

上記のような形式の問題を生成するには，文書の重要文の抽出，および，言い換え（パラフレーズ）が中心技術となる．

重要文の抽出は，自動要約等で良く研究されており，現在の主流は機械学習によるものである．しかし，学習者が選んだ任意の文書に対して重要文の抽出を機械学習に基づく手法で精度良く行うためには，様々なジャンルや内容の文書を対象とした自動要約の正解データが必要となり，実現は困難である．また，自動要約や情報検索の分野で古くから用いられている tf-idf に基づく語の重要度は，テーマが比較的類似した大量の文書が必要となるため，これも学習者が選んだ任意の文書に対して重要文を抽出する手法としては適していない．そこで，大量の学習データ（原文章と要約後の文章の組）やテーマが比較的類似する大量の文書を必要とせず，対象の文書だけを用いて語の重要度を推定することができる松村らの手法（松村他：「語の活性度に基づくキーワード抽出法」，人工知能学会論文誌 17 巻 4 号 F，pp.398-406(2002 年)）を利用することにした．この手法は，「文書は著者が自分の考えを読者に伝えるために書かれるものであるから，文書を読んだ後に強く読者の記憶に印象を残すような語が著者の主張を表すキーワードとしてふさわしい」という考えに基づいている．したがって，松村らの語の重要度に基いて重要文を抽出した場合，文書を読み進める上で整理し一時的にでも記憶しておくべき重要文を抽出できると期待され，この意味でも本課題の趣旨に合った手法と言える．

パラフレーズも自然言語処理の分野で良く研究されているが，Ｈ２４年度末の時点までの調査では，十分な精度が保証されている（評価実験がきちんと行われている）パラフレーズのための大規模な知識で無償公開されているものは発見できなかった．そこで，誰もが入手可能な情報のみを用いた，意味が変わらない言い換え手法を開発することにした．

【平成２５年度】

上記研究計画の(2)を実施した．ただし，意味を変える変換はルールベースの手法で実現可能なため，重要文の抽出と意味を変えない言い換え手法に関する研究を行った．

重要文の抽出法について

松村らの手法は，文書を（節や章などの）セグメントに分割し，セグメント内の１文中での語の共起頻度にしたがって，セグメント単位で，順次，語の活性度を伝搬させる．最後のセグメントに対する活性伝搬が終了した時点で，活性度が高い語，および，活性度を活性回数で割った値（鋭活性度と定義しておく）が高い語を重要語として抽出している．本課題では，松村らの手法による語の重要度を用いて，文中の語の重要度の和を $\log(n+1)$ で割った値（n はその文の単語数）を文の重要度とし，重要文の抽出を試みた．語の重要度としては語の活性度および鋭活性度それぞれを試みた．

“The Free Library” (www.thefreelibrary.com) より，1,500 単語程度からなる文書を 20 文書選択し，各文書を３名の被験者が読んで５つの重要文を選択し，１名でも重要文として選択した文を重要文とする評価データを作成した．１文書の平均の文数は 57 文で，平均の重要文数は 10.6 文であった．提案した文の重要度で上位５文を重要文として抽出したところ，語の活性度に基づいた文の重要度の方が若干 Precision が高く 46%であった．

意味を変えない言い換え手法について

言い換え後の表現の妥当性を出現頻度に基づく語の結びつきの強さで測ることができるように，言い換えの対象を元々結びつきが強い，「他動詞＋（冠詞＋形容詞＋）名詞」に限定した．他動詞を v，名詞を n とし，v の同義語の一つを v_i，n の同義語の一つを n_j とすると，“v +（冠詞＋形容詞＋）n” の言い換え候補として以下を生成する．

$$v_i + (\text{冠詞＋形容詞＋})\, n \quad (i = 1, 2, \cdots, I)$$
$$v + (\text{冠詞＋形容詞＋})\, n_j \quad (j = 1, 2, \cdots, J)$$

