

## NIE ワークシートの自動生成に向けた問題文の分析 Analysis of Questions for Automatic Generation of NIE worksheets

國方 海秀<sup>†</sup>      安藤 一秋<sup>‡</sup>  
Kunikata Kaishu      Ando Kazuaki

### 1. はじめに

近年、小学校や中学校の義務教育機関において、Newspaper In Education, 通称 NIE が実践されている。NIE とは、新聞を教育に活用する取り組みのことであり、教育的な効果も確認されている[1]。しかし、NIE を実践するにあたり、「教材研究の時間が足りない」、「新聞活用のための時間の確保が難しい」、「カリキュラムとの調整が難しい」といった課題がある。新聞社は、教師の NIE 実践を支援するために、新聞記事と問題で構成された NIE ワークシートを Web 上で配信しており、教師は NIE で利用できる。しかし、NIE ワークシートを配信していない新聞社もあるだけでなく、配信している場合でも、授業で扱いたい記事や問題を含むワークシートが存在していない場合も多い。そのため、NIE を実践する教師は、独自にワークシートを準備する必要がある。

そこで、本研究では、小学校教師に焦点をあて、NIE に対する授業準備の負担を軽減するために、NIE ワークシートの自動生成を目的とする。本稿では、問題文の自動生成に向けた初期検討として、NIE ワークシートで採用される記事のジャンルと問題文について分析する。

### 2. NIE ワークシートの分析

#### 2.1 分析の概要

Web ニュースから問題文を自動生成するための初期検討として、まず NIE ワークシートの記事ジャンル、出題方式、問の対象に着目して、傾向を分析する。その後、問題文の構造についての特徴を分析する。

分析対象のワークシートは小学生を対象にしたものであり上毛新聞 (30 件)、南日本新聞 (30 件)、神戸新聞 (30 件)、山梨日日新聞 (30 件)、静岡新聞 (30 件)、大分合同新聞 (30 件)、岩手日報 (10 件)、北海道新聞 (10 件)、下野新聞 (10 件) の合計 210 件である。また、問題数は 708 件である。

#### 2.2 記事ジャンルの分析

NIE ワークシートに採用されやすい記事ジャンルを確認するため、各記事のジャンルについて分析する。記事ジャンルは、Web ニュースの分類を参考に、14 ジャンル (社会、科学技術、自然、防災、スポーツ、農業、伝統、歴史、経済、国際、政治、教育、漁業、その他) を利用する。

分析の結果、社会記事を取り扱った社会ジャンル (66 件) が最も多く、その次に自然記事 (19 件)、スポーツ・科学技術・防災ジャンル (各 16 件) が多いことを確認した。

### 2.3 出題方式の分析

新聞記事からどのような形式で問題が出題されるのかを確認するため、NIE ワークシートの問題の出題方式を分析する。出題方式には明確な指標が無い場合、一般的な出題方式に分類する。

分析の結果、文章で解答する問題が 369 件と最も多くなった。次に単語で解答する問題が 265 件となり、以降は 80 件未満となった。このことから、NIE ワークシートに主に使われているのは、文章で解答する問題と単語で解答する問題の 2 つの方式が中心であることが分かった。

### 2.4 問題が問う対象の分析

問題の解答からの問題文生成の実現可能性を探るため、問題が問う対象について分析する。理由や目的などを問う問題以外に、固有表現を問う問題も存在していることから、RIKEN AIP の拡張固有表現[2]の第 2 階層を用いて分類する。また、事前調査により、理由と感想は、記事内に解答が存在するものと記事内には解答が存在せず、児童自身が考察するものが確認できたため、理由 (記事参照) と理由 (考察) のように細分類する。

分析の結果、製品名を問う問題が最多で 68 件確認できた。拡張固有表現における製品名は、一般的な製品のみならず、政策計画名や職業といった称号名なども含むため、多くなったと考える。次いで、地名が 54 件、理由 (記事参照)・時間・意見を問う問題がそれぞれ 49 件存在した。特に上位には、固有表現を解答とするものが多いことを確認した。理由や意見などを問う問題のほとんどが文章で解答する問題であったが、固有表現については、単語で解答する問題だけでなく、穴埋め問題や児童に調査を促す調査問題でも対象になるためである。

### 2.5 記事ジャンル別の分析

ジャンル別の出題傾向を確認するため、記事ジャンルごとに問題の出題方式と問題が問う対象を分析する。

分析の結果、出題方式は、14 ジャンル中 10 ジャンル (社会、科学技術、自然など) において、文章で解答する問題の割合が多く、4 ジャンル (農業、歴史、国際、漁業) において単語で解答する問題の割合が多いことを確認した。しかし、漁業のみは記事数が 2 件しかなく、今後の調査が必要である。他のジャンルは、30 件以上の記事が対象であるため、ジャンルごとの出題方式に傾向があると考えられる。

次に問題が問う対象について分析した結果、全体的に理由や意見、製品名、地名などを問う問題が多く見られた。単語で解答する問題が多い農業ジャンルでは、地名、自然物名、寸法表現、製品名、時間、期間と上位 6 種が固有表現であった。同様に、単語で解答する問題が多い歴史ジャンルにおいても、上位 3 件が固有表現であった。したがって、これらのジャンルにおいては問題が問う対象に傾向があると考えられる。しかし、防災ジャンルや政治ジャンルなど

<sup>†</sup> 香川大学 大学院 創発科学研究科, Graduate School of Science for Creative Emergence, Kagawa University

