

意味処理を用いた 算数文章題演算処理手法の提案

阿部一貴[†] 吉村枝里子[†] 土屋誠司[†] 渡部広一[†]

コンピュータとの円滑な会話の実現には、人間が学習する教養知識を理解する必要がある。例えば、数量が入った会話では算数に関する教養知識が必要不可欠になる。

そこで本稿では、算数文章題を取り上げ、様々な算数に関する教養知識や辞書を用いた言葉の意味処理、数量の論理関係の整理、文脈の判断を行うことで文章の意味を理解する。自然言語文章から算数文章題か否かを判断し、算数文章題を演算する手法を提案する。

Operation Processing of Arithmetical Problem Using Semantic Processing

Kazutaka Abe[†] Eriko Yoshimura[†] Seiji Tuchiya[†]
and Hirokazu Watabe[†]

As for the achievement of a smooth conversation with the computer, it is necessary to understand the education knowledge that man studies. For instance, the education knowledge concerning arithmetic is necessary and indispensable for the conversation that uses the amount.

In this paper the arithmetical problem expressed by Japanese sentences is treated proposes the method that the meaning of sentences is understood and operates the problem using the semantic processing of the words that uses the education knowledge concerning various arithmetic and the dictionary, using the arrangement related to the logic of the amount, and using the judgment of the context.

1. はじめに

近年、我々の周りではコンピュータが急速に発展し、多種多様な分野で活用されている。その結果、コンピュータは我々の生活において必要不可欠なものとなったが、定型的な入力しか受けつけず、定められた処理を行うのみである。本当の意味で、柔軟に人間の要求や意図を理解しているとは言い難いと言える。そこで、人間の要求や意図を理解し、人間とコミュニケーション（意思疎通）を行う知的なコンピュータが望まれる。そのためには、人間同士が日常行っているコミュニケーションの仕組みをモデル化し、人間が持つ常識的な知識や、社会生活を営む上で必要となる文化に関する広い知識である教養知識を理解することが必要となる。この常識的な知識と教養知識をまとめたものを常識知識と呼ぶこととする。

日常会話の中には数量を含む会話文がある。そのような会話のすべてが、人間がこれまでに学習してきた算数に関する教養知識が基となっている。具体的には、算数の知識を用いることにより演算を行うことができ、その結果を用いて適切な数量等の判断を行い、会話に活かしている。例えば、定員 20 名のバスに、5 名の団体 2 組、7 名の団体 1 組の予約が入っているという以下の図 1 のような状況があるとするとする。



図 1 バス予約のイメージ

客が「あと何名まで予約できますか?」というとき、バス会社は「あと 3 名まで予約できます。」と返事をして会話が成り立つ。これは「定員 20 名」、「5 名の団体 2 組、7 名の団体 1 組」から「残りの定員は 3 名である」ということを演算により把握してい

[†] 同志社大学大学院 工学研究科
Department of Intelligent Information Engineering and Sciences,
Graduate School of Engineering, Doshisha University

るためである。演算に加え、「定員」という語が「容量の限界」を表すことや、「名」は「人を数える単位」というように、算数の知識を知っておく必要がある。このように、数学的な側面から、一般的な会話文と同じように考えることは難しいため、教養知識、また問題特有の表現や知識を保持しておく必要がある。さらに、それに付随して論理的な思考力求められる。

また、このような会話は以下のように算数文章題として文書によって表現することもできる。

(例題) 定員 20 名のバスに 5 名の団体 2 組, 7 名の団体 1 組の予約が入っている。4 名の団体で予約をすることは可能ですか？

本稿では、このような算数文章題を対象とする。常識的な知識や算数の知識を用いることにより、自然言語文章から算数文章題か否かを判断する、算数文章題と判断された文章に関して意味処理することで、その答えを求める手法を提案する。扱う算数文章題としては、小学校で習う程度の算数文章題を対象とする。具体的には、加減算、乗除算、四則混合算などが挙げられる。このような算数文章題を、正しく理解することができれば、数量を含む会話文の理解などに役立つのではないかと考えられる。よって、そのような算数文章題が意味処理され、最終的に回答を導き出せるまでを研究の目標としている。

2. 研究概要

常識的な知識や算数の知識を用いてテーマ（算数文章題）の判定を行い、答えを求める一連の流れを「算数文章題の意味処理」と呼ぶ。算数文章題の意味処理のイメージを図 2 に示す。

