自然言語処理 92－7
(1992. 11. 20)

キーワード法による諺の検出のモデル

土井 晃一*, 金原 史和†, 田中 英彦†

*株式会社 富士通研究所 国際情報社会科学研究所
†東京大学 工学部 電気工学科

計算機上で自然言語理解を行なう際には文字通りの意味の解析だけでは不十分である。ここではそういうものの一つとして諺をあげる。

自然言語理解の応用としてまず考えられるのは機械翻訳である。全世界で刊行される雑誌、新聞等は発行部数が多い。毎日のように新しく刊行されている。なるべくなら母国語で読みたいものだが、人手ではとても間に合わない。しかも現在の機械翻訳ではほとんど扱えない、諺や比喩が頻繁に使われている。現在の機械翻訳では慣用表現は扱うようになってきた [1] が諺のように大きな単位はまだ扱われていない。文章の中心的意味がここで表現されていることが多い。多くは直訳できるがそうでないものも多い。ここの翻訳を間違えると文章全体の意味が通じなくなってしまう。現在の機械翻訳の仕組みから考えて、意味処理をしてから諺などの検出はしにくい。諺のところで構文解析、意味解析が失敗してしまうことが多い。早期に諺などを検出することにより、構文解析、意味解析、文脈・状況理解の助けになる。

諺を検出する際に必要なことは、どこで、どういう形で使われているかである。単に辞書をひくだけでよいこともある。この場合は形態素解析とおなじことになる。諺という品詞を一つ増やせば良い。例えば「馬の耳に念仏」という例だと、この通りにこの場合は「名詞」として辞書に登録すれば良い。しかし諺が少しでも変化するとこの方法は使えなくなる。特に会話文等にこの傾向は顕著である。例えば「何とかとハサミは使いよう」という例が挙げられる。本論文ではこのように変化した諺を可能な限り検出する方法を提案する。

A Model of Proverb Detection Using Keyword Method

Kouichi DOI*, Fumikazu KANEHARA† and Hidehiko TANAKA†

*International Institute for Advanced study of Social Information Science, Fujitsu Laboratories Ltd.
17-25, Shinkamata, 1-Chome, Ota-ku, Tokyo, 144, Japan
Phone: +81-3-(3735)-1111 ex.(2417)
Fax: +81-3-(3730)-6830
doy@iias.flab.fujitsu.co.jp, taknaka@mtl.t.u-tokyo.ac.jp
†University of Tokyo, Faculty of Engineering, Department of Electric Engineering.

Only the analysis of literal meaning comprehension is not enough for natural language comprehension on computer. In this paper, proverb is treated in the example of non-literal meaning.

Machine translation is considered that application of the research of natural language comprehension. There are many journals and newspapers which are published in the world every day. In these, proverb and metaphor are used frequently. They often have a central meaning of the sentences. The-state-of-art machine translation cannot treat such proverb and metaphor.

Syntax analysis, meaning analysis and context analysis can be easily done when the proverbs are detected in the earlier stage of natural language comprehension.

We research the variation of proverbs. We propose the keyword method to detect various proverbs.

1 はじめに

計算機上で自然言語理解を行なう際には文字通りの意味の解析だけでは不十分である。ここではそういうものの一つとして諺をあげる。

自然言語理解の応用としてまず考えられるのは機械翻訳である。全世界で刊行される雑誌、新聞等は発行部数が多い。毎日のように新しく刊行されている。なるべくなら母国語で読みたいものだが、人手ではとても間に合わない。しかも現在の機械翻訳ではほとんど扱えない、諺や比喩が頻繁に使われている。現在の機械翻訳では慣用表現は扱うようになってきた [1] が諺のように大きな単位はまだ扱われていない。文章の中心的意味がここで表現されていることが多い。多くは直訳できるがそうでないものも多い。ここの翻訳を間違えると文章全体の意味が通じなくなってしまう。現在の機械翻訳の仕組みから考えて、意味処理をしてから諺などの検出はしにくい。諺のところで構文解析、意味解析が失敗してしまうことが多い。早期に諺などを検出することにより、構文解析、意味解析、文脈・状況理解の助けになる。

