

名詞の自動処理における助数詞の有効性について

BLIN RAOUL

E. H. E. S. S.

blin@msh-paris.fr

本研究の目的は名詞とそれに対応する助数詞の組を、膨大なコーパスを用いた自動分析により、リストアップする方法を考察することである。また、その結果を利用して動詞の項として働く名詞句の意味素性を自動的に決定出来るのではないかを考察する。意味素性を決定するために助数詞を利用する理由は、助数詞の自動抽出は容易であり、助数詞が対応する名詞の意味を表すので、その意味を表層で捉えやすいからである。

本稿では、まず、名詞や動詞を分類する上で、助数詞がどのような役割を持ちえるかを検討する。筆者は膨大なコーパスでの作業は行わず、限られたコーパスのみ使用したが、その自動処理中に現れた問題点をまとめてみる。最後に、助数詞を用いて名詞(句)の意味素性を自動的に決定することの限界について述べる。

以下で使用する専門用語を確認する。助数詞は「人、個、...」、個体数量詞は「数詞+助数詞」、単名詞は体言、複合名詞は単名詞をノ、トなどで結んだものである。「三役ら幹部」のように幾つかの並列したものも複合名詞と呼ぶ。単名詞と複合名詞を名詞と呼ぶ。名詞句は名詞に格助詞を付けたものとみなす。また、ここでは、個体数量詞しか扱わず、(1センチ、キロ...)容量数量詞は考察しない。

1- 目的

文献を参照した限りでは、名詞全てを対象としてどの助数詞が対応するのかを示した辞典やリストは存在しない。現在、市場に出版されている一般辞典はそういう項目はないが部分的に記入されている(例えば:新明解国語辞典で「子ども」を数える助数詞は記入されている一方、「息子」に対応する助数詞は記載されていない)。しかし、テキストを自動修正する時や文章を生成する時に、このデータが重要で不可欠である。さらに、この名詞と助数詞の組をリスト化することで名詞句の解析においても、かなりのパーセンテージ問題を解決できる。BLIN(1996, 訂正版を参照)では、名詞句に対応する助数詞に関する知識をその名詞句の文法的な特徴とみなし、それを利用して先行詞を抽出する方法を提案した。同様に、例えば「5人の学生が受験をしたが、3人が合格、2人が不合格だった」の「3人が」のように(名詞が省略されて)数量詞のみが名詞句に現れる場合の処理においても、この文法的特徴を使用できる。

さらに、助数詞を使用することによって、名詞の意味素性も、ある程度まで決定できるのではないと思われる。現在、どういった統語論的立場からみても、動詞の特徴をまとめようとする時、動詞の特徴としてその項目である名詞句の意味素性(人間・非人間、アニメート・非アニメートなど)が含まれる。IPAL動詞辞書などがその例である。しかし、名詞句の意味素性を定めるために有効な統語的方法がないので、その作業は手作業となり、しかも直感的である。つまり、使用者が自分の知識と感覚で意味素性を決めるが、一般知識の上で考えられた現実と自然言語が表す現実とは必ずしも対応しない。したがって、意味素性を定義するには、より言語的な判断が求められている。そこで、助数詞を使用して名詞句と動詞句を分類することが可能ではないかと考えられるのである。その理由としては、助数詞を使用することで統語論的な分類を行うことができることがある。

しかし、助数詞は百個以上、名詞は数万あり、さらに、名詞1個には助数詞1個しか対応しないとは限らないので、なんらかの方法でそれぞれの名詞や名詞句に付きうる助数詞を自動的に検索することが出来れば大変な手作業が避けられる。助数詞は数の限られた、比較的簡単な構文で現れるので、このような自動的処理が可能となると思われる。以下、その手続きを考察したい。

ここで著者の立場を簡単にまとめると次のようになる。本研究の目的は、できるだけ多くの名詞を検討し、その名詞に対応する助数詞の情報をまとめることである。その方法としては、膨大なコーパスを使用し、名詞と付属助数詞が共起する文構文を見つけ、その名詞と付属助数詞を辞書に記入する。以下論議するように、修正の手作業が必要となることもあり得るが、この手作業を最小限にするためには、文構造の選択が大変重要になる。

2- 対象の名詞句

日本語の助数詞に関して、言語学、自然言語処理学の分野での文献が多数あるので（末尾文献参照）、ここでは詳細な説明はしない。本研究では、全ての名詞句と助数詞が既に確認され、一覧表にされていることを前提とする。

奥津（1969）が次のように個体数量詞をまとめた：

- a) NQC型 本3冊を読んだ
- b) QのNC型 3冊の本を読んだ
- c) NCQ型 本を3冊読んだ
3枚紙を用意した
- d) NのQC型 本の3冊を読んだ

（Q：数詞＋助数詞、N：名詞句、C：格助詞）

上のそれぞれの構文によって、自動検索上で成功の効率がかなり異なるので、本研究は「QのNC」型と「NQC」型に限定する。（c）は例えば名詞と助数詞との間に他の語句が入った場合、文章の統合パーズンが困難になる可能性がある。（d）は依存関係を表すので不正な関連を招く可能性が高い。例えば：

- (1) 彼の本の一冊...

