

物語文章の要約化処理について

目本俊亮 阿部純一

北海道大学文学部行動科学科

文章(テキスト)の要約化とは、文章中に内在する多くの命題情報の中から互いに意味的に集約できるものを探し出し、それらをよりマクロな命題に変換していく処理過程といえる。本稿では、対象とする文章を“物語文章”とし、その要約化の際に用いられる知識、すなわち、要約化の規則について提案する。さらに、その規則を簡単な物語に適用して、物語文章の意味は様々な抽象度のレベルで表象され得ることを示す。また、要約化処理の結果として得られる心内の意味表象と実際に産出される要約文章の違いについて考察する。

SUMMARIZING PROCESS FOR NARRATIVE TEXT

Toshiaki Muramoto Jun-ichi Abe

Department of Behavioral Science, Faculty of Letters, Hokkaido University,
N-10, W-7, Kita-ku, Sapporo 060, Japan

In summarizing a narrative text, we integrated underlying propositions into fewer macro-propositions. In this paper, we propose the knowledge structure used in the process of summarization of a narrative text, as 'summarizing rules'. Applying the rules to a simple story, we illustrate that a meaning of the story can be represented in various levels of abstraction. We also discuss the difference between the underlying meaning and the summary text itself which is really produced.

1. はじめに

自然言語理解に関わる問題の1つとして、“文章の要約化”がある。次の4つの文章を見てみよう。

- (1) あるところにきつねとくまがいました。彼らは友達でした。ある日、彼らは夕食のためににわとりをつかまえることにしました。どちらもひとりぼっちになりたくなかったので一緒に行くことにしました。彼らにはわとりが大好物だったのです。彼らは夜になるまで待ちました。そして、彼らは農場のそばまで大急ぎで駆けて行きました。彼らはそこにわとりがいることを知っていたのです。くまはたいへんなまけものだったので、屋根の上に登って見張っていました。きつねはたいへん注意深くにわとり小屋の戸を開けました。彼は一羽のにわとりをつかまえて殺しました。彼がそれをにわとり小屋の外へ運び出そうとしている時、屋根の上にはいたくまの体重で屋根がみしみしと音をたてました。きつねはその音を聞いてびっくりしました。けれども、もう逃げ出すことはできません。くまが落っこちて来て五羽のにわとりが死んでしまいました。すぐに農夫が出てきて、何事が起こったのかを知りました。(約400字)
- (2) 友達同志のきつねとくまは、ある日夕食にするにわとりをつかまえに、二匹で夜になってから農場のそばまで大急ぎで駆けて行きました。なまけものくまは、屋根の上で見張りをし、きつねは注意深くにわとり小屋に入り一羽のにわとりをつかまえて殺しました。きつねがにわとりを運び出そうとしている時にくまの体重で屋根がぬけ、くまが落っこちて五羽のにわとりが死んでしまいました。この事はすぐに農夫に知られてしまいました。(約200字)
- (3) きつねとくまは夕食にするにわとりをつかまえに農場へ行きました。きつねが一匹のにわとりを小屋から運ぼうとした時、屋根の上で見張りをしていたくまが落っこちてしまい、すぐに農夫が出てきて事態を知られてしまいました。(約100字)
- (4) きつねとくまが、農場にいるにわとりをつかまえに行ったが、くまが屋根から落ちて農夫に見つかってしまいました。(約50字)

(2)～(4)は、ある心理学的実験において、文章(1)を読んだ同一の被験者によって実際に産出された長さの異なる3種類の要約文章である。人間は、要求されれば様々な長さの要約文章を産出することができる。それでは、そのとき人間はどのような内的処理をしているのであろうか。異なる長さの要約を生成するときには、まったく異なった処理をしているのであろうか。

もとの文章と産出された要約文章を比較することによって分かるのは、ある情報は抜け落ち(情報の削除)、あるいはいくつかの情報はまとめられ(情報の集約)、またある情報は被験者の解釈によって歪められる(情報の歪曲)、などといった現象が起こっているということである[1]。そのような現象は、いったいどのような内的な処理のメカニズムによってもたらされたのであろうか。どのような情報がどのような理由によって削除され、どのような情報同士がどのような理由で集約されるのであろうか。本研究では、対象とする文章を“物語文章”とし、このような要約化の処理過程についての基礎的な考察を行う。そして、特に、要約化の際に用いられるであろう知識(要約化の規則)について具体的に記述することを試みる。

