

以降, 3. では, 新聞記事から学習課題を自動生成するための予備調査として, はてなカードで出題された学習課題の分析を行う。

3. 調べ学習課題の分析

3.1 分析の目的と対象データ

テンプレートが作成できれば, 記事から抽出した重要語を適応することで学習課題が自動生成できる。そこで, テンプレートが作成可能であるかを調査するために, はてなカードのアンカー文字列(新聞記事の重要語)が課題文に含まれているかどうかを調査する。なお, 重要語と完全に一致する語句が含まれている場合のみ, 課題文に重要語が含まれていると判定する。ただし, 重要語を部分文字列として含む複合語が課題文に出現する場合は, 含まれていないと判定する。例えば, 重要語 <被爆>(以下, 重要語は<>の中に表示)を含む複合語「被爆国」は重要語と異なる語句と判定し, 調べ学習課題の中に重要語 <被爆> は含まれていないと判断する。このように判断する理由は, 新聞記事中に出現する重要語 <被爆> と課題文に出現する <被爆国> では, 語句の概念の抽象度が異なるため, 新聞記事から抽出した重要語を単純にテンプレートに適応するだけでは生成できない課題とみなされるためである。

なお, 分析対象のデータは, 小学校6年の記事76件に付与されたはてなカード228件に記載されている811題の学習課題とする。

3.2 重要語の有無による分類結果

811題の学習課題を, 重要語の有無によって分類した結果, 重要語が含まれている課題が587題, 含まれていない課題が224題になった。この結果より, 72.3%の学習課題文には重要語が含まれていることになる。つまり, 課題文をテンプレート化することができれば, 重要語を当てはめることで新たな課題文を生成できるといえる。

重要語が含まれている課題文には, 課題文中の重要語を別の語句に置き換えることで, 新しい課題文を生成できると考えられるものが多数確認できた。例えば, 課題文「<パラグアイ>ってどんな国だろう?」において, 重要語 <パラグアイ> を <日本> に置き換えることで, 「<日本>ってどんな国だろう?」という新課題文が生成できる。

そこで, はてなカードの課題文をテンプレート化することが可能であるかを分析するために, まずは課題文を疑問詞タイプで分類する。

3.3 疑問視タイプによる分類

重要語が含まれている587題の課題文を, 疑問詞タイプで分類する。疑問詞タイプは, When, Where, What, Who, Why, How, Which とし, 課題文に含まれる疑問詞を手掛かりに分類する。なお, 課題文に疑問詞が含まれない場合は Other に分類する。

分類結果を表1に示す。表1より, How や What に分類される課題が多いことがわかる。また, Other が109件であることから課題文に疑問詞が存在しないものも多い。各疑問詞に分類された課題の中にも, 「<坂本龍馬>ってどんな人?」や「<地盤>ってなに?」などのように, テンプレート化が容易な課題と, 「<びょうぶ絵>で有名な江戸図びょうぶで江戸の様子を調べてみよう!」や「なぜ, 畠山重忠は<一ノ谷の戦い>で馬を背負って下りたのだろう?」などのように, 課題文の構造に汎用性がなくテンプレート化が困難な課題がある。

表1. 疑問詞による分類

How	What	Why	Where	Who	When	Which	Other
249	200	21	5	2	1	0	109

3.4 調査の考察

分類された課題数が多い How, What, Other について考察する。まず, How に分類される課題で最も多いタイプは, 「<裁判官>ってどんな職業なの?」や「<クニマス>ってどんな魚だろう?」など, 「<重要語>ってどんな{...}だろう」タイプで39.4%を占めた(以下, {...}の部分は被修飾語と呼び, {}の内に表示)。「<重要語>ってどんな{被修飾語}だろう」タイプの課題を生成するためには, 重要語と被修飾語の関係を考える。例えば, 重要語と被修飾語の関係は, 「<日本>ってどんな{国}だろう」という課題では適切だが, 「<領有権>ってどんな{国}だろう」という課題では適切でない。よって, 重要語と被修飾語の関係が適切になるようにしなければならない。重要語と被修飾語は, 概念が上位語と下位語の関係になっていることが多い。例えば, 先程の例では, 重要語<日本>に対して, 被修飾語{国}は上位語となっている。因みに, 「<古代オリンピック>ってどんな{内容}だったの?」など, 重要語と被修飾語の概念が上位下位関係になっていないものがあるが, 僅かである。そこで, 「<重要語>ってどんな{被修飾語}だろう」タイプの課題を作成するためには, 重要語に対する適切な上位語を選択して被修飾語の部分に当て嵌める必要がある。