では、もちろん「冊」で「本」を数えるが、

- (2) ホテルの一室に
- (3) 東富士演習場地域農民再建連盟の十カ所に

「室」ではホテル自体は数えられない。格助詞の付いていない「名＋助数詞」は、時制句が多いため今回は対象外とした。

「NQC」と「QのNC」構造にはいくつかの変化がある。主に、助数詞の前後に接頭語（4、5）、名詞（6、7）、数量名詞（8、9）が入るものである。

- (4) 市民<約>310人から
- (5) <約>310人の市民
- (6) 高校生ら1000人<以上>が
- (7) 1000人<以上>の高校生
- (8) 前職14人<全員>が
- (9) 14人<全員>の前職が

このような助数詞のリストとそれぞれの現れる位置を知れば、自動処理上特に不正な組み合わせを起こさないだろう。また、個体量数詞には数詞の代わりに「数」が現れうる。

3- 検索中の失敗

自動検索が次のように行なわれるのが理想である。例えば、「庭に、リンゴ3個が落ちている」を分析する時、まず数詞を検索して、その右に助数詞と格助詞があるかを確認する。ない場合は、また「Q/NC」ではないと断定すれば、この名詞句の分析を中止する。助数詞と格助詞がある場合、この名詞句が「NQC」型だと断定し、数詞の左にある名詞を抽出する。この段階で、検索を単独の名詞に限るか修飾名詞まで行うかは、パーザーの精密さによって決まる。基本的に、研究は膨大なコーパスの上で行われるので、処理を単名詞に限定しても問題はないはずである。さらに、以下で論ずるように、修飾名詞の研究で誤解が生じ易いので、出来れば単名詞のみを研究する方が無難であろう。以上のプロセスで名詞と助数詞の組み合わせ抽出作業が終了したとする。この段階で、「リンゴ」を「個」で数えると自動的に分かる。その後、助数詞と意味素性をどう結ぶかは、特に、意味素性の数は研究者によってかなり見解が異なるので、その定義や意味素性と助数詞との組み合わせ方も異なるだろう。本稿では、詳しいことを述べないが、以下、その組み合わせを決定するとき注意すべき点をいくつか挙げる。

ここでは、限られたコーパスの分析の際に現れたいくつかの不正な結果を紹介し、その原因を考えてみる。最も注目したいのは、複合名詞である。まず、「NQC」の結果をみる。

ノ格によって修飾された名詞句の場合は、それを特定する助数詞がどの名詞に対応するかに関して、問題がある。次の例を見られたい：

- (10) 本社の車1台を

(11) おもちゃの車1個を

(10)は、主名詞「車」に対応していると考えられるが(11)は、「車」＝「一個」という対応を考えるよりは、「おもちゃの車」＝「1個」というように、助数詞が名詞句全体に対応すると考えた方が妥当であると思われる。より詳細な助数詞の付き方について、さらに意味論的な研究が必要であろう。

並列の場合は、助数詞は名詞句の主名詞（幹部）と修飾名詞（三役ら）、両方に対応すると断定出来そうである：

(12) 三役ら幹部10数人と2時間近く話し合う。

以下見るように、主名詞が属性を表すと検索が失敗しやすいのだが、並列の場合は違うようである。

[N1トN2ト..トNmQC]のような単純な「名詞＋ト格」句の列挙に基づいた名詞句の後ろにある助数詞は、基本的に右の単名詞に対応する。

(13) 犬と子ども1人が...

(14) さらに議員と秘書2人が...

ただし、(14)の場合は、「議員1人と秘書1人が」という解釈を排除すべきかどうかは決定できない。しかし、この解釈は特に作文で許容度が低いため、無視する方がよいであろう。

「など」を使用した名詞句では、助数詞は名詞句の最後（家電）とその他の名詞（洗濯機、炊飯器...）に対応する：

(15) 洗濯機、炊飯器などの家電4台が...