2. 要約化の処理の過程

自然言語処理のシステムは、文章中の言語表現から個々の命題を抽出して解釈し、さらにそれらが相互にどのように関連し合っているのかを解析しなければならない。要約化処理は、その解析結果に基づいて行なわれると考えられる。ところで、人間が要約を行なう場合には、重要でない情報を積極的に削除して重要な情報のみを残すというよりは、もとの文章中に明示されているできるだけ多くの情報を(たとえ、その中にあまり重要でない情報が含まれようとも)集約・凝縮して要約文章中に取り込もうとする傾向がある。このことは、要約化処理の1つの特徴を表わしている。すなわち、要約化処理とは、文章中の個々の命題情報の中から意味的に集約できるものを捜し出し、それらをより少数のマクロな命題に変換していく処理過程である、といえるであろう。もちろん、変換されたマクロ命題同士もまたさらに、よりマクロな命題へと変換され得る。情報の抜け落ち現象や抽象化現象は、このような処理の結果として必然的にしかしかくまでも副次的に生じてくるものと考えられる。

では、どのような命題とどのような命題がどのような意味的關係にあるときに、より大きな意味的なまとまりを作るのか。言い換えるならば、どのような意味的結合パターンの命題群からひとつのマクロな命題が構築されるのか。それは、要約化に関する知識、すなわち、要約化の規則に依存する。

3. 要約化の規則

多くの文章理解の研究者は、彼ら自身の文章理解のモデルを提案するとともに、文章の要約化の現象についての議論も行なっている[2,3,4,5,6,7,など]。しかしながら、それらの多くは、彼らの理解モデルで容易に扱える構造をもつ文章に限定された議論であり、極めてトップダウン的な要約化処理についての議論である傾向が強い。また、ボトムアップ的に要約化の処理をとらえている場合でも、その要約化を支える知識について具体的に定式化した研究は少ない。

我々は、文章中に明示された個々の情報から、どのようにして“要旨”が抽出され、また、要約文章が産出されるのか、その一般法則やアルゴリズムに興味がある。したがって、我々は、多くの要約産出の心理学的実験を行い、そこで得られた大量のデータに基づいて、文章の要約化の規則の定式化を試みている[1]。本稿では、現在までに明らかになった要約化の規則について、以下に述べていく。

(a) 共通の項または述部を持ち、「共起」関係にある命題群を1つにまとめる。

ここでいう「共起」関係とは、時間的に同時に生起（開始、継続、終了）し、それらの間には何ら因果的あるいは含意的関係を持たない命題間の関係を指す。

例) 花子の父は先生である。花子の母も先生である。====> 花子の両親は先生である。

[状態(花子の父, 先生である) <-- 共起 --> 状態(花子の母, 先生である)]

====> 状態(花子の両親, 先生である)

太郎はその荷物を持って移動した。====> 太郎はその荷物を運んだ。

[状態(太郎, 持っている(荷物)) <-- 共起 --> 過程(太郎, 移動する)]

====> 過程(太郎, 運ぶ)

(b) ある状態を表わす2つの命題が、「変化」関係にあるならば、それらを「状態変化」として1つにまとめる。

ここでいう「変化」関係とは、2つの対立する（同時には成立し得ない）命題が時間的に重複することなく前後して成立するとき、それらの関係を指す。

例) 太郎は健康であった。今、太郎は病気である。====> 太郎は病気になった。

[状態(太郎, 健康である) --- 変化 ---> 状態(太郎, 病気である)]

====> 状態変化(太郎, 健康である, 病気である)

そのバスは停止していた。今、そのバスは動いている。====> そのバスは動きだした。

[状態(バス, 停止している) --- 変化 ---> 状態(バス, 動いている)]

====> 状態変化(バス, 停止している, 動いている)

信号は赤だった。今、信号は青である。====> 信号は赤から青に変わった。

[状態(信号, 赤である) --- 変化 ---> 状態(信号, 青である)]