諺を検出する際に必要なことは、どこで、どういう形で使われているかである。単に辞書をひくだけでよいこともある。この場合は形態素解析とおなじことになる。諺という品詞を一つ増やせば良い。例えば「馬の耳に念仏」という例だと、この通りにこの場合は「名詞」として辞書に登録すれば良い。しかし諺が少しでも変化するとこの方法は使えなくなる。特に会話文等にこの傾向は顕著である。例えば「何とかとハサミは使いよう」という例が挙げられる。本論文ではこのように変化した諺を可能な限り検出する方法を提案する。

諺の検出方法には次の二つの方法が考えられる。

- 文脈と状況を理解してから、文字通りの意味との矛盾を見い出して諺を検出する方法
- 文字列のパターンから諺を検出する方法。

前述のように1の方法は色々な困難が生じる。本論文では2の方法を用いる。基本的に文字列のパターンから諺を検出するので、形態素解析直後に検出することにする。その具体的な方法としてはキーワードを使う方法を提案する。ここで言うキーワードとは名詞、動詞、形容詞、形容動詞の語幹部分を指す。例えば「馬の耳に念仏」という例だと「馬」、「耳」、「念仏」の三つがキーワードとなる。

本論文ではまず諺が現実の文の中でどのように現れるかを分析する。次にそれらの変化に対応できるキーワード法を提案する。

2 検出における問題

まずどういう形で諺が文章中に現れるかを考察する。その現れ方に対応できるモデルを次に考える。

諺を文や会話で使う時、必ずしも辞典に載っているとおりに使うとは限らない。更に辞典によって表記が違う諺も存在する。諺がそのまま文字通りの形で現れないで変化した場合、データベースの中に諺として登録してある文字列との単純なマッチングでは諺を検出する事は出来ない。例えば、『二兎を追う者は一兎を得ず』が『二兎追う者は一兎を得ず』となっただけでも、マッチングに失敗してしまう。また、この問題とは別に、検出したつもりでも、本当に諺として使われているのかどうかという問題も生じる。

このように、変化した諺を検出するには、単純なマッチングでは無理であるので、他の方法が必要となる。ここではまず諺がどのように変化し得るかを議論する。

なお、ここで言う変化とは形態の変化であり、意味の変化ではない。従って、『二兎を追う者は一兎を得る』のような、皮肉や冗談を目的として故意に意味を変えたようなものは取り扱わない。

グループ μ の分析によると [2]、一般に文彩は変化という観点から、大きく削除と付加に二分される。それをさらに細別すると以下のようになる。

2.1 助詞の付加・挿入・省略・置換

1. 挿入：目糞鼻糞を笑う → 目糞が鼻糞を笑う
2. 省略：馬鹿があって利口が引き立つ → 馬鹿あって利口が引き立つ
3. 置換：二兎を追う者は一兎を得ず → 二兎を追う者は一兎も得ず
4. 付加：急がば回れ → 急がば回れよ

助詞は格助詞・接続助詞・副助詞・終助詞の四種類に分類されている。諺で使われている助詞は、主に格助詞と、副助詞の『は』や『も』であり、また、これらの助詞全てがこのように変化出来るわけではない。例えば、格助詞『が』は副助詞『は』に置換出来るが、格助詞『と』に置換する事はまずあり得ない。仮にあったとしても、頻度は非常に低いと考

えられる。また、例4のように付加し得る助詞は終助詞がほとんどである。

2.2 用言 (助動詞・動詞・形容詞・形容動詞) の活用

1. 馬鹿と相場には勝てぬ → 馬鹿と相場には勝てず
2. 雨が降ろうが槍が降ろうが → 雨が降っても槍が降っても

諺が文中において使われる場合、単語の語尾変化が起こりやすくなる。また、動詞などが活用を起こす時は、新たに助詞や助動詞を伴う場合が多い。

2.3 語順の変化

1. 馬鹿とハサミは使いよう → ハサミと馬鹿は使いよう
2. 雨が降ろうが槍が降ろうが → 槍が降ろうが雨が降ろうが
3. 鼻糞で鯛を釣る → 鯛を鼻糞で釣る

諺において語順の変化は主に名詞同士で起こる。例2のように動詞が語順の変化に関わるような諺は少ない。なお、『猿も木から落ちる』が『木から落ちる、猿も』のようになる変化はありえない。[3]