同義語は WordNet (wordnet.princeton.edu) を利用して求める．動詞あるいは名詞をその同義語で置換した動詞句は，必ずしも自然な表現とは限らない．しかし，置換後の他動詞句が自然である場合，意味も保存されている可能性が高い．そこで，自然さを２つの語の結び付きの強さである PMI(Pointwise Mutual Information) で計測し，生成された候補に，置換した語（同義語）の頻度が閾値 H 以上で，PMI が閾値 θ 以上である候補があれば，PMI が最大の候補を言い換えとして出力する手法を提案した．置換した語の頻度が H 以上という制限を課すのは，低頻度の語は PMI の信頼度が低いためである．また，PMI を計算するための語や語の共起頻度は，検索エンジンのヒット数で近似した．

“The Free Library” の記事から 150 個の動詞句（「他動詞＋（冠詞＋形容詞＋）名詞」）

を抽出し，評価対象のデータとした．評価対象の動詞句の他動詞または目的語の名詞をその同義語で置換してできるすべての言い換え候補を，英文校正の専門家に依頼して，

- Natural and similar meaning
- Unnatural and similar meaning
- Natural and different meaning
- Unnatural and different meaning

の４つに分類した．生成した言い換えが元の表現と同じ意味（上記の Natural and similar meaning）になっている割合（Precision）を評価指標とする．また，作成する選択肢の文の内，最低１つは意味が変わらない言い換えを施す必要があるため，言い換えが生成される割合（Gain）も評価指標の一つとする．トレーニングデータに対して，Gain が G %以上（G = 10，20，30，40）で Precision が最大となるように２つの閾値 θ,H を設定し，テストデータの Gain と Precision を求めるという交叉検定により評価した（分割数５）．結果を表１に示す．

表１．言い換え手法（H25 年度版）の性能

G (%)	Gain (%)	Precision (%)
10	9.3	66.7
20	27.3	59.0
30	30.7	56.4
40	37.3	50.7

【平成２６年度】

計画では平成２６年度は，被験者（学習者）実験により，多読支援システムの評価を行うことにしていたが，前述のように実用に耐える性能ではなかったため，計画を変更し，重要文の抽出手法，意味を変えない言い換え手法の改善を図ることにした．また，選択問題の選択肢の文は読んだ文書中の文から抽出した文を加工したものとなるため，著作権等を研究対象とする研究者を研究分担者に加え，著作権上の問題がないかを検討することとした．

重要文の抽出法について

活性度が高い語も鋭活性度が高い語も，どちらも重要語と考えられる．鋭活性度の定義から，２つの指標はレンジが異なる．そこで，活性度が最大の語の活性度 a_{max} で語 w の活性度 a_w を割った値と，鋭活性度が最大の語の鋭活性度 s_{max} で語 w の鋭活性度 s_w を割った値の内，大きい方（小さくない方）を w の重要度として，重要文の抽出を行った．文の重要度は前年度の定義と同様である．前年度と同じ評価データを用いて評価した結果，Precision は 47%で，わずかに 1%の向上に留まった．

意味を変えない言い換え手法について

語や語の共起を検索エンジンのヒット数で近似していることが，性能が低くなっている最大の原因と考えられた．そこで，Wikipedia の body 部分の全文をローカルに保存し，Tree Tagger を用いて品詞を同定し，品詞列の単純なパターンマッチにより，他動詞と名詞句からなる動詞句を抽出し，他動詞の頻度，名詞の頻度，他動詞と名詞の共起頻度を求めた．

評価は前年度と同じ評価用データと手法により行った．結果を表２に示す．表１と２を比較すると，15～20% 程度 Precision が向上していることがわかる．

表 2．言い換え手法（H26 年度版）の性能

G (%)	Gain (%)	Precision (%)
10	10.0	81.8
20	18.7	77.3
30	32.7	75.3
40	40.0	72.1

４．研究成果

重要文の抽出手法に関して，最終年度に手法の改善を試みたが，わずかに 1%の精度向上に留まった．しかし，学習者が選んだ任意の英文書に対して内容に関する問題と解答を生成する必要があり，そのためには，松村らの手法のように，対象とする文書だけを利用して語の重要度を求める手法に基づいた重要文抽出法が望ましい．しかも，松村らの手法は，語の重要度の考え方からも，問題の選択肢の文を生成するための原文（選択肢原文）を選択する尺度として有効に働くと期待できる．