====> 状態変化(信号, 赤である, 青である)

(c) 同一の動的状態の開始を表わす命題と終了を表わす命題を「過程」として1つにまとめる。

例) その風車は回り始め、その後、停止した。====> その風車は回った。

[状態変化(風車, 停止, 回転) + 状態変化(風車, 回転, 停止)]

====> 過程(風車, 回転する)

花子は泣き始めたが、すぐに泣き止んだ。====> 花子は泣いた。

[状態変化(花子, 泣いてない, 泣いている) + 状態変化(花子, 泣いている, 泣いてない)]

====> 過程(花子, 泣く)

(d) 一連の「過程」から「状態変化」にいたる「因果」関係の連鎖を、「出来事」として1つにまとめる。

ここでいう「因果」関係とは、何らかの「過程」が何らかの「状態変化」や別の「過程」の開始を直接的に生起させる場合、それらの命題同士の関係を指す。

例) 太郎は移動した。そして、太郎は空港に着いた。====> 太郎は空港へ行った。

[過程(太郎, 移動する) --- 因果 --> 状態変化(太郎, 空港にいない, 空港にいる)]

====> 出来事(太郎, 行く(X, 空港))

花子は踊った。花子は疲れてしまった。====> 花子は踊り疲れた。

[過程(花子, 踊る) --- 因果 --> 状態変化(花子, 元気である, 疲れている)]

====> 出来事(花子, 踊り疲れる)

太郎が岩を押したら、岩は動いた。====> 太郎は岩を動かした。

[過程(太郎, 押す(岩)) --- 因果 --> 状態変化(岩, 止まっている, 動いている)]

====> 出来事(太郎, 動かす(岩))

太郎がボールを打つと、ボールは飛び、窓ガラスにぶつかった。窓ガラスは割れてしまった。

====> 太郎は窓ガラスを割った。

[過程(太郎, 打つ(ボール)) --- 因果 --> 過程(ボール, 飛ぶ)

--- 因果 --> 過程(ボール, ぶつかる(窓ガラス))

--- 因果 --> 状態変化(窓ガラス, 割れていない, 割れている)]

====> 出来事(太郎, 割る(窓ガラス))

上述の例で分かるように(3番目あるいは4番目の例のように)、集約された結果が、もとの命題情報のすべてを表わしきれないことはしばしば生ずる。これが要約における情報の欠落現象の原因の1つとなっている。

(e) ある登場人物が何らかの「過程」を願望し、それを実行した場合、それらを「意図的過程」として1つにまとめる。

例) 太郎は動こうと思った。太郎は動いた。====> 太郎は(意図的に)動いた。

[願望(太郎, 過程(太郎, 動く)) --- 実行 --> 過程(太郎, 動く)]

====> 意図的過程(太郎, 動く)

太郎はその岩を押そうと思い、実際に押した。====> 太郎はその岩を押した。

[願望 (太郎, 過程 (太郎, 押す (岩))) --- 実行 --> 過程 (太郎, 押す (岩))]
====> 意図的過程 (太郎, 押す (岩))

(f) ある登場人物が何らかの「出来事」を願望し、それを実現した場合、それらを「意図的出来事」として1つにまとめる。

例) 太郎は倒れようと思った。太郎は倒れた。====> 太郎は (意図的に) 倒れた。

[願望 (太郎, 出来事 (太郎, 倒れる)) --- 実現 --> 出来事 (太郎, 倒れる)]
====> 意図的出来事 (太郎, 倒れる)

花子は窓を開けようと思い、窓を開けた。====> 花子は窓を開けた。

[願望 (花子, 出来事 (花子, 開ける (窓))) --- 実現 --> 出来事 (花子, 開ける (窓))]
====> 意図的出来事 (花子, 開ける (窓))

(d)、(f)の規則は、その例からも分かるように、多くの場合、言語表現上は願望を表わす情報が削除され、過程や出来事を表わす情報のみが残される、といった情報の取捨選択現象を生起させる。