How に分類された「<重要語>ってどんな{被修飾語}だろう」タイプ以外の課題は, 「<比例選得票数>って, どんな意味が有るんだろう?」や「<盧舎那仏>にはどんな思いがこめられているの?」など様々で分類が難しいため, テンプレート化も困難なものが多い。

What に分類される課題で最も多いタイプは, 「<提訴>ってなに?」や「<元寇>って」などの「<重要語>って何だろう」タイプで93.5%を占めた。このタイプはテンプレート化が容易であり, どのような重要語にも対応できる。

疑問詞が存在しない Other に分類される課題で最も多いタイプは, 「<化石>について調べよう!」や「<家紋>について調べてみよう!」などの「<重要語>について調べてみよう」タイプで68.8%を占めた。このタイプも What と同様, テンプレート化が容易であり, どのような重要語にも対応できる。

以上より, What と Other は, テンプレート化が容易な課題が多いが, How は, テンプレート化が困難な課題が多い。本稿では, How の中で最も多い, 「<重要語>ってどんな{被修飾語}だろう」タイプに注目し, このタイプの学習課題を自動生成する手法について検討する。

4. How 型調べ学習課題の自動生成の検討

4.1 導入

How 型の「<重要語>ってどんな{被修飾語}だろう」タイプ学習課題を生成するためには, 新聞記事から抽出した重要語に対応する被修飾語を選定する必要がある。被修飾語の中には, {もの}や{こと}のように汎用的な語がある。例えば, 重要語が<クニマス>の場合, 被修飾語{もの}を使って, 「クニマスってどんなものだろう?」という学習課題を生成できる。学習課題を抽象的にすることで, 小学生に自由に調べさせる教育も考えられる。小学校教師に対するインタビューの結果, 語彙力・調査力に個人差が大きい小学生の場合, 調査した内容が発散したり,

間違っ事柄を調べたりする可能性が高く、あまり教育効果が得られないことがわかった。

そこで、被修飾語を{もの}や{こと}というような抽象的な語ではなく、「<クニマス>ってどんな{魚}だろう?」のように、重要語が何であるかを限定する被修飾語を使用することによって、小学生が的確に調べられると考えられる。本研究では、「<重要語>ってどんな{被修飾語}だろう」タイプの学習課題の自動生成の際、汎用的に使用できる{もの}、{こと}以外の被修飾語を使用して学習課題を生成することを目指す。

4.2 シソーラスを利用した課題生成の検討

「<重要語>ってどんな{被修飾語}だろう」タイプの学習課題の場合、被修飾語は重要語の上位概念になることが多い。例えば、日本語語彙大系[池原 97]において、「<日本>ってどんな{国}だろう」の被修飾語{国}は、重要語<日本>の下位概念となっている。このような上位概念と下位概念の関係を利用し、重要語に対する被修飾語をシソーラスから抽出することで学習課題が生成できる。しかし、シソーラスを使用して「<重要語>ってどんな{被修飾語}だろう」タイプの学習課題を自動生成するためには、重要語と被修飾語がシソーラス上に掲載されている必要がある。

そこで、よみうり博士のアイデアノートに掲載されている小学6年生用の記事96件のはてなカードを利用して、日本語語彙大系でのカバー率を調査する。対象記事96件には、「<重要語>ってどんな{被修飾語}だろう」タイプの課題が90題存在する。この中から、重要語と被修飾語のセットを抽出し、同じ語句の組み合わせであるセットを集約した80セットを調査対象とする。本調査では、重要語、被修飾語、重要語と被修飾語のセットの両方が日本語語彙大系に掲載されているかどうか(含有率)を調査する。

調査結果を表2に示す。調査結果より、日本語語彙大系には、重要語の30.0%、重要語と被修飾語のセットの20.0%しか掲載されていない。新聞記事の重要語には、“大日本沿海與地全図”や“全国ディベート甲子園”などの固有名詞や、“迷惑メール”などの最近使われるようになった語、“公家装束”などの普段使われる頻度が少ない語が多く存在する。このような語は日本語語彙大系に掲載されていないことが多く、重要語の含有率が低くなったと考えられる。

