

小学2年生の問題を解く - 電脳優子2年生国語・概要

関根聡*1 齋藤真実*2 岡田美江*2 井佐原均*3

*1株式会社ランゲージ・クラフト研究所 / ニューヨーク大学

*2長岡技術科学大学電気系

*3独立行政法人情報通信研究機構

1. はじめに

小学校2年生の国語の問題を解くシステム『電脳優子2年生国語』を作成している。本稿では、その目的と概要、関連研究を紹介し、使用した知識や解決できなかった問題に対して考察を行う。本年次大会のデモンストレーションで発表する別の稿(齋藤他 2005)で、実際の問題の解き方と評価結果について詳しく述べ、システムを紹介している。

2. 目的

本プロジェクトの主な目的は、以下の2点である。

1) 自然言語処理の技術を一般の人にも分かりやすい形で実現すること

現在、例えばロボットの技術は二足歩行ロボットの実現や進歩により、一般の人にも分かりやすく、注目を集める対象となっている。自然言語関連の技術では、音声認識、機械翻訳、対話処理などが実現されてきている。しかし、残念ながら、二足歩行ロボットも実は人間ほどの歩行能力がないという点では同様であるにも関わらず、一般論としては、別の形での評判を得ている感がある。それには、自然言語処理の能力を一般の人にも分かりやすい形で評価することが難しいことも一つの原因ではないかと考えられる。そこで、我々は一般の人にも分かりやすいような応用の一例として、小学2年生の国語の問題を解くという課題を設定した。この課題ではシステムの能力は誰にでも馴染みの深い「得点」という形で示される。2年生というレベルは、優し過ぎず難しすぎず、総合的に70点くらいとれることを見込んで設定した。

2) 自然言語処理の対象のレベルを下げ、問題を見つめ直すこと

現在の自然言語処理は、新聞記事や技術文書などの難しいテキストを対象に、複雑な世界知識や意味処理が必要な高度な処理を目指している。そこで、我々は、対象とするテキストのレベルを下げてみると問題点がより明確に見えてくるのではないかと考えた。もちろん、小学生2年生レベルであっても、様々な世界知識を用いて言語活動を行っている。したがって、新聞記

事の記事を対象にしていた際にあった困難な問題がなくなるものではない。しかし、文章や問題のレベルが低くなった分、その問題の内容が比較的容易なものとなり、いくつもの問題が複雑に絡まることなく現出するのではないかと期待できる。また、小学生の問題を対象とすることにより、特定の課題に偏ることなく言葉について幅広く対象にできるのではないかと考えられる。この試みを通じて、言語の全ての問題を見出そうというのではないが、対象問題をシンプルにすることによって、内在している言語の問題の一部がより鮮明に見えてくるのではないかと、という試みである。

3. 関連研究

関連研究として、米国で“Reading Comprehension (読解問題)”を解くという試みが提案されている(Hirschman et al. 1999) (Charniak et al. 2000)。Hirschmanは、Deep Readという読解問題を解くタスクを提案し、予備実験として、新聞などの文章を対象としWhen, Whoなどの質問に対し解答のある文を抜き出すシステムを作成した。システムは、基本的に固有表現や意味の拡張を行ったBag-of-wordsタイプの情報検索(正確には文選択)である。小学3年、6年のレベルの問題で30-40%の質問に対して正しい答えのある文を抜き出すことができたと報告されている。文章問題以外には対応しておらず、ここで提案している「電脳優子」に比較して、より限定されたシステムでしかない。Charniakはこの技術を洗練する試みを行い、動詞や主語に重きを置いたり、Whyの質問に対して特別なルールを導入したりしたシステムを作成したが、本質的なターゲットや手法はDeep Readのシステムと変わりはない。

算数の文章問題を理解しようという試みもある(安西ら 1985) (稲葉ら 2004)。安西らは小学校1, 2年の文章問題を13種類の問題形式の内部表現かその複合表現に帰着し、問題を解くシステムを提案した。語彙は限定され、評価はされていないが、教育目的のシステムとして興味深い。稲葉らは、小学校1年から4年までの文章問題に対して、構文解析などを用いずに、

概念ベースにある単語と、単位を伴う数量表現を抽出し、ヒューリスティックな規則によって、加算減算か乗算除算か四則演算混合問題かを分類する。分類問題の精度は90%で、加減算と乗除算に限った答えの正解率は81%と報告されている。