(g) 一連の「可能」関係にある「意図的出来事」および「意図的過程」を「意図的出来事群」として1つにまとめる。

ここでいう「可能」関係とは、ある「状態」がある「過程」が生起するための必要条件となっている場合、その「状態」あるいはその成立を含意する命題と、その「過程」を表わす命題の間の関係を指す。

例) 太郎はボールを取って、投げた。====> 太郎はボールを投げた。

[意図的出来事 (太郎, 取る (ボール)) --- 可能 --> 意図的過程 (太郎, 投げる (ボール))]
====> 意図的出来事群 (太郎, 投げる (ボール))

太郎はお金を稼ぎ、車を買って、車で道東旅行をした。====> 太郎は車で道東旅行をした。

[意図的出来事 (太郎, 稼ぐ (お金)) --- 可能 --> 意図的出来事 (太郎, 買う (車))
--- 可能 --> 意図的過程 (太郎, 車で旅行する (道東))]
====> 意図的出来事群 (太郎, 車で旅行する (道東))

これらの例から分かるように、規則(g)は、言語表現上は最終的な「意図的出来事」あるいは「意図的過程」のみを選択するように機能する。

(h) ある登場人物が何らかの事実を正しく認知した場合、その認知とそれを生起させた原因 (伝達行動、知覚、思考など) およびその事実を「認知的出来事」として1つにまとめる。

例) 太郎は病気であった。太郎の母はそのことを先生に告げた。それで、先生は太郎が病気であることを知った。====> 太郎の母は、先生に太郎が病気であることを知らせた。

[状態 (太郎, 病気である) + 過程 (太郎の母, 告げる (先生, 状態 (太郎, 病気である)))
--- 因果 --> 状態変化 (先生, 認知 (状態 (太郎, 病気である)))]
====> 認知的出来事 (太郎の母, 知らせる (先生, 状態 (太郎, 病気である)))

ありが池に落ちた。はとは、それを見ていたので、ありが池に落ちたことが分かった。

====> はとは、ありが池に落ちたことを知った。

[出来事(あり, 落ちる(池)) + 過程(はと, 見る(出来事(あり, 落ちる(池)))]

--- 因果 --- 状態変化(はと, 認知(出来事(あり, 落ちる(池))))]

====> 認知的出来事(はと, 認知(出来事(あり, 落ちる(池))))

- (i) 「認知的出来事」あるいは何らかの生理的欲求とそれによって動機づけられた「意図的出来事群」を「エピソード」として1つにまとめる。

例) はとはありがおぼれそうになっているのを見て、ありが助けてやった。====> はとはありが助けた。

[認知的出来事(はと, 認知(状態(あり, おぼれそうである)))]

--- 動機づけ --- 意図的出来事群(はと, 助ける(あり))]

====> エピソード(はと, 助ける(あり))

太郎はのどが渇いていたので、水を飲んだ。====> 太郎は水を飲んだ。

[状態(太郎, のどが渇いている) --- 動機づけ --- 意図的出来事群(太郎, 飲む(水))]

====> エピソード(太郎, 飲む(水))

- (j) 「認知的出来事」あるいは何らかの生理的欲求と、それによって動機づけられた「願望」、それを実現するための「意図的出来事群」、およびその結果を「失敗エピソード」として1つにまとめる。

例) 父は、太郎の成績が悪いことを知って、太郎が成績を上げるようにと、参考書を与えたが、太郎は勉強しようとはしなかった。====> 父は、太郎に勉強させることに失敗した。

[認知的出来事(父, 認知(状態(太郎, 成績が悪い)))]

--- 動機づけ --- 願望(父, 意図的出来事(太郎, 上げる(成績)))]

<--- 理由 --- 出来事(父, 与える(太郎, 参考書)) + 状態(太郎, 勉強しない)]

====> エピソード(父, 失敗する(意図的出来事(太郎, 勉強する)))

太郎はおなかがかすいていたので、何か食べようと思い、戸棚の中を捜したが、手に入れたのはかびのはえたパンだけであった。====> 太郎は、空腹を満たすことができなかった。

[状態(太郎, 空腹である)]

--- 動機づけ --- 願望(太郎, 意図的出来事(太郎, 食べる(X)))]