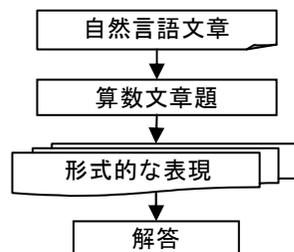


図 2 算数文章題の意味処理のイメージ

辞書や常識知識を用いて、自然言語文章から、「日本でお酒が飲める年齢は何歳

か？」のような算数文章題ではない文章を判別する。算数文章題であると判断されたものに関しては、文章全体を演算処理するために必要となる情報を整理した形式的な表現に直す。そして、常識知識の 1 つである算数知識などを用いて解答を求める。

3. 関連技術

3.1 概念ベース

概念ベース[1]とは、語（概念）と意味（属性）のセットを蓄積している国語辞書等から自動構築された知識ベースである。ある単語 A をその単語と関連の強いと考えられる単語 a_i と重み w_i の対の集合として定義する(式 1)。

$$A = \{(a_1, w_1), (a_2, w_2), \dots, (a_m, w_m)\} \quad (1)$$

ここで、 A を概念表記、 a_i を一次属性と呼ぶ。このような（属性、重み）の対の集合を大量に集めたものを概念ベースと呼ぶ。ただし、任意の一次属性 a_i は、その概念ベース中のある概念表記の集合に含まれているものとする。従って、一次属性は必ずある概念表記に一致するため、さらにその一次属性を抽出することができる。これを A に対する二次属性と呼ぶ。概念ベースにおいて、「概念」は n までの属性の連鎖集合により定義されている。

3.2 関連度計算方式

関連度[2]とは、定義された知識の関連しか特定できない意味ネットワークのようなものとは違い、任意の概念と概念の関連の強さを定量的に評価するものである。具体的には概念連鎖により概念を 2 次属性まで展開したところで、最も対応の良い 1 次属性同士を対応付け、それらの一致する属性個数を評価することにより、関連の強さを定量化するものである。関連度は、概念間の関連の強さを 0 と 1 の間の実数値で表す。表 1 に例を示す。

表 1 関連度計算方式の例

概念 A	概念 B	概念 A と概念 B の関連度
自動車	車	0.919
	馬	0.031

3.3 シソーラス

シソーラス[3]とは、広く同義語、類義語を整理したもので、約 2700 の意味属性の上下位関係、全体部分関係が木構造で示されたものである。約 13 万語が登録されており、親子・兄弟関係を持つ語についてはその関係も記述されている。例えば、「酒」と「ビール」は親子関係、「ビール」と「ウイスキー」は兄弟関係である。シソーラスの親とはシソーラス上で上位（親）に当たる関係のものを言う。

3.4 意味理解システム

意味理解システム[4]とは、単文入力に対して [主体], [何], [時間], [場所], [方法], [理由], [誰に], [用言] という8個の要素で構成されるフレームに分けることができるシステムである。図3に「昨日、本棚に1冊260gの本が20冊あった」という例を用いたフレーム分けの例を示す。

主体	何	時間	場所	方法	理由	誰に	用言
1冊260g の本		昨日	本棚				20冊ある

図3 意味理解システムの例

3.5 質問文意味理解システム

質問文が何を求め、そこにどのような条件が含まれているかを知ることができれば、質問文の内容を理解できたと言える。質問文意味理解システム[5]に質問文を入力すると、質問対象語（質問文が求めている対象）とその尺度（質問対象語「価格」の場合は尺度「通貨の価値」など）、その条件（質問対象語にかかっている条件）を抽出することができる。例えば、「昨日買った林檎は何個あるか?」という質問文を入力すると、質問対象語は「個数」で、尺度は「数量」、条件は「昨日買った」となる。

4. 算数文章題の意味処理

4.1 算数文章題の定義

本稿では、「問題文にある概念の数量を問う質問文を1つ含み、答えを求めるために演算処理を行う必要がある問題」を算数文章題と定義する。また、算数文章題には文章の内容または、解答方法によって様々な種類が存在している。本稿では、その種類を大まかに「一般問題」と「特殊問題」の二つに整理し、意味処理を行う。

