

日本語の数字表現の意味的な分類について

5 F - 2

旭岡佳美, 平川秀樹, 天野真家
(株) 東芝 総合研究所

1. はじめに

機械翻訳などの自然言語処理システムの実用的な対象である科学技術文献の特徴の1つは、数字関連の表現が多く、かつ、重要な情報を担っていることである。日本語における数字関連の表現では、意味関係が構文的に明示されておらず、また、数字表現の表わす意味は多様である。数字表現の意味を決定するためには、どのような意味関係が存在するのかを適切に分類しなければならない。意味関係の分類という観点から考えると、数量表現の研究では、意味関係は、数量詞後置の研究との関連において議論されてきた。^[1]また、「名詞+の+名詞」の意味関係に関する研究がいくつか行なわれている。^{[2][3]}しかし、数字表現における意味関係の網羅的な分類はこれまで行なわれるに至っていない。

本稿では、数字表現においてどのような意味関係が存在するのかを、構文の面と意味の面の両方向の視点で分類する。2.では日本語の数字表現の構文上のパターンを整理する。3.では意味関係の分類について述べる。4.では実際の文書での意味分類について述べる。

2. 数字表現の構文

日本語の数字表現の構文上のパターンを[1]にならって整理すると次のようになる。

- (1) Q/N型 3冊の本を読んだ。
- (2) NQC型 ケーキ2個を食べた。
- (3) NCQ型 紙を5枚用意した。3枚紙を用意した。
- (4) N/Q型 データの3つを選んだ。
- (5) N/Qダ型 長さは3000mだ。
- (6) QN型 21インチTVを買った。
- (7) QC型 3軒を訪ねた。4mを使った。

ここでNは名詞のことであり、Qは数字+助数詞(2つまとめて数量詞と呼ぶ)のことである。Cは助詞である。(3)の一部と(6)の間には構文上の曖昧さが存在するが、他は構文的に明確である。

3. 意味関係の分類

3.1. 分類の動機

日本語の数字表現においては、いくつかの構文パターンに同種の意味関係のパターンが存在すると共に、1つの構文パターンにいくつかの意味関係のパターンが存在する。自然言語理解の立場から考えると、構文解析を終了した時点を対象として、異なった統語構造から同一の意味構造を抽出し、かつ同一の統語構造から異なった意味構造を抽出しなければならない。この処理を行うためには、まず、数字表現の構文パター

ンを整理し、それぞれの表現を意味的に分類することが必要である。以下では、各構文パターンについて意味関係を分類する。なお、図1に構文パターンと意味分類の関係をまとめている。

3.2. 意味関係の分類

今回、様々な数字表現を解析した結果、以下のような分類が得られた。はじめにQ/N型についての分類を示し、次にその他の型の分類を示す。Q/N型の表現は最も多様な意味関係を持ち、その他の型の分類の基本となる。

<Q/N型の表現>

- (1) 数量 3冊の本、3単語のデータ、200ccの水
- (2) 属性 200ccのビン、3個の人、10年の研究
- (3) 値-関数 4mmの長さ、20Vの電圧
- (4) 値のみ 30個の値段、20ccの影響
- (5) 順序 3台目の車

・数量表現と属性表現の違い：数量表現はNQC型やNCQ型でも表わすことができるが、属性表現をNQC型やNCQ型で表わすことはできない^[1]。また、属性表現では、「容積200ccのビン」、「正解3個の人」のように属性名を補った表現があるのに対し、「数3冊の本」、「量200ccの水」といった表現は通常用いられない。以上のような違いから、数量表現と属性表現を区別する。(量と属性の違いは、例えば、「200cc of water」と「a 200cc bottle」というように英語表現の違いとして現われることもある。)しかし、数量表現であっても数量の情報のみを背負っているわけではなく、例えば、

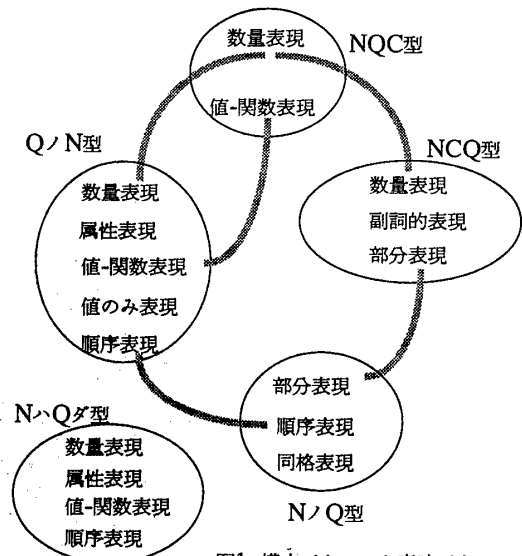


図1. 構文パターンと意味パターン

「1枚のチョコレート」という表現では、それが(恐らく)板チョコであるという情報も担っている。

・数量表現・属性表現と値-関数表現・値のみ表現との違い：前者ではQがNに修飾的に働いており、後者では関数的^[3]に働いている。(N^hQ^d型の数量・属性表現に対する翻訳では関数に当たる"number"や"length"などを補う必要がある。)