表2. 日本語語彙大系の含有率

	含まれる個数	含有率(%)
重要語	24	30.0
被修飾語	59	73.8
重要語と被修飾語	16	20.0

この調査結果を踏まえると、多くの重要語の場合、日本語語彙大系の上位下位概念を利用して重要語に対応した被修飾語を選定することは難しい。シソーラスとして日本語語彙大系以外にも、日本語 WordNet などが利用可能であるが、同様の問題が存在すると考えられる。

そこで本稿では、シソーラスだけでなく、Web上のテキスト情報も活用することで、「<重要語>ってどんな{被修飾語}だろう」タイプの学習課題を自動生成する手法を検討する。

4.3 Web上のテキスト情報を活用した課題生成の検討

Web上には、「<重要語>ってどんな{被修飾語}だろう」タイプの学習課題が存在する可能性もある。そこで、Web検索エンジンを使用して、「<重要語>ってどんな*(*印はワイルドカード)」

というフレーズで検索し、上位 n 件の検索結果から頻度情報を基に重要語に対する被修飾語候補を選出する。そして、{もの}や{こと}以外で出現頻度が最大の語を被修飾語とする。

例えば、重要語<クニマス>の場合、「クニマスってどんな*」でフレーズ検索を行う。検索結果全28件から、被修飾語候補として、{魚}、{味}、{おさかな}、{マス}、{サカナ}が抽出できたとする。これらの被修飾語の出現件数は{魚}が18件、{味}が6件、{おさかな}が2件、{マス}が1件、{サカナ}が1件である。一番出現頻度が高い被修飾語候補を被修飾語とする。この結果を利用し、「クニマスってどんな魚だろう」という学習課題を生成する。

しかし、上記の手法では、「築城家ってどんな*」というクエリで検索した場合、検索結果が0件となるため、被修飾語が得られない問題がある。そこで、対応策について検討する。被修飾語には、{人}や{国}などの多く使用される被修飾語が存在する。そこで、重要語に対して、上記の方法で網羅的に収集した被修飾語を候補として、重要語と被修飾語候補の関連度を求め、被修飾語を選定する手法を検討している。

今後は、被修飾語候補のデータの作成と、関連度を求める手法について検討を進め、上記で述べた手法を段階的に評価する予定である。

5. おわりに

本稿では、小学校でのNIEを対象とした新聞記事読解システムの機能の一つである調べ学習支援について概説した。新聞記事から学習課題を自動生成するための予備調査として、既存の調べ学習課題の分析した。その結果、文中に重要語が含まれている課題の中で、How型が一番多いことがわかった。また、How型の中でも、「<重要語>ってどんな{被修飾語}だろう」タイプが多いことがわかった。このタイプの課題を生成するため、シソーラスを利用する手法について検討した結果、含有率の問題があることを確認した。そこで、シソーラスだけでなく、Web上のテキスト情報を併用する手法について検討した。

今後は、検討した手法の実装・評価を行うと共に、その他の調べ学習支援機能について検討を進める。

謝辞

本研究の一部は、文部科学省科学研究費補助金(若手研究(B)22700813)の助成を受けて実施した。

参考文献

- [NIE 13] <http://nie.jp>
- [読売 13] よみうり博士のアイデアノート:
<http://www.yomiuri.co.jp/nie/note/top.html>
- [小林 13] 小林 健, 安藤 一秋: 小学生を対象とした新聞読解支援のための説明語抽出手法, 情報処理学会研究報告, Vol.2013-CE-119, No.17, 2013.
- [藤沢 12] 藤沢 祐輔, 相原 慎太郎, 安藤 一秋: Web一般新聞記事を子供向けに言い換える知識の抽出, 言語処理学会第18回年次大会論文集, P2-5, pp.751-754, 2012.
- [小田 00] 小田 光宏, 北本 正章, 古賀 節子: マルチメディア教材による学習課題設定支援の有効性, 日本教育情報学会年次論文集(16), P50-53, 2000.
- [池原 97] 池原 悟, 宮崎 正弘, 白井 諭, 横尾 昭男, 中岩 浩巳, 小倉 健太郎, 大山 芳史, 林 良彦: 日本語語彙大系, 岩波書房, 1997.