<--- 理由 --- 過程(太郎, 捜す(戸棚の中))]

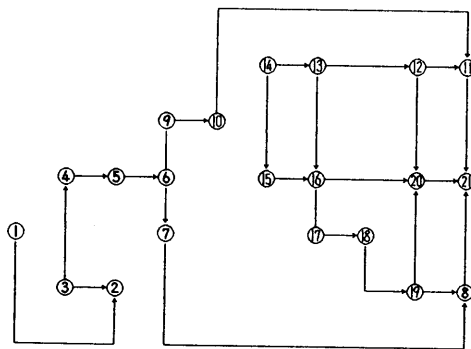
--- 因果 --- 状態変化(太郎, 持っていない(かびたパン), 持っている(かびたパン))]

====> エピソード(太郎, 失敗する(意図的出来事(太郎, 空腹を満たす)))

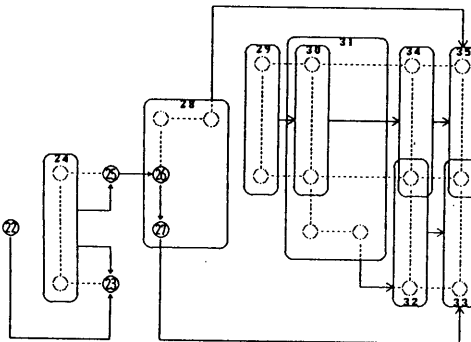
これらの規則のうち、(a)、(b)、(c)は比較的局所的な規則であり、(d)から(j)へと進むにつれて次第により大きなレベルの要約化規則となっている。

4. 要約化の規則の適用例

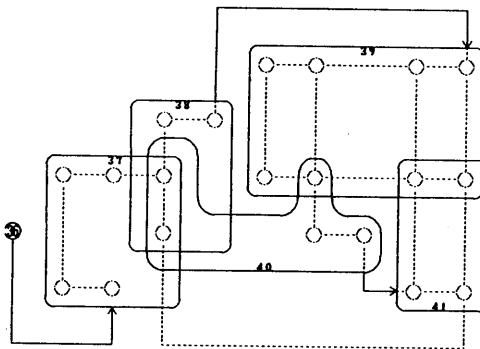
以上の要約化の規則を適用することによって生成された、ある物語の様々なレベルの意味的なまとまりの様子を、図1に示す。なお、図中の各表現の右には、そのほぼ相当と思われる言語表現が記されている。



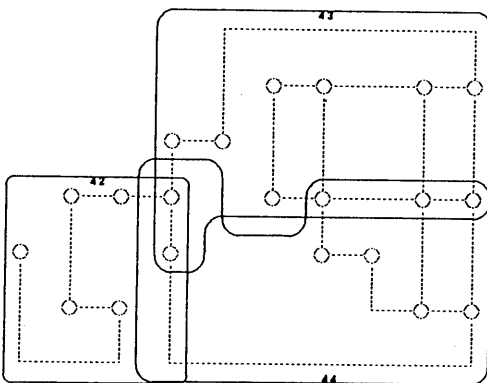
- 1 ありはのどが渴いている
- 2 ありは水を飲むことを望む
- 3 ありは池へ行くことを望む
- 4 ありは池へ行く
- 5 ありは池の中に落ちる
- 6 ありはおぼれそうになる
- 7 ありは苦しい
- 8 ありは助かることを望む
- 9 はとはありがおぼれそうであるのを見る
- 10 はとはありがおぼれそうであることを知る
- 11 はとはありが助かることを望む
- 12 はとはありが木の小枝の上に乗ることを望む
- 13 はとは木の小枝を池の中に投げ入れることを望む
- 14 はとは木の小枝を取ってくる
- 15 はとは木の小枝を取ってくる
- 16 はとは木の小枝を池の中に投げ入れる
- 17 ありはのどが木の小枝を池の中に投げ入れるのを見る
- 18 ありはのどが木の小枝を池の中に投げ入れたのを知る
- 19 ありは木の小枝の上に乗ることを望む
- 20 ありは木の小枝の上に乗る
- 21 ありは助かる