2.4 辞典によって表記の異なるもの

1. 馬鹿の大食 ←→ 馬鹿の大食い
2. 馬鹿があって利口が引き立つ ←→ 馬鹿があればこそ利口が引き立つ

諺の辞典によっては表記の異なる諺も存在する。このためデータベースにどちらを登録すればよいか困るし、諺を使う側も必ずしも登録した方を使うとは限らない。

2.5 自立語の付加

1. 馬の耳に念仏 → 馬の耳に念仏を唱える
2. 鬼の目にも涙 → 鬼の目にも涙が有る

付加とは、文字通り、諺の先頭か末尾に単語が付くことである。例でもわかるように、変化した形の中に元の形がそのまま残る場合が多い。当然、終助詞の付加と違って、例1を『(諺：馬の耳に念仏) を唱える』と捉えてはいけないので、『唱える』まで正しく検出してやる事が重要である。しかし、実際問題、このように付加した諺を目にする事は少ない。[3]

2.6 自立語の挿入

1. 石橋を叩いて渡る → 石橋をよく叩いて渡る
2. 火事あとの火の用心 → 火事が起こったあとの火の用心

形容詞や副詞などは例1のように名詞や動詞に付く事が可能なので、自立語の挿入は、文法的にはいくらでも起こり得る。しかし、実際に日常では例1のように使われる例というのは、付加同様、極めて少ない。[3]

2.7 自立語の省略

1. 馬の耳に念仏 → 馬に念仏
2. 二兎を追う者は一兎を得ず → 何とかは一兎を得ず
3. 二兎を追う者は一兎を得ず → 二兎を追う者は何とやら

例1は純粋な省略である。例2，例3は省略と付加が同時に起こったものであるが、省略の場合、『何とか』とか『何とやら』とかいう言葉が省略された部分に入る事が多い。これは、置き換えが起こったともみなせるが、一対一の置き換えではないので、置換の部類には入れない事にする。

2.8 置換

1. 鬼の首を取ったよう → 鬼の首を取ったみたい
2. 転ばぬ先の杖 → 転ばぬ前の杖
3. 手も足も出ない → 手も足も出せない
4. 馬鹿とハサミは使いよう → 太郎とハサミは使いよう
5. 鬼の目にも涙 → 部長の目にも涙
6. 嘘つきは泥棒の始まり → 嘘つきは何とかの始まり

ここで言う置換とは、自立語同士の一対一の置き換えを指す。普通は、名詞には名詞がというように、同じ品詞同士で置換が起こる。『始まり』が『始ま

る』に変わるように、転用名詞が元の動詞に変わったものは、置換とはせず、語尾変化を起こしたものとして捉える。置換には、例1、例2、例3のように単に同じ意味、または似た意味の単語を置き換えたものと、例4、例5のように、情報の提供や強調などを目的として置換されたものとの二種類ある。例6は双方に属せるような例である。後者の場合、名詞対名詞の置換が主になり、それも固有名詞、あるいはある特定人物を指す名詞に置き換わる場合が多い。例4では、諺を使う側は太郎が馬鹿である事を言いたいのであり、例5では、部長が鬼のような人物である事を言わんとしている。[4]