4.1.1 一般問題

一般問題とは、小学校1～6年生で習う「加減算」、「乗除算」、「四則混合算」を用いて解くことができ、特殊な公式などを用いない問題である。以下にその一例を示す。
例①「林檎が3個ある。蜜柑が5個ある。果物は全部で何個あるか?」（加減算）
例②「1個75円のケーキを8個買う。ケーキの代金はいくらか?」（乗除算）
例③「1個40円の蜜柑を4個買った。さらに1本30円のバナナを5本買った。果物の代金は全部でいくらになるか?」（四則混合算）

4.1.2 特殊問題

特殊問題とは、主に一般常識問題であるSPI (Synthetic Personality Inventory) 問題などに見られる問題文の表現が特殊な問題である。ある程度定式化されており、公式など

を用いて解くことができる問題である。具体的には、「速度算」、「濃度算」、「数列算」、「N進数算」、「仕事算」などがあげられる。以下にその一例を示す。

例①「4, 7, 10, 13, 16...という数列がある。第50項はいくつ?」（数列算）

例②「2進法の110011は10進法ではいくつになるか?」（N進数算）

例③「時速100kmの電車は3時間40分間で何km進みますか?」（速度算）

例④「食塩20gを水100に溶かすと何%の食塩水が出来ますか?」（濃度算）

本稿では、一般問題に関してのみ演算処理を行うので、特殊問題の演算処理に関しては対象外とする。

4.2 算数文章題の意味処理の流れ

本研究では、算数の知識を集めた知識ベースや、辞書である概念ベース、シソーラス、などを用いて、算数文章題の意味を理解することを目的としている。また、構文解析などを行わずに、演算対象となる概念や数量を表す語の位置関係など、算数文章題の文章の傾向に着目した手法を考案することで極めて多種多様な表現に対応することを目指している。算数文章題の理解の流れを図4に示す。

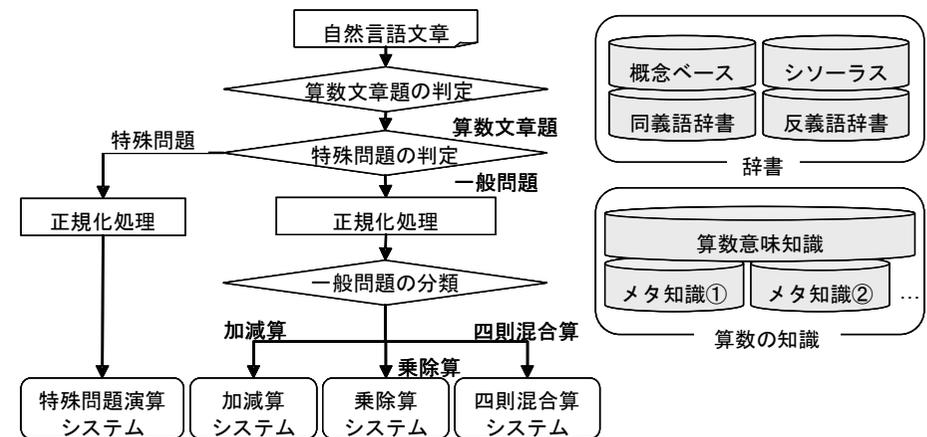


図4 算数文章題の意味処理流れ

まずは、入力文が算数文章題であるか否かの判断を行う。この処理を「算数文章題の判定」と呼ぶ。また、この処理は「質問文の意味理解を用いた判断」、「数量を表す語の有無の判断」、「単位尺度による判断」の3つの流れで構成されている。

次に、「算数文章題の判定」で算数文章題と判定された問題は「算数文章題の分類」を行う。この処理では、算数文章題が4.1節の「一般問題」か「特殊問題」のどちらの

文章題であるのかを判断している。

最後に、分類された算数文章題は、演算しやすいように正規化処理を行う。正規化処理された問題は算数知識を用いて演算処理を行い、質問文に適した解答を導き出す。この一連の流れを「算数文章題の意味処理」と呼び、次の章より詳しく説明を行っていく。

5. 算数文章題の判定

入力される自然言語文章は算数に関する文章のみとは限らない。そこで、入力された自然言語文章が算数に関する文章であるかどうかを質問文の意味理解によって判断し、さらに単位などのチェックを行うことにより算数文章題であるか否かの判断を行う。算数問題の判定の流れを図5に示す。