最後に、誤解を招きやすく注意を要するいくつかの名詞句や表現がある。例えば：

(16) アトラクタ五輪を

しかし、こういった構文は既に知られているので、検索中無視されるだろう。

以上述べた失敗の多くは、一番安全な解釈を選択することにより、また処理できない構文を無視することにより、解決可能である。しかし、以下紹介するものは、統語論の面だけでは解決が困難である。特に集合名詞の場合、Nは集合を示し、Qはその意味素性を数える。このように依存関係が表れる可能性があるので、必ずしも助数詞が名詞に対応しない：

(17) 人口200万人と

(18) 家族3人が

一方、辞書で集合名詞がその他の名詞と区別されていたら、単に集合名詞を含んだ名詞句を無視することができる。しかし、重要な情報を失う可能性がある：

(19) 家族3世帯が

このような構造で、名詞が集合名詞であるのか単なる個体名詞であるのか解釈するには、意味論的な分析がどうしても必要となるであろう。

「QノNC」は、「など」に基づいた名詞句を除くと、複合名詞はかなり扱いが困難である。

(19) 3匹の犬、牛、子豚、などの動物が泳いでいた

(20) 3匹の子豚と犬が

(21) 3人の子のおもちゃが

(22) 3人のおばさんの兄弟が

(23) 3個のビスケットの車が

(24) 3台のビスケットの車が

「ト」接続の場合の例では「3匹」で、子豚しか数えないか、あるいは子豚と犬を含めて動物を数えるかは表層では決められない。

「ノ」の場合は、一番左（21）あるいは（一番右にある）主名詞（24）の名詞を数えるか、名詞句全体（23、24）を数える。また、（22）の解釈があいまいで、一番左か右を数えられる。

4- 助数詞と名詞の意味素性

作業中に表れた問題点をまとめた。以下、これらの問題点を解決したとしても、助数詞の特徴に起因すると思われる問題を考察する。

意味素性と助数詞の関係は、必ずしも単純なものではない。しかも、それぞれの助数詞によってかなり差がある。例えば「本」は、対象の名詞は「ホームラン」、「論文」、「棒」など、かなり異なった性質の対象に対応することが出来るが「人」で数えられるものは、人間であるということがあげられる。つまり、動詞の項としてかどうかは別にして、名詞を分類するために、その助数詞の多義性を考慮せざるを得ない。「人」が対応する名詞に確実に[+humain]の意味素性を特徴づけられるが[本]の対応する名詞に[+concret]などはつけられない。ただ[-humain]が対応するのは確かである。つまり、助数詞と意味素性との関連を慎重に扱わなくてはならないのである。助数詞を選ぶ際にいくつかの可能性があれば、どれにするか意味を考慮に入れた選択方法が必要である。

また、名詞の意味以外にも助数詞の選択に影響しうる要因を考慮する必要がある。場合によって、文脈も考慮することが必要である。それは文脈によって、助数詞の有無が決まる可能性があるからである。例えば、普段「車」を「台」で数えるが、車のおもちゃや車形のビスケットであれば「個」でも可能で、場合によってはその方がふさわしいこともある。

(25) 子供達がいろんな形のビスケットを数えながら食べていた。車1個、トラック3個...

さらに、助数詞の使用は話し手の社会環境にも左右されている。例えば、技術専門用語として使用され、一般には使用されていないような助数詞があるであろう。さらに、時代的にも、使用に変更が見られる。例えば明治の小説では、「個」で人間を数えられるが現在は不適当になった（岡本1996）。助数詞の使用法だけでなく、助数詞句にさえ変化がみられる。小説「坊っちゃん」ではNQC構文が現れる頻度は他の構文に比較して低いのに対し（沖久1986）、本研究での新聞記事を中心としたコーパスでは「QのN」とほぼ同じ頻度で現れた。

これらのことは、文章の生成で助数詞を選択する際注意すべき点である。

終わりに

名詞とそれに対応する助数詞の組を自動的にリストアップすることは可能であるが、意味的分析を行わずに名詞の意味素性を細かく決定することは、本稿で見たように困難である。しかし、意味素性が少ないほど、動詞の項目としての名詞の分析が有効であるという環境（M. Gross 1975）では、このような大ざっぱな分類でも十分であるかも知れない。

参考文献

- 岡本勲、近代作家の数詞・助数詞、「日本語学」5-1、明治書院、1996年
沖久雄、数詞・助数詞の文法、「日本語学」5-1、明治書院、1986年
奥津敬一郎、数量詞的表現の文法、「日本語教育」14、1969年
北原博雄、連用用法における個体数量詞と内容数量詞、「国語学」183、1996年9月
新明解国語辞典4版、1989年
R. BLIN、日本語の名詞句における照応計算のモデル、電子情報通信学会技術研究報告、1997年7月
M. GROSS, Methodes en syntaxe (統合研究方法), Hermann, 1975 (フランス語)
I. TAMBA, Specificatifs numeraux, Unite et pluralite en japonais et expression du nombre en francais, (日本語の助数詞、単位、複数、とフランス語で数える)「C. R. L. A. O」XXI, 1992. (フランス語)