2.9 二重に真のもの

1. 馬鹿の大足 …… 馬鹿というあだ名の男がいるかもしれない。
2. 隣の花は赤い …… 本当に、庭に赤い花があるかもしれない。

これは、変化の問題ではないが、形は諺で、意味は諺でない例である。このような文字列を逆に諺でないと判断するには文脈と状況の理解がどうしても必要になる。

2.10 その他

1. 二兎を追う者は一兎を得ず → 二羽の兎を追う者は一羽の兎を得ず
2. 蟻の這い出る所もない → 蟻の這い出られる所もない

例1は接尾語の挿入、例2は助動詞の挿入である。このような、細かい変化は、この他、多々ある。

2.11 組合せ

当然の事ながら、以上の変化は組合わさる。

1. ハサミと馬鹿は何とやら (語順の変化、自立語の省略と付加)
2. 馬に念仏を唱える (自立語の省略と付加)
3. 二兎追っている太郎は一羽も得られない (助詞の省略・挿入、動詞の活用、補助用言の挿入、自立語の置換、接尾語の挿入、助動詞の挿入)

以上、検出における主な問題点を列挙した。これらの問題の中には、頻度の差がある。日常よく使われる変化の形から、そうでない形まである。例えば、自立語の挿入や付加のように、諺の元の形に含まれない自立語が現れる変化というのは少ない。逆に、省略や用言の活用は頻度が高い。そこで、この諺の変化の問題を、頻度の高いものから優先的に対処していく事を考える。

本論文では、これらの問題のうち、問題2.6の自立語の挿入以外のすべての問題を取り扱う。ただし、問題2.8の置換に関しては、例2のようなものは取り扱わない。つまり、名詞や動詞などの自立語の置換は、先述したように、何かを強調したり、新しい情報を提供しようとして起こった置換に的を絞る。具体的には自立語では名詞をその対象とし、名詞の中でも具体物 (特に動物) を表す名詞に限って考える。詳細は3章で述べる。

3 検出の方法

3.1 キーワード

2章で挙げたように、諺の形の変化は、単語の置換・付加・挿入・省略・語順の変化、またはこれらの組合せに属する。しかしどのように変化しても、原形の一部は必ず残っている。そうでなければ、人間が見ても検出できない。助詞、助動詞などを残して名詞、動詞などをそっくり入れ替えて、状況、文脈などから元の諺を連想させる場合もあるが、ここではそういう例は扱わない。また、諺の一つの特徴として、諺は平常文ではあまり見られないような単語の組合せから構成されているものが多い。例えば、『鬼に金棒』ならば、普通、桃太郎の話でもない限り、『鬼』とか『金棒』とかいう単語が出現する頻度は低い。『鬼』と『金棒』が並んで出てくるとなるとさらに頻度が低くなる。このため、人間は諺の一部を見ただけで、それが諺である事を認識できる。それが諺でなくとも、少なくともその諺を連想する事はできるはずである。『馬の耳に念仏』という諺ならば、『の』『に』などの助詞や『耳』が欠落していても、『馬』と『念仏』だけで、『馬の耳に念仏』なる諺を連想できる。だが諺の一部ならば何でもかんでもいいというわけではない。『落ちる』から『猿も木から落ちる』を連想する事は普通できない。『猿も木から落ちる』の中で占める『落ちる』の量的 (要するに文字数) な割合が少ないためでもあるが、それ以上に『落ちる』という動詞が平常文で頻繁に使われるからである。逆に、原形の中で占める量的な

割合が低くても使用頻度の少ない単語(または熟語・複合語)ならば、それだけで元の諺を連想できる場合が多い。『二兎』を見て『二兎を追う者は一兎を得ず』を連想できるのがその例である。

そこで、本研究では、このようないわゆるキーワードを利用して、2章で挙げたような変化した諺の検出を試みる。つまり、諺を単語単位に切り分け、その単語ごとのマッチングを行なうことにする。