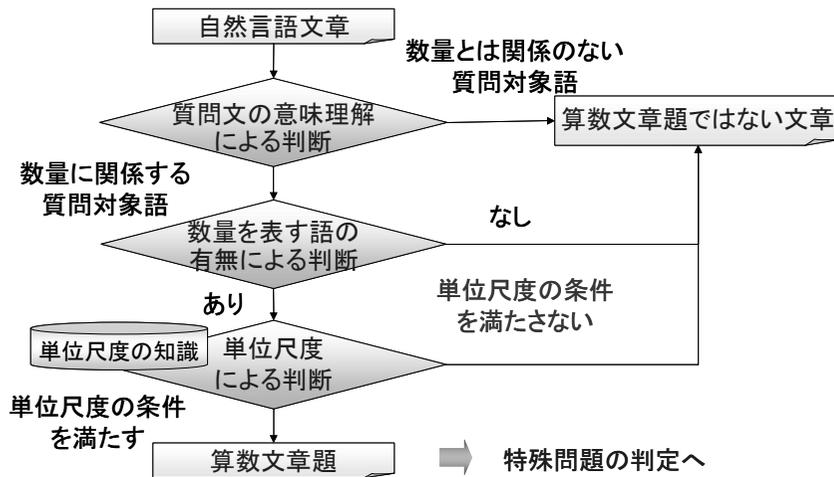


図5 算数文章題の判定の流れ

算数文章題の判定は、「質問文の意味理解による判断」、「数量を表す語の有無による判断」、「単位尺度による判断」の3段階の判断で構成される。これらの判断を経て、入力文である自然言語文章を「算数文章題ではない文章」、「算数文章題」に分類する。

次に、「質問文の意味理解による判断」、「数量を表す語の有無による判断」、「単位尺度による判断」の各処理について説明する。

5.1 質問文の意味理解による判断

入力文中における質問文の内容に着目し、算数文章題ではない文章か否かの判断を行う。判断の際に、関連技術である質問文意味理解システムを用いている。質問文に着目し、質問文意味理解システムにより得られた質問対象語が、「長さ」、「質量」（尺度はそれぞれ「数量」）などの、数量に関係するものである場合を考える。その場合は、数量についての質問であると考えられるため、この段階では算数文章題であると判断し、次の数量を表す語の有無による判断を行う。次に、質問対象語が、「人物」、「場所」などの数量に関係しないものである場合を考える。この場合、数量とは関係のない質問であると考えられ、4.1 節で述べた算数文章題の定義を満たさないため、算数文章題ではない文章であると判断する。判断例を以下に示す。

例①「国語の本の値段は何円か？」 → 質問対象語「値段」

この問題での質問対象語「値段」は数量に関係するため、質問文は算数文章題であると判断する。

例②「果物はどこで買ったか？」 → 質問対象語「場所」

この問題での質問対象語「場所」は数量に関係しないため、質問文は算数文章題ではない文章であると判断する。

5.2 数量を表す語の有無による判断

ここでは、入力文中に「3個」、「100円」などの数量を表す語が含まれるかどうかを調べることにより、算数文章題ではない文章か否かの判断を行う。判断例を以下に示す。

例①「昨日、算数の本を買った。

算数の本の値段はいくらであったか？」 → 数量を表す語「なし」

数量についての質問であるが、入力文中に数量を表す語が1つも含まれないため、算数問題ではない文章であると判断する。

例②「林檎が3個ある。蜜柑が5個ある。果物は全部で何個あるか？」 → 数量を表す語「あり」

数量についての質問であり、入力文中に数量を表す語「3個」、「5個」が含まれるため、算数文章題に関する文章であると判断する。

5.3 単位尺度による判断

この処理を行う前に、「質問文の意味理解による算数文章題か否かの判断」と「数量を表す語の有無による算数文章題か否かの判断」が行われる。よって、入力文は数量についての質問を含み、かつ数量を表す語を含む文章であると考えられる。ここで、「1.5kgは何mか？（数量を表す語「あり）」という文を考える。これは数量についての質問であり、入力文中に数量を表す語が含まれるため、算数文章題であると判断される。しかし、「尺度の異なる単位の変換」に関する質問であり、算数文章題であるとは言えない。このように、判断に「単位の知識」を必要とする入力文に関しては、

単位尺度による算数文章題か否かの判断を行う必要がある。ここでは、単位・単位尺度の観点から、入力文が算数文章題か否かの判断を行う。判断知識として単位とその意味を格納した「単位尺度の知識」を用いている。