4. 概要

まず、我々は市販の小学2年生の問題集を5冊買い集めた。その中で「株式会社くもん出版」が出版している「小学ドリル 学力チェックテスト 国語2年生」の問題集が問題形式のシンプルさ、問題の難易度の点から本研究の目的にもっとも適していると考え、これを主な対象にした。問題の種類を分析、考察する際には、他の4冊の問題集も参考にしている。まず、我々は問題の種類を分類した。国語の問題は大きく、漢字の問題、言葉の問題、読解問題、作文問題に分かれる。それぞれの問題を詳細に分類し、問題集全体を、日本電子出版教会の設定した電子出版交換フォーマットである JapaX 準拠の XML の形式で書き直した。このファイルでは、問題がそれぞれの分類ごとに決まった形式で書かれており、問題をテキストから読み取り構造化する部分はすでにこのように人手で行っている。

つぎに、それぞれの問題を解く別々のプログラムを作った。このプログラムの詳細については別稿の論文を参照していただきたい。プログラムが使用した知識やツールは、大きなものでは、辞書、形態素解析、同義語、反意語辞書、シソーラスと大規模コーパスである。漢字、言葉の問題に対しては大規模コーパスが非常に役に立った。文章問題については質問応答と同様の技術でかなり解けたが、問題も残った。

5. 問題の分類

問題集にある問題を分類した。大分類、中分類を表1に示す。実際には、それぞれの中分類に対して、記述問題か選択問題か × 問題なのか、どのような形のヒントが提供されているか等で細かく分類されている。実際に、対象とした問題集では、約100種類の小分類に対応する問題が存在した。違った種類の問題集を見ても、主な所は似ているものが多いが、細かく見ると違った種類の問題も存在している。

現在、問題は様々な書式に対応するために JapeX 準拠のフォーマットで書かれている。しかし、一般に問題は問題用紙に自由に書かれており、最終的なデモシステムでは、これを直接入力とし、それを JapeX などのフォーマットに落とし込むことが望ましい。同時に、そこから自動的にどの分類の問題であるかを判断

する必要がある。現在、このサブシステムも開発している。この部分が作成されれば、将来的には、問題集の1ページをOCRで読み取り、適切な場所に答えを書き出すというシステムも実現できる可能性がある。

表1. 問題の分類

大分類	中分類
漢字	読み、書き、部首、書き順、分類
言葉	カタカナ、仮名遣い、適切な名詞、動詞、形容詞、副詞の穴埋め、類義語、反意語、助詞、接続詞、擬音擬態語、丁寧語、句読点
読解	散文(誰、何、いつ、どこ、どんな、なぜ、どのように、どう、どれくらい)の疑問詞による内容確認; 気持や様子の書き出し、ストーリー経過順、韻文
作文	文の組み立て、書き方

6. 知識・ツール

使用した主な知識やツールは以下の通りである。

単語辞書

単語辞書として、難しい単語が入っていない子供用の辞書を作成した。参考にした辞書は、JUMAN 辞書、くもんの国語辞書、オノマトペ辞書、固有表現辞書などである。特に後者の2点は一般の辞書にはほとんど含まれないが、問題集では頻繁に出現する。単語辞書には、一般的な表記、読みなどの他に、問題中にある一部のみ漢字で表されている単語(「問だい」というような書き方)に対応するための別表記の情報がある。

漢字辞書

漢字の読み、書きに利用するための漢字辞書。読みの他に部首の情報がある。

形態素解析

形態素解析システムはJUMANを基に自ら作成した。JUMANをそのまま使わなかったのは、子供用の文章を対象にしていることで、特別な処理が必要かと思ったためだが、最終的にはその必要はなかった。

同義語、反意語、シソーラス

意味的な辞書は、対象単語の制約もあり、既存のいくつかの辞書やWEBにある知識を基に上記の単語辞書に連携するような形で自ら作成した。

大規模コーパス

漢字の書き取り、言葉の補完など様々な場面で大規模コーパスを文字列頻度検索、共起頻度検索すること

により答えを導いた。このための大規模コーパスとして、新聞記事の延べ 38 年分と自ら集めた 350GB 相当の WEB コーパスを利用した。検索はこの大規模コーパスを対象に検索可能なサフィクスアレーに基づくシステムを利用している（吉平ら 2004）