何をキーワードとするかをまず議論する。上で述べたように、諺を構成する単語には、それぞれ、その単語単独で元の諺をどのくらい連想できるかという重み付けがあると考えられる。例えば、『捕らぬ狸の皮算用』という諺なら、『狸』より『皮算用』の方が原形を連想しやすい。この重み付けは、単語の使用頻度が低くなればなるほど、重くなる。そこで、使用頻度の高い単語はキーワードとせず、また、諺から省略されにくい単語をキーワードとする必要がある。ここでは、単語ごとにいちいち重み付けを決める事はせず、品詞ごとに考え、どの品詞をキーワードにするかを考える。まず、種類が限られ頻繁に用いられる助詞と助動詞はキーワードとはしない。次に、諺ではあまり使われない、感動詞と連体詞と接続詞及び副詞らも含めない。さらに、省略や付加や挿入の起こりやすい、接尾語と接頭語もキーワードの対象にはしない。結局、キーワードと成り得るのは、名詞と動詞と形容詞並びに形容動詞の四品詞となる。ただし、動詞と形容詞の補助用言(『追っている』の『いる』や、『投げてみる』の『みる』等)はキーワードとはしない。

以下、キーワードといえば、これら四品詞を指すことにする。

3.2 キーワードによる検出方法の概要

図1はキーワードによる検出の概要を例で示した図である。検出は形態素解析直後に行なうので、検出の段階では入力文は既に品詞分解されている。まず、その形態素解析結果からキーワードのみを抽出する。諺本体も、形態素解析をしてキーワードを抽出したデータを用意し、キーワード単位のマッチングを行なう。つまり、入力文のキーワード群の中に諺のキーワード群が存在するかどうか識別することになる。その際、諺のキーワードが連続して並んでいる必要がある。もちろん、諺のキーワードが全て、入力文のキーワード群の中になくともよい。また並んでいる順番がバラバラでもよい。なお、形容詞・形容動詞はその終止形を、動詞はその語幹の先頭文字をキーワードとする。この理由は3.3節で述べる。

3.3 キーワード検出方法による各問題への対処

本節では、今回取り扱う問題に対して、このキーワード検出方法がどう役に立つかをみていく。

3.3.1 助詞の付加・挿入・省略・置換

助詞はキーワードの対象外であるので、助詞の変化は、マッチングには影響を及ぼさない。

3.3.2 用言(助動詞・動詞・形容詞・形容動詞)の活用

先述したように形容詞と形容動詞はその語幹をキーワードとするので、どのように活用しても構わない。動詞はその語幹の先頭文字をキーワードとするが、これにより、活用による語尾変化に耐えられると共に、可能動詞・使役動詞・自動詞・他動詞による差異をなくす事ができる。例えば『泳ぐ』と『泳げる』は活用の型が違い、終止形に直すだけではマッチングに失敗する。語幹は同じなので語幹を動詞のキーワードにする事も考えられるが、『飛ぶ』と『飛ばす』のように語幹まで異なるものも存在するため、先頭文字をキーワードとした。なお、口語では、『泳げる』も『飛ばす』も、動詞と助動詞の組合わさったものではない。

- 可能動詞:『泳ぐ』と『泳げる』、『出す』と『出せる』
- 使役動詞:『飛ぶ』と『飛ばす』、『行く』と『行かす』
- 自・他動詞:『出る』と『出す』、『始まる』と『始める』

また、転用名詞はすべて動詞とみなす。いささか不自然な例ではあるが、

『嘘つきは泥棒の始まり』→『嘘つきから泥棒が始まる』

というような変化にも対応できる。先頭文字をキーワードにする事で、他の単語との取り違いが懸念されるが、実際、先頭文字が同じで意味が全く違う動詞は存在しないし、他の品詞との取り違いは、品詞の情報を残しておけばそれで解消される。