今回の評価が，選択肢原文の選択の評価として適切であったか疑問が残る．選択肢原文は，対象とする文書の非常に重要度が高い文である必要はない．提案手法により選択された文を，「読み進める上で，整理し，一時的にでも記憶しておくべき文か否か」という観点で直接評価する必要がある．ただし，今回の評価用データ作成に利用した３名の被験者の重要文判定の一致率をκ統計量で見てみると，0.27 前後であり，非常に低い一致率となっている．このことは，被験者が変わると重要と判断する文が異なることを意味しており，「読み進める上で，整理し，一時的にでも記憶しておくべき文か否か」という観点での評価も被験者毎に大きく異なる可能性がある．このことを考慮した評価データ作成を検討する必要がある．

意味を変えない言い換え手法に関しては，簡便な手法にしては非常に高い Precision が得られた．問題の選択肢の文を生成するための原文を 10 文程度抽出し，そのうち最低１文は意味を変えない言い換えをするとするならば，Gain は 10% 程度で良い．このときの Precision は 81.8% と高い．しかし，これでも十分な精度とは言えない．誤った問題と解答の組が生成されては，学習に逆効果であるため，Precision は 100% を目指す必要がある．誤りの原因を調査してみると，提案手法で生成された言い換えが Natural and similar meaning でないものは，ほとんどが

Natural and different meaning と評価されているものであった．したがって，さらなる改善のためには，自然な（つまり他動詞と名詞の結びつきが強い）候補の動詞句の内，元の動詞句と意味が同じ動詞句を選択するように手法を改良する必要がある．これは，分布類似度に基づく多義語の語義解消手法が参考になると考えており，課題終了後も引き続き研究していく予定である．

選択問題の選択肢の文は読んだ文書中の文から抽出した文を加工したものとなる．本課題が目指した問題と解答の自動生成システムではなくても，問題の生成に読んだ文書中の表現を利用する場合は，やはり著作権上の問題がないか検討しておく必要がある．検討結果は以下の通りである．

まず，「読んだ文書の内容に関する問題と解答の自動生成システム」を実際に学習者が個人で利用する場合については，私的使用の範囲と考えられるため，著作権上の問題は起きない．ただし，参加人数が多い授業において，同一の文書を一斉に読ませ，問題と解答の自動生成システムを利用したテストを行うような場合は，教育目的でも，著作権者の利益を不当に害するおそれがあり，著作権上の問題が残る．

一方，システムの開発過程で，生成される問題の質や解答が正しいか否かの調査・検証のため，専門家に評価してもらう必要があるが，この際，問題と解答の自動生成システムを利用することは，私的使用の範囲は超えるものの，技術の開発又は実用化のための試験の用に供するための利用に該当すると考えられ，著作権上の問題は生じない．

また，本自動生成システムが完成し，それを Web 等で公開する場合には注意を要する．悪意ある第３者が，本システムを利用して生成した問題と解答を著作物とする場合は，著作権侵害となる．このような目的での利用を防ぐように公開に際しては利用規約に私的使用のみを許すことを明記する等が必要である．

５．主な発表論文等

〔学会発表〕（計 2 件）

① S. Okaku, Y. Tomiura, E. Ishita, S. Tanaka, Towards Generating Multiple-Choice Tests for Supporting Extensive Reading, The 7-th International Conference on Mobile, Hybrid, and On-line Learning (eLmL 2015), Feb. 23, 2015, Lisbon (Portugal).

② S. Okaku, Y. Tomiura, K. Shu, S. Tanaka, Towards Generating Multiple-Choice Tests for Evaluating Comprehension of Arbitrary English Texts, 5th International Conference on E-Service and Knowledge Management (ESKM 2014), Sep. 3, 2014, Kitakyushu-shi (Japan).

〔その他〕
ホームページ等
以下の Web ページに平成２６年度までの成果を公開予定.
http://nlp.inf.kyushu-u.ac.jp/project.html

６．研究組織

(1)研究代表者
冨浦　洋一　（TOMIURA，Yoichi）
九州大学・システム情報科学研究院・教授
研究者番号：１０２１７５２３

(2)研究分担者
田中　省作　（TANAKA，Shosaku）
立命館大学・文学部・教授
研究者番号：００３２５５４９

安東　奈穂子（ANDO，Nahoko）
九州大学・法学研究院・研究員
研究者番号：５０３８０６５５
（平成２６年度のみ）