7. プログラムと評価

問題を解くプログラムと評価については、別稿（齋藤ら 2005）により詳しく述べるが、ここでは全体的な話を簡単に説明する。「問題の分類」の部分で、全部で 100 程度の小分類に分かれたと述べたが、その中には漢字の書き順など計算機が扱うことが困難な問題などがある。その部分を除いた、全体の約 90%の問題を対象に、それぞれの種類の問題ごとに解答を求めるプログラムを作成した。必ずしも全ての小分類に対して別々のプログラムを作成するわけではなく、ある程度共通化できる部分は共通化したが、全体では 47 種類のプログラムからなっている。

評価は、プログラムを作成する時に参照した問題集とは別の問題集に対して行った。その結果を表 4 に示す。全体の問題数は評価の対象となった問題の数であるが、その中には訓練データ中に入らない形式の問題で、今回作成したプログラムでは対処できない問題もあった。これを除いて、プログラムが対処できる問題の数が対象の問題数に示してある。言葉、読解の問題で半分近く減っている。対象になった問題での正解率は、漢字、言葉では非常に高く、読解では低くなっている。総合での正解率は、全体に対しては 55%、対象となった問題に対しては 83%であった。

表 4 . テストデータでのテスト結果

	全体の 問題数	対象の 問題数	正解数	正解率 (対象)[%]	正解率 (全体)[%]	訓練データ全体の 正解率[%]
漢字	76	74	69	93.2	90.8	89.8
言葉	245	140	117	83.6	47.8	71.5
読解	34	22	10	45.5	29.4	63.8
全体	355	235	196	83.4	55.2	80.3

8. 考察

「電脳優子 2 年生国語」の作成を通して見つけた問題について考察を述べる。

問題の認識、分類

まず、国語の問題を解くためには、その問題が何を求めているかを理解しなければいけない。我々は問題集を参考に、約 100 種類の分類を作成し、それらの分類に対応した個別のプログラムで問題を解くという戦略をとった。しかし、ここには 2 つの問題がある。ま

ずは、分類が適切か、未知の分類がないか、ということである。それは、評価結果から明らかな様に否定的な結果である。今回の分類でカバーしきれない問題はたくさんあり、その範囲は今の所見えていない。また、問題を分類するプログラムの精度という問題もある。このプログラムは現在開発中である。

漢字の読み

漢字の読みは複数あるため、辞書を引けば常に答えが得られるというものでもない。「ゆびの間」における「間」は「あいだ」か「ま」か「かん」と読むのか難しい。漢字の書き取りでは、すでに漢字で書かれている文章がコーパスとして大量に存在していたため、そこで頻度を調べれば良かった。しかし、読みがそのまま載っている大規模なコーパスがないため、読みをコーパスで調べることはできない。

大規模コーパスの有用性

新聞記事や WEB の大規模コーパスは様々な場面で非常に有効であった。いくつか具体例を挙げる。

[漢字の書き取り] 「人にあう」「答えがあう」に「会う」「合う」が正しい答えであるということは、それぞれを代入し、コーパスで頻度を調べることによって、かなり正確に解けた。

[助詞の補完] 「かさ{にとを}家{にとを}わすれる。」といった助詞の補完の問題では、候補を代入した文章の頻度をコーパスで調べることにより解ける。この方法は（松井 2004）にも言及されているが、正確な評価をしたものではない。また、全ての組み合わせの代入によっても見つからない場合があるが、適切なコンテキストの削除を行い、解答を得た。また、小学生の問題であるので、一般には当然漢字で書かれるものも漢字で書かれていない場合があるため、かな漢を利用して例が見つかる工夫を施した。言葉の問題にある他の穴埋め問題もほぼ同様の方法で実現した。

[反意語の選択] 例えば「まどをあける」というコンテキストにおいて「あける」の反意語は「しめる」か「とじる」かという問題があるが、これもそれぞれの語を代入し、その頻度を調べることによって解決した。

[関連語の選択] ある刺激語（目や耳）に対して関連する単語（見る、聞く）を選ぶような問題である。これはコーパス内での共起頻度を調べることにより答えを求めた。しかし、「目で見ると」というような当たり前の表現はコーパスにもないという問題があった。辞書の定義文を使うことも試みたがそれでも不十分であった。

このように、漢字や言葉の問題で大規模コーパスは非常に有用であった。同じ問題を、例えば、動詞の格辞書や反意語、類義語、関連語、シソーラスなどの辞

書的な知識として揃え、それによって問題を解く方法もあるだろうが、今回のプログラムを見る限り、ある目的においては大規模なコーパスは人間が編集した知識に勝ると考えられる。コーパスにある知識を予め編集しておけば等価であるとも考えられるが、どのように編集すべきなのか、本当に編集できるのか、何のために編集すべきなのか、は現在の所、不明であり、今後考えてみるのに非常に面白いテーマであると思う。