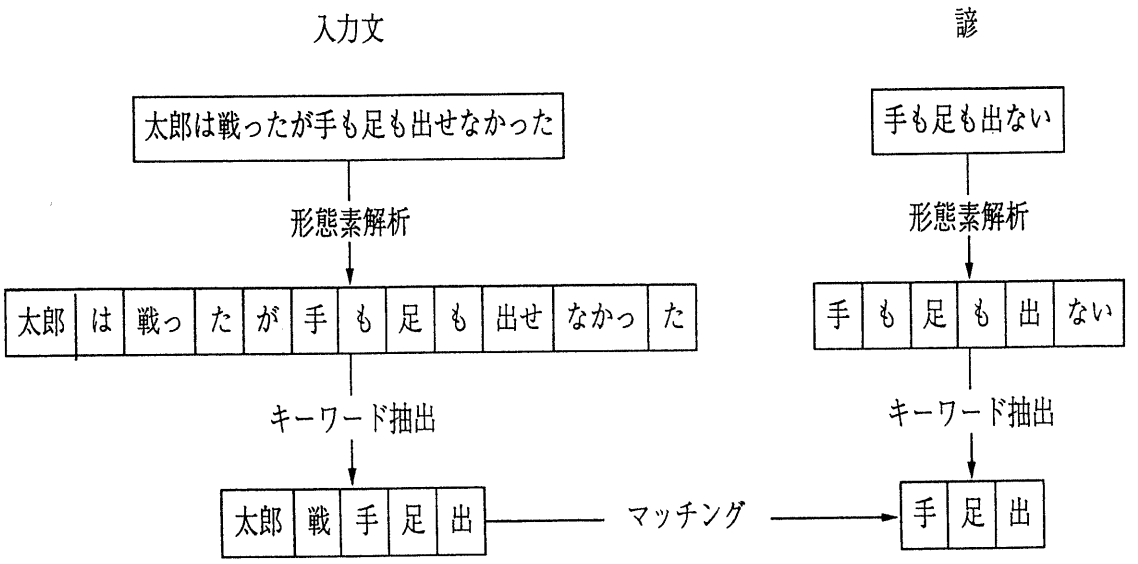

図 1: キーワード抽出による検出の概要

3.3.3 語順の変化

語順の変化もキーワード検出方法により解決される問題である。キーワードを語の組として扱うからである。ただし、許されない語順の変化も許してしまうという欠点がある。これについては、3.3節で述べる。

3.3.4 辞典によって表記の異なるもの

辞書による表記の差異の問題は、結局、助詞の置換・付加・省略・挿入、用言の活用、送り仮名の違いのものだけを扱うので、これもキーワードを名詞、あるいは用言の語幹部分とすることで解決される問題である。

3.3.5 自立語の付加

もともとの諺のキーワード群にないキーワードがその諺に付加した場合、当然、その新たなキーワードを、キーワードのマッチングだけで捕まえる事は無理である。そこで、キーワードとして抽出されなかった単語にも着目する必要がある。まず、どういう場合に自立語の付加が起こり得るかを見てみる。もともと自立語の付加はあまり起こらない変化であるが、起こるとすれば、2節の項目 2.5で示したように、諺の末尾に、しかも、動詞が付加する場合が多いと思われる。さらに、そのような付加を許す諺は名詞で終る諺に多い。また、動詞が名詞の後に付加する場合、両方の単語間には助詞が入らなければならない。この助詞は何でも構わないというわけではなく、主に格助詞が入る。以上の事に着目して、助詞の種類に注目すれば、付加が起こったか否かが判断しやすくなるはずである。

3.3.6 自立語の省略

純粋な省略の場合は、新しいキーワードは登場しないので、残ったキーワードのみを捕まえてそれを諺とすればよい。また、2節の例 2、例 3のように、『何とか』や『何とやら』が省略された位置に入る場合には、そのような単語が諺のキーワードになり得るという情報を与えておいてやればよい。

3.3.7 置換

2章でも述べたように、具体物 (特に動物) を表す名詞の置換を取り扱う。この置換を取り扱う場合、当然、キーワードと入れ替わった単語は、マッチングでは捕まえられない。そこで、諺を構成するキーワードの中で、予め置換できるキーワードとできないキーワードとの区別をつけておけば、対応できるはずである。例えば、『鬼の目にも涙』ならば、『鬼』は置換され得るが、『耳』と『涙』は置換対象とはならない。この区別は、人間に置き換わるかどうかを主な基準としている。よって、『部長の目にも涙』ならば、『鬼』は『部長』に置換できるという情報から、『部長』を諺の一部と判断するのである。