例①「1個40円の蜜柑を4個買った。さらに1本30円のバナナを5本買った。果物の値段は何円になるか？」

例①では、質問文で問われている単位「円（尺度：値段）」と同じ尺度である単位を含む語が、質問文以外にも存在する。具体的には、数量を表す語「40円」、「30円」である。このことから、解答を求めるためには、質問文で問われている単位の尺度と同じ尺度の単位が、質問文以外にも必要になると推測できる。しかし、以下の例②では質問文で問われている単位の尺度と同じ尺度を用いている単位を含む語が、質問文以外に存在しない。

例②「食塩60gを180gの水に溶かす。この時、何%の食塩水ができるか？」

これは、単位尺度の関連性を考慮しなければならないと考えられる。図6に単位の関連性のイメージを示す。

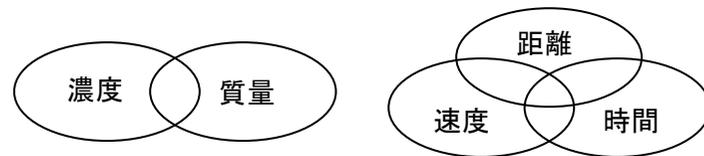


図6 単位尺度の関連性のイメージ図

図6の様に、単位には様々な関連性があると考えられる。「g（尺度：質量）」と「%（尺度：濃度）」には単位尺度の関連があると考えられる。この単位尺度の関連をまとめたものである「単位尺度の関連の知識」を作成した。単位尺度の関連は他に「長さ、時間、速さ」、「長さ、面積」などがあり、[長さ/時間/速さ]、[長さ/面積]という形式で、計10パターン登録している。

ここでは以下の条件のいずれかを満たすものを算数文章題と判断する。

I. 質問文の単位の尺度と同じ尺度の単位を含む語が、質問文以外に少なくとも1つ存在する

例①「1個40円の蜜柑を4個買った。さらに1本30円のバナナを5本買った。果物の値段は何円になるか？」（単位「円」が2つ存在する）

II. 入力文中に存在するそれぞれの単位の尺度が、単位尺度の関連を満たしている

例②「食塩60gを180gの水に溶かす。この時、何%の食塩水ができるか？」（単位尺度の関連「質量、濃度」を満たしている）

I, IIのいずれの条件も満たさないものに関しては、「算数文章題ではない文章」と判断する。また、この段階で、数量を含むが単位が存在しない文章に関しては、単位尺度による判断を行わずに、算数文章題として判断する。これは、数列算やN進数算などの問題によく見られる傾向である。

6. 算数文章題の分類

算数文章題は様々な表現で構成されるが、4.1.2節で述べた数列算やN進数算などの特殊問題においては、一般問題とは異なり、特徴的な表現が多いという傾向がある。よって、7章で述べる正規化を行いやすくするために、特徴的な表現が多い特殊問題を判別する。判定された特殊問題に関しては、本稿では演算処理を行わない。

分類方法は、算数文章題と判定された入力文に対して、まずは形態素解析を行う。そして、形態素解析した単語一つ一つに対して特殊問題判定知識に格納されているキーワードと関連度計算を行う。図7に例を示す。

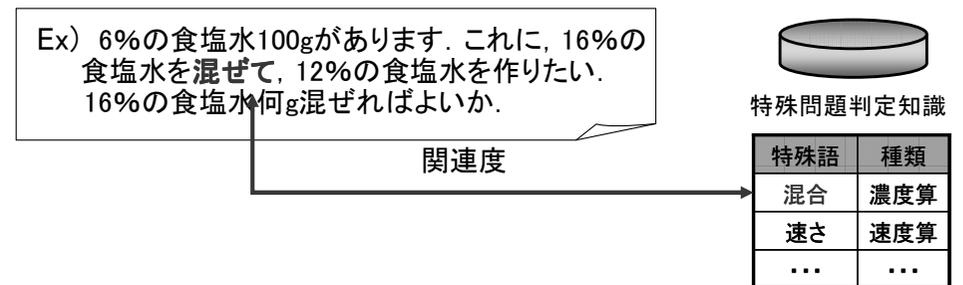


図7 算数文章題の分類のイメージ

図7の様に、濃度算のキーワードと文中の語の関連度が閾値よりも高ければ、濃度算と判定する。また、数列算とN進数算は、このようなキーワードだけでは分類できない。数列算では「3, 7, 11, 15, ...」, N進数算では「1101001」のような数字の羅列部分が問題文中に存在している。そこで、この二つの問題に関しては、あらかじめキーワードによる分類の前に数字の羅列部分による分類を行っておく必要がある。