言葉の問題

言葉の問題は、言葉の意味や遣い方を様々な角度から聞く問題であり、問題のパラエティが非常に多い。それぞれの問題の種類に対してプログラムを用意するという方法が最も難しい種類の問題である。今回は予め分類してしまうという方法でこの問題を避けてしまったが、今後取り組んでいきたい課題である。

世界知識の必要性

小学校2年生の問題を選んだのは世界知識の必要性が低いだろうという予想であったが、それでも様々な世界知識が必要であった。例えば、「小学校を卒業すると中学校に入学する」、「欲しい物を他人からもらうと人は嬉しいと感じる」、「子供は母親に会いたいと思うものである」といった知識である。どのように記述し、どのように使用すべきか難しい問題である。また、世界知識というよりは言葉の知識でもあるが、カタカナ語の候補の中から「外国から来た言葉」や「動物の鳴き声」を選ぶ問題は単なるシソーラスだけでは解決できない問題である。言葉それぞれの属性を設計し、その知識を埋め、それを柔軟に利用する方法を考える必要がある。

読解問題と質問応答の違い

現在、質問応答の技術が対象にしているのは名詞句が答えになるようなシンプルな質問である(QAC HP)。このような問題も存在したが、それを拡張しなければいけない問題も多くあった。例えば、質問応答では固有表現のタイプを認定し、そのタイプにあった候補を探すが、2年生の読解問題では「誰」の質問の答えは、人間と組織に限ることはできない。「裏のタヌキ」や「お月様」さえも答えになるからである。また「それはいつの頃ですか」という時間を聞く質問でも「わかばがしげりだすころ」という節で答とする場合もあり、名詞句のみが答えではない。また、「この文章はいつのお話ですか」というように、ある特定の出来事があった時間だけではなく、話の全体の時を問う問題もあり、単にキーワード近傍の同じタイプの名詞句を抜き出すだけでは無理な問題があった。しかしながら、「では、きちんとした文章理解をしないと解けないのか」とい

う疑問には、直感的にはすぐにはうなずけない。人間として問題を解いてみても、文章を理解しないでも、かなりの問題は解けるからである。では、何が必要なのかというのはこれからの課題である。

その他の種類の技術(談話解析、照応解析)

形態素解析、コーパスでの頻度、質問応答の技術以外にも、自然言語処理で培われた技術を使う場面がいくつかあった。特に大きな2つの例を挙げる。1つは、談話解析の技術である。これは段落ごとに切られた塊を順番に並べ替えるという問題や、文と文をつなぐ接続詞を選択する問題である。また、今回の問題集では直接聞く質問はなかったが、照応解析の問題は非常に重要である。問題が聞いている対象を特定するのも照応の問題であるし、代名詞のリンクを辿ることにより正解が得られるべき問題もあった。

9. まとめ

小学2年生の国語の問題を解くシステムを作成している。一般の人にもわかりやすい応用システムの作成を目指し、かつ、自然言語処理の問題点を見つめ直す機会としている。今後を負うところは多く、精進して行きたいと考えている。

参考文献

- Hirschman, L., Light, M., Breck, E. and Burger, J. D. "Deep READ: a Reading Comprehension system". ACL 1999.
- Charniak et al., "Reading Comprehension Programs in a Statistical-language-Processing Class". Workshop on Reading Comprehension Tests as Evaluation for Computer-based Language Understanding Systems. 2000.
- QAC HP: <http://www.nlp.cs.ritsumei.ac.jp/qac/index-j.html>
- 安西祐一郎、上里譲、田村淳「算数の文章題を解くシステムにおける問題の内部表現について」知識工学と人工知能 41-1. 185
- 稲葉栄美子、渡部広一、河岡司: 2004. 「常識知識を用いた算数問題の意味理解」第159回自然言語研究会
- 齋藤真実、岡田美江、関根聡、井佐原均「小学校2年生の国語の問題を解く - 電腦優子2年生国語・デモ」第11回言語処理学会年次大会
- 松井くにお「検索ロボット技術を活かしたWWW検索技術」日本語学 2004年2月号.
- 吉平健治、武田善行、関根聡: 「WEB文書を対象としたKWICシステム」第10回言語処理学会年次大会