<sup>‡</sup> 香川大学 創造工学部, Faculty of Engineering and Design, Kagawa University

は、件数に差があまり見られず、傾向を確認することができなかった。

## 2.6 問題文の分析

問題文の自動生成方式として、テンプレートを用いる方法と言語モデルを利用した文生成が考えられる。そこで、問題文の構造的な特徴を確認するため、まずは文末表現に注目して分析する。

問題文を文節で分割し、文末部分の件数を確認する。問題文を形態素解析し、名詞を L、数字を 0 に一般化して分析する。また、「書きましょう」や「書いてください」などの意味的に類似するものは同一対象としてカウントする。

結果を表 1 に示す。表中の割合は、問題文 708 件に対する割合である。表 1 より、何かを書かせることを促す文末「書きましょう」が最多で、全体の 21%存在している。また、2 番目は「L ですか」(例：意見ですか)であり、記事中のキーワードを問う問題が 10%を占めた。このタイプの問題には、文末のみでなく、それ以前の表現にも特徴があると考えた。そこで、このタイプの問題文を確認した結果、「どんな」や「どこで」といった表現が確認できたため、文末以前の表現も考慮した調査が必要であると考えた。

表 1 文末表現の出現数調査の結果

問題text	件数	割合
書きましょう	151	21%
L ですか	71	10%
調べてみよう	67	9%
何ですか	57	8%
挙げましょう	16	2%
ありますか	14	2%
考えてみましょう	12	2%
何L ですか	9	1%
思いますか	7	1%
なぜですか	7	1%
説明してください	7	1%
合計	418	59%

そこで、文中表現として疑問詞に着目して、表 1 の上位 4 件の文末表現を含む問題文に、疑問詞が出現する件数を調査する。疑問詞には、一般的な 5W1H を用いる。

文末表現別の結果を表 2 に示す。表 2 から、全体的に how の要素が最も多いことがわかる。さらに内容を調査した結果、how の中でも「どんな」と「どのような」といった名詞に係る表現が多く確認できた。「L ですか」タイプの問題の場合、72%で 5W1H が使われており、その他のタイプの問題では 30%程度あることから、差があることを確認した。また、「何ですか」タイプでは、文末に「何」を含んでいるため、文中には 5W1H 表現があまり出現しなかったと考える。「調べてみよう」や「書きましょう」タイプでは、5W1H が使用されている割合が約 20%と少なく、5W1H 以外の特徴的な表現の有無を今後は調査する。

以上の結果から、文中表現 how の要素である「どんな」、「どのような」と文末表現「L ですか」を組み合わせることで、テンプレート形式でも問題文を生成できる可能性があるといえる。

表 2 文末表現と 5W1H の共起数調査の結果

	what	who	when	where	why	how	合計	割合
書きましょう	0	3	0	1	0	30	34	23%
L ですか	0	5	2	8	0	36	51	72%
調べてみよう	0	0	1	2	1	10	14	21%
何ですか	1	0	1	1	0	2	5	9%
合計	1	8	4	12	1	78	104	30%

## 3. 分析のまとめ

本稿では、Web ニュースから問題文を自動生成するための初期検討として、5 つの分析を実施した。まず、小学生を対象とした NIE ワークシートに採用される記事ジャンル、出題方式、問題が問う対象について調査することで、どのような記事にどのような出題がされるかを分析した。その結果、農業、歴史、国際のジャンルのみに単語で答える問題が多く、単語としては、拡張固有表現を問う傾向が見られた。したがって、これらのジャンルについては、傾向に沿った問題を生成する必要がある。

次に、問題文の構造を分析するため、文末と文中表現に注目して調査した。文末表現では、「書きましょう」、「L ですか」、「調べてみよう」、「何ですか」の 4 種で約 50%を占めることを確認した。次に、文中表現として 5W1H と文末表現を組み合わせることで分析した結果、「どんな」や「どのような」といった how の要素と「L ですか」タイプの組み合わせが最も多いことを確認した。したがって、テンプレート形式でも問題文を生成できる可能性があると考えられる。例えば、記事中から重要語を抽出し、文末表現「L ですか」と how を組み合わせることでテンプレートによる問題文生成が考えられる。

今後は、記事数・問題文数を増やして分析を継続するとともに、記事内に回答が存在する、もしくは存在しない問題文に特徴があるかといった視点でも分析する。そして、全体の分析結果を踏まえて、テンプレート型と言語モデルによる自動生成型などにより、問題文を自動生成する手法について検討する。

## 4. おわりに

本稿では、NIE ワークシートの自動生成を目的に、問題文の自動生成に向けた初期検討として、NIE ワークシートで採用される記事のジャンルと問題文について分析した。記事ジャンル、出題方式、問題が問う対象の分析では、農業ジャンルなど一部のジャンルで傾向を確認した。問題文の構造分析の結果、文章で解答する問題と単語で解答する課題の両方で利用できる「書きましょう」タイプが最も多いことを確認した。また、5W1H の文中表現に着目した結果、how の要素が多く用いられていることを確認した。

今後は、問題文の構造分析を継続し、その分析結果を踏まえて、テンプレート型と言語モデルによる自動生成型などにより、問題文を自動生成する手法について検討する。

### 謝辞

本研究の一部は、JSPS 科研費 19K12271 と 22K12294 の助成を受けて実施した。

### 参考文献

- [1] 教育に新聞を, <https://nie.jp/>
- [2] 拡張固有表現 - RIKEN AIP, <http://liat-aip.sakura.ne.jp/wp/拡張固有表現-2/>