<NCQ型の表現>

- (1) 数量 本3冊、データ3単語、水200cc
- (2) 値-関数 長さ4mm、電圧20V

<NCQ型の表現>

- (1) 数量 本3冊読んだ。データを3単語選んだ。
- (2) 副詞的 本を3m移動させた。
- (3) 部分 生徒が3人欠席した。

・数量表現と副詞的表現の違い：副詞表現ではQが動詞の程度などを修飾しており、NとQの間の直接の関係は薄い。

<N/Q型の表現>

- (1) 部分 生徒の3人が欠席した。
- (2) 同格 参考書の5冊を買った。
- (3) 順序 走っている車の3台目に注目した。

<N^hQ^d型の表現>

- (1) 数量 本は3冊だ。
- (2) 属性 ビンは200ccだ。
- (3) 値-関数 電圧は200Vである。
- (4) 順序 車は3台目だ。

<QN型の表現>

- (1) 属性 21インチTV
- (2) 順序 3台目モニターテレビ

<QC型の表現>

- (1) 数量 4枚を使用した。3mを用意した。
- (2) 属性 200Vまで使用できる。
- (3) 順序 3台目を購入した。

4. 実際の文書での意味分類

4.1. 実際の文書の分析

実際の文書において本当にさまざまな表現が用いられるかという点は、上記の分類がどの程度必要であるかという点で重要である。そこで、上記の分類に基づき、実際のマニュアル文書(約1200文)の数字表現を意味的に分析した。結果を表1に示す。この分析結果から、実際に多様な構文と意味関係が表われるということが確認できる。

実際の文書での数字表現例を幾つか挙げる。

Q/N(1) 1本の走査線、32チャンネルの信号

	Q/N	QN	QC	NCQ	NCQ	N/O	N ^h O
数量	36		14	14	8		5
属性	11	13	26				3
値-関数	5						24
値のみ	2			4			
副詞					4		
部分					0	2	
同格						4	
順序	2	4	7			0	0

表1. 実際のマニュアル文書での意味関係

- Q/N(2) 4チャンネルの回路、6ビットの分解能
- Q/N(3) 30dBの増幅度、3分の休止時間
- QN(1) 3芯電源コード、9インチTVモニタ
- QC(1) 32素子、8回路、1超音波フレーム
- NQC(1) モニタ2台、IC16個
- NCQ(1) 素子を1個ずらす。回路を16系統設ける。
- NCQ(2) メータが0.1mA振れる。
- N/Q(2) ブローブの128素子

また、筆者が読んで意味を決定できなかった例もある。

- Q/N(1) or Q/N(2) 1ブロックの回路
- Q/N(5) or Q/N(4) 4段目のゲイン
- QN(1) or QC(1) 1受信レート

4.2. 実際の文書での意味関係の解析の問題点

数量表現の理解では語彙の情報、特に語の使い方に関する知識が解析の基本となる。例えば「1本の走査線」が"数"表現であることは、「線」は"本"と数えるという言葉の使い方に関する知識を使って理解できる。しかし、実際のデータを人間が解釈するには語彙知識以外の知識が決め手となっている場合も多い。

(1) 背景知識から推論する必要がある表現の例

- 「4チャンネルの回路」
- 「3芯電源コード」

「4チャンネルの回路」が"属性"表現であることは、回路には出力があるという背景知識と、出力は"チャンネル"と数えるという言葉の知識を使って理解できる。「3芯電源コード」が"属性"表現であるということは、コードには芯があるという背景知識を使って理解できる。システムはこれらの知識を持つ、または推論する必要がある。

(2) 文脈中の他の記述により理解する必要がある表現の例

- 「16チャンネルのecho信号、ECHO1~16」
- 「128素子のブローブのうち38素子を選択する。ブローブの128素子は～」

「16チャンネルの信号、ECHO1~16」が"数"表現であることは、「ECHO1~16」という表現を見てはじめて理解できる。「ブローブの128素子は～」が同格表現であることは、その前の文章がなければ理解できない。

以上のように数字表現を理解するには、語彙知識の他に背景知識や文脈知識が利用されている。これらの知識は単独で用いられるばかりでなく、複合的にも用いられる。

5. おわりに

以上、数字表現の意味関係を解析するための意味分類を提案した。今後は、語彙知識を整理すると共に、数字表現の意味の表現形式や後置数量詞が埋める格の問題などここで述べなかった問題、また、文中の知識を使った解析の問題など、数字表現の意味解析方法について研究をすすめる予定である。

参考文献

- [1] 奥津 敬一郎: "数量詞移動再論", 都立大学人文学報, 第160号, 1983, 1-24.
- [2] 島津 明, 内藤 昭三, 野村 浩郷: "日本語文意味構造の分類-名詞句構造を中心に-", 情報処理学会自然言語処理研究会 47-4, 1985.
- [3] 平井 誠: "日本語文における「の」と連体修飾の分類と解析", 情報処理学会自然言語処理研究会 58-1, 1986.