このように、算数文章題と判定された問題に対して、その問題文中に数字の羅列部分や特殊問題判定知識の語が含まれる場合は、特殊問題であると判断する。

7. 算数文章題の演算処理

算数文章題の解答を求めるためには、算数文章題であると正しく判断された入力文から演算に必要な情報を抽出する必要がある。解答を求める演算に必要な問

題文中の様々な情報を「整理した形式的な表現に直すこと」を算数文章題における正規化と定義する。本稿の正規化では、「言葉の意味」、「文法」、「文脈」を理解することができるような正規化表現を提案する。ここでは関連技術として、意味理解システムを用いている。

7.1 正規化処理

7.1.1 正規化表現の定義

演算を行うためには問題文から様々な情報を抽出する必要がある。抽出した情報をまとめたものが正規化表現である。正規化表現の項目に関して、図8に示す。

演算対象語	修飾語と明示語				数量	用言
	連体修飾語	時間語	場所語	…		

図8 正規化表現の項目

問題文から、正規化表現を取得する際の規則を以下に述べる。

1. 「演算対象語＝数量」の関係になるようにする。(数量は必ず1つ)
2. 基本的に演算対象語は1つ。
3. 演算対象語が1つに切り離せない場合は、複数の演算対象語の関係を取り出す
 ※ 正規化表現の関連が強いと思われるものに関しては、「関係付け」を行う
 「関係付け」とは、ある2つの正規化表現が、乗除算の演算で必要となる比の関係 ($A1 : A2 = B1 : B2$) のペア ($A1 : A2$ など) であるという情報を持たせておくことである。

図10の修飾語と明示語の中の「…」には主に明示語を格納する。明示語とは、問題文中にある情報の中で、以下のような語であると定義し、「明示語の知識」を作成する。

- ① 規則3に関する正規化表現の各項目を取得する際の手掛かりとなる語 (複数明示語)
 → 「みんな」、「全部」、「より」などの『合計』や『比較』など演算処理の方法を表現する語。
- ② 正規化表現の関係付けを行う際の手掛かりとなる語 (乗除明示語)
 → 乗除算の演算で必要となる『比の関係』を表現する語。現在は「ずつ」の1語のみ。
- ③ 演算を行う際の手掛かりとなるもの (演算明示語)
 → 「合わせて」、「残る」などの『質問文に存在し、加減算のどちらの演算であるか』を示唆する語 (演算明示語)。

7.1.2 正規化表現の取得方法

問題文から正規化表現を取得する方法を説明する。

1. 問題文を一文ずつ意味理解システムにかけ、意味フレームの成分に分ける。
2. 「明示語の知識」を参照し、意味フレームの各成分に格納された情報の中から各明示語を取得する。
3. 数量に着目し、数量とその単位を調べる (例: 「12枚」→数量「12枚」、単位「枚」)
 ※数量1つにつき、1つの正規化表現を取得する。(規則1より)
4. [主体]、[何]、[方法]、[理由]、[用言]の各成分中の名詞で、演算対象語の知識(「定価」や「代」など)を取り出す。その知識に存在しなければ、具体物である名詞を演算対象語の候補とし、持ちうる単位を単位知識ベースより調べる。(例: 「色紙」→「数量:枚 | 質量:g | 長さ:m…」)
5. 4で調べた単位が3の単位と一致したものを演算対象語とする
6. 各修飾語(「連体修飾語」、「時間語」、「場所語」など)を取得する
 図9に「今月の水道代は2500円です」という文を用いて、正規化表現の例を示す。

演算対象語	修飾語と明示語				数量	用言
	連体修飾語	時間語	場所語	…		
水道代		今月の			2500円	です

図9 正規化表現の例

演算対象語知識から「代」を抽出し、その前に接続している「水道」と合わせて抜き出す。「代」に関しての単位知識として「円」がデータベースに格納されている。よって、「水道代＝2500円」のように演算対象語と数値を抜き出すことができる。また、演算対象語の修飾語として「今月の」が抽出される。これは、「今月」の部分が時間に関係する事柄であるということが意味理解システムよりわかるので、時間語の項目に分けられる。このように、数値一つ一つについて正規化処理を行っていくことで、演算処理により解答を求めていく。