3.3.8　二重に真のもの

この問題は字面を正しく捉えるだけでは解決されない。どうしても文脈や状況を知る必要がある。

3.4　キーワードによる検出方法の問題とその対策

3.4.1　諺の末尾の捕捉

この問題は助動詞や終助詞をキーワードの対象外にしたために起こる問題である。キーワードは捕まえられても、助動詞や助詞を捕まえないために検出した結果が、元の諺の意味と違う事が起こり得る。まず、以下の例文を見てみる。

1. 『大船に乗ったよう』→『大船に乗ったみたい』(比況)
2. 『老いては子に従え』→『老いては子に従うべし』(命令)
3. 『開いた口には戸は立たぬ』→『開いた口には戸は立たない』(打消)
4. 『手も足も出ない』→『手も足も出す事が出来ない』(打消)
5. 『酒は飲むとも飲まるるな』→『酒は飲んでも飲まれてはいかん』(打消命令)

掲げた例は、どれも助動詞や助詞が変化した例でキーワード自体は残っている。よって、キーワードのマッチングによって、諺の存在は認識できる。しかし、諺がどこから始まってどこで終るかという判断はキーワードのマッチングだけでは出来ない。この諺の範囲を正しく検出しない事には意味を正しく把握できない。例4の場合だと、キーワードだけを諺と捕らえると『手, 足, 出』が捕まるので、『手も足も出す』が諺となってしまう。今、仮に『手も足も出ない』の内部の意味の表現を『全く通用しない』とすると、全体として『全く通用しない事が出来ない』となって、不自然な文になってしまうばかりでなく意味も逆転してしまう。このように、キーワードを捕まえて諺の存在を認識できても、諺の範囲を正確に決めるのはキーワードを使うだけでは難しく、裏を返せば、キーワード検出の欠点といえる。では、どう対処するか? 主にこういう問題は諺の末尾に起きる。普通、諺はキーワードで始まるものがほとんどだからである。変化が起こっても同じ事が言える。そこで、諺の末尾の形に注目すると、名詞で終るもの、動詞・形容詞・形容動詞の終止形で終るもの、動詞の命令形で終るもの、助動詞で終るもの、終助詞『な』で終るものが多い。名詞で終るものについて、起こり得る諺の末尾に関する問題は自立語の付加であるが、これについては既に述べた。今、問題となるのは、助動詞で終るもの、動詞の命令形で終るもの、終助詞『な』で終るものである。助動詞で終るものは、例に示したような、比況または打消の助動詞で終るものが圧倒的に多い。

この点に着目して、元の諺の末尾の単語の意味を考慮して、それと同じ意味の単語を、マッチングした最後のキーワードより後ろの所で探してやり、最初に見つけたものまでを諺とすれば、諺の範囲を正確に把握する事ができるであろう。実際に例4を用いて説明してみる。まず、入力文の中に「手も足も出す事が出来ない」という文があったとすると、キーワードとして『手』『足』『出』を捕らえ、『手』から『出』までを諺と判断する。その後、諺の原形の末尾の単語『ない』をみて、『ない』が打消の助動詞である事から、打消を意味する単語(主に助動詞)が『出』の後にないかどうかを探す。その結果、『出来ない』の『ない』を見つけて、新たに、そこまでを諺とするのである。他の例に関しても同様である。ちなみに例1に出てくる『よう』や『みたい』は正確には接尾語に属するが、この場合は助動詞の語幹として扱う。

この他、接尾語や助詞で終る諺もあるが、これらについては、ほとんどの場合、そのまま形が残るか、省略されるかのいずれかであるので、一度、キーワードで検出した後、その接尾語か助詞が、検出されたキーワード群の中での最後のキーワードのすぐ後にあるかどうか探してやればよい。

3.4.2　偽のキーワード

この問題には二種類ある。一つは、原文の中に諺は含まれているが、その諺の前後にまぎらわしい偽のキーワードがあるために、諺の範囲が決めにくくなるという問題である。もう一つは、原文が諺を含まない平常文であるのに、文中にまぎらわしい偽のキーワードがあるために、諺として検出してしまうという問題である。前者の例として、図2を見てみる。ここで、考えられる諺の変化は以下の通りである。