- 22 ありはのどが渴いている
- 23 ありは水を飲むことを望む
- 24 ありは池へ行く
- 25 ありは池の中に落ちる
- 26 ありはおぼれそうになる
- 27 ありは苦しい
- 28 はとはありがおぼれそうであることを知る
- 29 はとは木の小枝を取ってくる
- 30 はとは木の小枝を池の中に投げ入れる
- 31 ありはのどが木の小枝を池に投げ入れたのを知る
- 32 ありは木の小枝の上に乗る
- 33 ありは助かる
- 34 はとはありを木の小枝の上に乗せる
- 35 はとはありを助ける



- 36 ありはのどが渴いている
- 37 ありは水を飲むことに失敗する
- 38 はとはありがおぼれそうであることを知る
- 39 はとはありを助ける
- 40 ありはのどが木の小枝を池の中に投げ入れたのを知る
- 41 ありは助かる



- 42 ありは水を飲むことに失敗する
- 43 はとはありを助ける
- 44 ありは助かる

図1. 要約化の規則を適用することによって得られた、ある物語の4つのレベルの意味的なまとめ。

5. 把握される要旨と産出される要約文章の違い

以上述べてきた要約化の規則に基づく処理の結果は、“要旨”と呼ばれるものにはほぼ相当する。人が文章に対する要約を求められたとき、そこで産出される要約文章の内容は、基本的には、把握され心内に表象される“要旨”を基礎とするものであろう。しかしながら、要約文章は以下のようないくつかの点で心内の“要旨”とは若干異なる性質をもつようである。

- (7) 言語化の際の制約：要旨を構成するようなマクロ命題は往々にして簡潔かつ確かな言語表現で表わすことが困難な情報内容をもつことが多いため、要約の産出にあたっては、より下位の詳細なレベルの命題が参照され、その内容が言語化されることになる。
- (4) 文章のまとまりを維持するための制約：要約は全体としてまとまった文章でなければならない。したがって、要約は、要旨には含まれないような、導入のための背景的情報や要旨の骨子をなす部分を接続するための情報を含まなければならない場合がある。
- (7) 要約産出方略に関わる問題：もとの文章を見ながら要約を産出することができる場合には、文章に内在する情報の集約化、抽象化をあまり行わず、主として文章中の表現の取捨選択を行なうことで要約を産出しようとする人がかなりいる。また、個人内でも、そうする場合がある。これは、要約生成という問題をより容易に（安直に？）行うための一つの方略といえる。

6. おわりに

本稿では、物語文章の要約化処理に焦点を当て、その処理を支える知識（要約化の規則）の一部を具体的に定式化した。さらに、それらを実際の簡単な物語に適用し、物語の様々なレベルの意味的なまとまりの存在を示した。今後は、要約化処理の計算機上での実現を目指す予定である。その際には、各規則の精緻化あるいは修正と新たな規則の追加が改めて必要となるであろう。

参考文献

- [1] 邑本俊亮・阿部純一（1987）物語要約における情報集約の過程。阿部純一（編）、昭和61年度科研（特定研究2）報告書「人間の言語理解過程についての認知論的研究」、北海道大学文学部。
- [2] Rumelhart, D.E. (1977) Understanding and summarizing brief stories. In D. Laberge & J. Samuels(Eds.), Basic processes in reading: Perception and comprehension. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- [3] Thorndyke, P.W. (1977) Cognitive structures in comprehension and memory of narrative discourse. Cognitive Psychology, 9, 77-110.
- [4] Lehnert, W.G. (1982) Plot units: A narrative summarization strategy. In W.G. Lehnert & M.H. Ringle(Eds.), Strategies for natural language processing. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- [5] DeJong, G. (1982) An overview of the FRUMP system. In W.G. Lehnert & M.H. Ringle(Eds.), Strategies for natural language processing. Hillsdale, N.J.: Lawrence Erlbaum Associates.
- [6] van Dijk, T.A., & Kintsch, W. (1983) Strategies in discourse comprehension. New York: Academic Press.
- [7] Graesser, A.C., & Clark, L.F. (1985) Structures and procedures of implicit knowledge. Advances in discourse processes, Vol. 17. Norwood, N.J.: Ablex.