7.2 演算規則

加減算、乗除算、四則混合算の一般問題の解答を、演算に関係すると判断された正規化表現の情報をもとに求める。各演算システムの演算手法を以下に述べる。

7.2.1 加減算の演算手法

加減算は、「林檎が5個ある。メロンを3個買う。果物は全部で何個になるか?」のように状態が変化することが特徴として挙げられる。この状態の変化に対応できるような手法を考案する必要がある。状態の変化は演算明示語により、判断を行っている。その判断方法に関して、「飴が12個ある。飴を5個食べた。残りの飴は何個あるか?」

の例を用いた図 10 によって説明を行う。

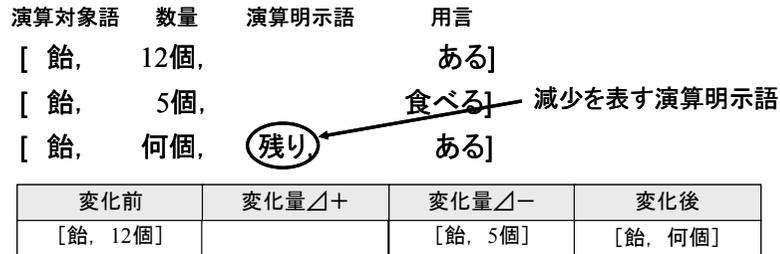


図 10 加減算の演算処理例

正規化表現 [飴, 12 個, ある] は [変化前] に格納される ([変化前] を示唆する演算明示語「ある」より)。質問部分の正規化表現 [飴, 何個, 残り, ある] は [変化後] に格納される ([変化後] を示唆する演算明示語「残り」より)。また、この質問部分の正規化表現にある演算明示語「残り」は、減算 (△-) を示唆する演算明示語でもあるので、正規化表現 [飴, 5 個, 食べる] は [変化量△-] に格納される。よって、演算を行うことで「7 個」という解答が求められる。

7.2.2 乗除算の演算手法

一般問題の分類により、すでに比の関係となる正規化表現のペアは取得できている。比の関係より、以下の数式が成立する。

$$a_1 b_2 = a_2 b_1 \quad (a_1, a_2 : \text{一方の関係付けた正規化表現の数量}, b_1, b_2 : \text{他方の} \textit{''})$$

この数式により解答を求めることができる。「1 冊 260g の本が 20 冊ある。本の全体の重さは何 g あるか?」という文を用いた例を図 11 に示す。

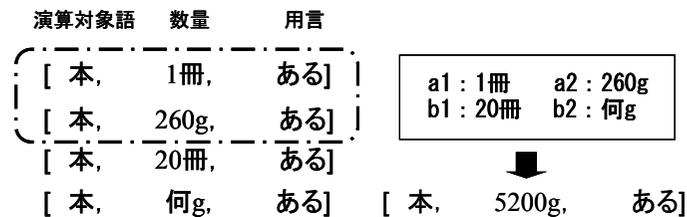


図 11 乗除算の演算処理例

正規化処理を行うことで、「1 冊の本」と「260g の本」がペアになっていることがわかる。また、「20 冊の本」という正規化表現の情報より乗除算の演算処理であると

ということがわかる。よって、質問文の正規化表現により「何 g」と重さを聞いているので、「5200g」と演算処理することができる。

7.2.3 四則混合算の演算手法

四則混合算システムは、加減算システム、乗除算システムの各演算手法を組み合わせることにより実現できると考えられる。基本的な考え方は、乗除算部分の演算を先に行い、その結果を用いて加減算を行うというものである。例を図 12 に示す。