1. 『馬鹿, ハサミ』: 省略が起こったと判断
2. 『ハサミ, 次郎, 使』: 語順の変化・置換が起こったと判断
3. 『次郎, 使』: 省略が起こったと判断

4. 『使』: 省略が起こったと判断

5. 『使, 三郎』: 省略・語順の変化・置換が起こったと判断

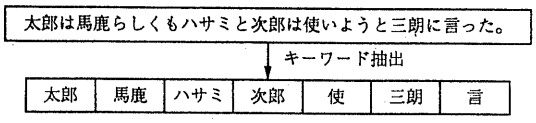

図 2: 偽のキーワード例 1

このように、まぎらわしい文では色々な可能性が生まれる。そこで、これらの可能性の中から正しい場合を検出するために、優先順位を決める作業を行なう。これにより可能な限り正しいものを選べるようになる。

次に後者の例を図 3で見てみる。この場合、迷う事なく諺と判断してしまう。この問題がこのキーワード検出の最大の弱点である。これの対応策として、助詞や助動詞を使った検査方法を採り入れる。以下、この方法について説明する。

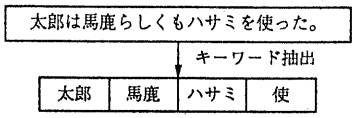

図 3: 偽のキーワード例 2

1. 助動詞の助詞への置換、または助動詞の挿入

   元の諺の中に『名詞+助詞』の形が含まれている場合、これは『名詞+助動詞』には、まずなり得ない。例えば、入力文「馬鹿らしくもハサミを使って…」は『馬鹿とハサミは使いよう』の諺の変化ではない。しかし、キーワードだけを見た場合、諺であると判断してしまう。そこで、キーワード間の単語を調べ、元の諺の形と比べて、このような助詞と助動詞の置換、または名詞と助詞の間への助動詞の挿入が起こっているものは諺とみなさないようにする。

2. 助詞同士の置換

   助詞同士でも置換できないものがある。特に諺で使われる助詞は格助詞と副助詞の『は』と『も』が多いので、格助詞同士の置換に注目する。例えば格助詞『が』は格助詞『から』や『に』などとは基本的に置換し得ない。前者の用法は主語・連体修飾語であり、後者のそれは連用修飾語だからである。このように共通の用法を持つ助詞同士は置換可能となり、それ以外は置換を許さないものとする。

3. 引用の『と』

   『馬の耳に何とかというように…』のように諺は文中で引用的に用いられる事が多いが、逆に諺の中に引用文が入る事はない。引用文が用いられる時には、格助詞『と』がよく使われる。そこで、この引用文の『と』が入り込んでないか、検査する。

4 おわりに

本論文では、文章中に諺がどのように現れるかを考察した。グループ μ の分析により、変化は基本的に削除、付加が基本となることを参考にした。変化として考察したのは、助詞の付加・挿入・省略・置換、用言 (助動詞・動詞・形容詞・形容動詞) の活用、語順の変化、辞典によって表記の異なるもの、自立語の付加、自立語の挿入、自立語の省略、置換、二重に真のものについてである。

次に検出の方法としてキーワードによる方法を提案した。キーワードとして、名詞、動詞、形容詞、形容動詞を挙げ、このキーワード法を用いることにより、諺の変化に対応できることを確認した。

参考文献

[1] 奥雅博. 日本語解析における述語相当の慣用的表現の扱い. 情報処理学会論文誌, Vol. 31, No. 12, pp. 1727–1734, 1990.

[2] Le groupe μ. *Rhétorique Générale.* Librairie Larousse, 1970.

[3] 山梨正明. メタファーと認知のプロセス. 自然言語処理技術 シンポジウム論文集, volume 88-1, 1983.

[4] 篠塚英明, 土井晃一, 田中英彦. 状況意味論を用いた諺理解. 情報処理学会第 41 回全国大会, Vol. 3, No. 3S-9, pp. 149–150, 9 月 1990.