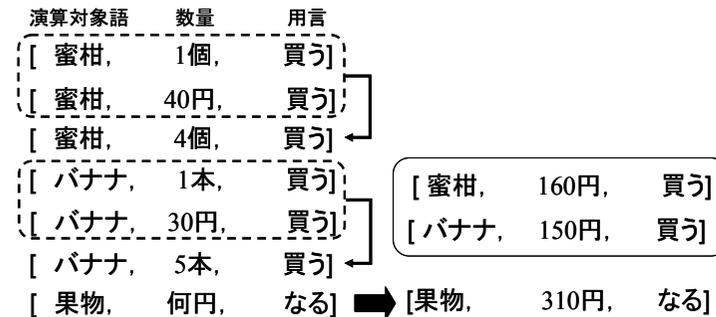


図 12 四則混合算の演算処理例

関係付けた正規化表現は図の点線に囲まれた部分であり、その上側は質問部分の正規化表現と [蜜柑, 4 個, 買う] と比の関係に、下側は質問部分の正規化表現と [バナナ, 5 本, 買う] と比の関係になる。この段階で乗除算の演算を行うと、正規化表現 [蜜柑, 160 円, 買う], [バナナ, 150 円, 買う] が得られる。次に、実線に囲まれた部分に着目すると、これは加減算の演算手法で述べた「状態が変化する場合」に相当するので、この場合の数式に各数量を代入することにより、解答「310 (円)」を求めることができる。

8. 評価

5 章の算数文章題の判定と 6 章の算数文章題の分類において、算数文章題であり一般問題であると判断された 100 文を対象に演算処理における表かを行った。正規化処理を正しく行なって演算処理により解答を求めることができた場合を「正解」、正しく処理できなかった場合を「不正解」として評価を行う。評価結果を図 13 に示す。

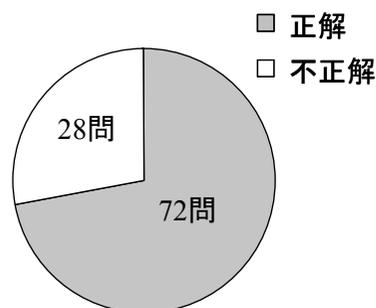


図 13 演算処理に関する評価結果

全体の精度としては、72%という結果を得ることができた。また、正解例と不正解例を以下に示す。

正解例「原価 1000 円の商品があります。これに 500 円の利益をのせました。この商品を定価の 20%引きにしました。売価は何円ですか？」

不正解例「91 人のグループ全員の血液型を調べた。B 型の人数は 18 人います。O 型の人は 30 人います。AB 型の人数は 8 人でした。グループ全員を基にした A 型の人数の割合を答えなさい？」

正解例は演算処理により解答を求めることができた。「定価」や「原価」などの表現を正しく理解でき、四則混合算として意味処理できていると考えられる。

次に、不正解例では、正規化は正しく行えるが、その後の解答を導き出すところで失敗した。その理由としては、システムが血液型に関する常識的な知識を持ってないという点が挙げられる。「血液型は A 型、B 型、O 型、AB 型の 4 種類しかない」という知識があれば、91 人から B 型と O、AB 型の人数を引けば残りの A 型の人数を求めることができる。また、そこから A 型の人数とグループ全員の人数を比較して割合を求めることができると考えられる。しかし、「血液型の知識」だけをシステムに組み込むことは簡単であるが、それをを用いて解答できる問題は限られる。よって、提案したシステムと常識的な判断が行えるシステムと連動させて用いれば、意味処理を行える問題は増えるのではないかと考えられる。

9. おわりに

本稿では、辞書や常識知識に加え、様々な関連技術を用いることにより、自然言語文章から算数文章題か否かを判断し、算数文章題と判断された問題についてはその解

答を求める手法を提案した。演算するために必要な正規化処理の手法を導入した結果、演算処理に関して 72%の精度を得ることができた。算数文章題の意味処理において、問題文から演算に必要な情報を抽出することにより、言葉の意味、数量の論理関係、文脈の理解から演算を行うことが実現できた。

参考文献

- 1) 眞鍋康人, 小島一秀, 渡部広一, 河岡司, “概念間の関連度やシソーラスを用いた概念ベースの自動精練手法”, 同志社大学理工学研究報告, Vol.42, No.1, pp.9-20, 2001.
- 2) 渡部広一, 奥村紀之, 河岡司, “概念の意味属性と共起情報を用いた関連度計算方式”, 自然言語処理, Vol.3, No.1, pp.53-74, 2006.
- 3) NTT コミュニケーション科学研究所監修, 「日本語語彙体系」, 岩波書店, 東京, 1997.
- 4) 篠原宜道, 渡部広一, 河岡司, “常識判断に基づく会話意味理解方式”, 言語処理学会第 8 回年次大会発表論文集, B6-2, pp.651-654, 2002.
- 5) 古川成道, 渡部広一, 河岡司, “概念ベースを用いた知的検索における曖昧な質問文の意味理解”, 第 18 回人工知能学会全国大会, 2D1-10, 2004.