情 報 学 基 礎 5 － 1
（ 1 9 8 7 ・ 6 ・ 2 5 ）

慣用表現について　— 収集と整理 —

田 中 康 仁　　吉　田　将
（姫路短期大学）　　（九州工業大学）

自然言語の処理で大きな問題となる慣用表現について研究した成果を述べる。
慣用表現の取り扱いは重要であると考えられているが，その場その場の処理がなされ大系的な研究までには至っていない。ここでは慣用表現を多量に集め，コンピューターファイルにまとめ，各種の属性を付加している。又，例文ファイルも作成し慣用表現の属性が適切であるか，処理が実際に行えるか等についても検討している。
最後に慣用表現ファイルの応用分野について検討した。

Idiomatic Expressions　— Systematic Gathering —

YASUHITO TANAKA
Himeji College
1-1-12 Shinzaike Honmachi
Himeji-city Hyogo-ken
670 JAPAN

SHO YOSHIDA
Kyushu Institute Technology
1-1 Sensuicho Tobataku
Kitakyushu-city Fukuoka-ken
804 JAPAN

This paper describes the results of a study on idiomatic expressions, which have been obstacles in natural language processing.
Idiomatic expressions have been considered as important, but only adhoc approaches had been taken up till now, and no systematic researches had been made.
In this study, a large number of idiomatic expressions were gathered put into a computer file, and various attributes were added to them. Sample sentence files were also created to test the pertinence of the attributes of the idiomatic expressions and also to see if the and also to see if the processing can be actually performed. Lastly, the areas to which the idiomatic expressions file could be applied was also studied.

1. はじめに

日本語文の中には慣用的に用いられる句が多くある。これらを構成している一語一語を取りあげて，その意味を考えても全体の句には結び付かない場合が多い。意味の深い部分に立ち入れば，また，語の連想から，各語の意味を結合すれば全体が合成できるものもある。このようなものについて言語工学の面から資料を整理・収集し慣用表現について考えてみる。

2. 言語工学の観点からの問題点

慣用表現についてどのような問題点があるか考えてみる。

(1) 何を慣用表現と考えればよいか？

(2) 慣用表現を機械的に収集するにはどのようにすればよいか？

(3) 慣用表現は文の中にどの程度出現するのであろうか？

(4) 辞典，等で集められているものはどの程度あるのであろうか？

(5) 慣用表現を機械可読ファイルとしたときこれらはどのような分野に応用できるであろうか？どの程度有効であろうか？

(6) 慣用表現の訳語（英語）はどのようになるか？

(7) 他の言語（特に英語）の慣用表現と日本語を比較してみる。

(8) 慣用表現を一つの意味の塊としたとき，それ全体としてはどのような品詞成分を付ければよいのであろうか？

(9) 今までの構文解析は単語を主体としたものであったが，その枠組のままで構文解析は良いのであろうか？

(10) 慣用表現と他の語との共起関係はどのようになっているのであろうか？

(11) 慣用表現として別な意味を持つ場合と慣用表現中の個々の語を合成した意味で使われる場合とではどちらが多いか，それらの場合分けは何によって判別できるか？

例

鼻が高い。
- ① あの人は鼻が高いが目は小さい。
- ② むずかしい問題を解いたので鼻が高い。

これらは言語工学的な見方からの問題提起である。もっと解決しなければならない問題は多い。ここではその幾つかについての解決が述べればと考えている。

3. 慣用表現について

慣用表現には色々なものがある。ここでは大きく3種類にわけてみる。

(1) 慣用的に用いられている句で，これらを構成している一語一語を取りあげてもその意味を考えて句の全体の意味に結びつかないもの。

例　口が悪い，　首があぶない。

(2) 文章の中で，特に文末などに助詞，助動詞等が一定の表現をするもの。

例　～となるはずである。　～のようになっている。

これらは個々の言葉を解析しても意味を理解できるが，全体としてどのような使われ方をするか等について研究する必要がある。特に他の言語から日本語の末尾表現を合成するという点についての研究である。

(3) 自然言語の表現の中には，ある特定の現象や事象について特定の言葉を用いるというものがある。言葉の共起の強いものがある。

例　雨が降る。　影響を受ける。

このように共起の強いものまでも慣用表現と考えるのは問題があると考えるひともあろう。しかし，計算機で言語を処理するには語と語の共起についても多量に集めておかなければならない。

ここでは(1)を研究対象とする。

4. 慣用表現の収集

4.1 収集方法

慣用表現の収集方法としては大別して次の二つが考えられる。

(1) 既存の資料，本を参考にする。

(2) 新聞，小説，科学雑誌の中から慣用表現を収集する。

この二つの方法が考えられるが，(1)の方法を採用し，電子化されたファイル等により(2)の方法をコンピュータ処理により実験することを考えている。

4.2 慣用表現データ・ファイル

慣用表現を分析するためには次の二つのファイルを作成する必要がある。

(1) 慣用表現ファイル

このファイルは慣用表現と，その属性を入力する。

(2) 慣用表現例文ファイル

この例文ファイルは例文と慣用表現に対応するインデックスを持っている。

これら二つのファイルの具体的内容は次の項で述べる。

4.3 収集の基礎となる資料について

収集の基礎となる資料としては次の8種類の資料を対象とした。AからFまでの6種類のものは購入することが可能である。

A	必携慣用句辞典	三省堂
B	国語慣用句辞典	東京堂出版
C	擬声語擬態語慣用句辞典	東京堂出版
D	慣用句の意味と用法	明治書院
E	語源・慣用語	教育出版
F	国語慣用句大辞典	東京堂出版
G	慣用句の調査	国立国語研究所
H	日本語慣用句用例集	大阪大学文学部

表1　慣用表現資料

Gは国立国語研究所の内部調査報告書である。Hは大阪大学文学部の宮地　裕先生（昭和62年3月定年退官）が作成したもので限定印刷であり，借りてコピーするより方法がない。C，Fは作業の途中で判ったことであるが古典が多く省いた方がよいものであった。多くの資料を混ぜると不統一をまねき，混乱する面もある。しかし，一方で各資料の不備をうめてくれるため，よりよいファイルになる。

AからFまでの資料を複写し，この中から入力部分に赤線を引き，次の項目を入力した。

① 登 録 No. （数　字）
② 資料記号 （A～H）
③ Page No. （数　字）
} 出 典
④ 慣用表現 （漢　字）

慣用表現の出典にさかのぼることができるようにした。このようにして出来たファイルのデータ件数は26,939件であった。出典は異なるが同一の慣用表現がある，これらをまとめると約22,000件になった。

さらに，慣用表現ファイルに次のような付加情報を付けた。

① 読み仮名
② 自立語の見出し（検索のため）　　例　手を出す。 → 手，出す
③ 品詞，又は品詞相当のもの　　語幹の抽出，活用付け
　例　黒山を築く → 黒山を築:く　（動詞，カ行，五段）

さらに，入力時に発生した各種誤りを除去した。

次に，以下のような項目を入力しなければならないと考えている。

① 分かち書きの区分
② 文，句の区分（慣用表現の種類）
③ 置換可能語
④ 置換にあたって文字列操作 (1) 前
⑤ 置換にあたって文字列操作 (2) 後
⑥ 訳　語
⑦ 共起語
⑧ その他の区分

置換可能語は慣用表現をできるだけ簡単な表現に書き変えるためのものである。

例　骨身を削る　　→ 苦労をする
　　どしゃぶりの雨 → はげしい雨

慣用表現を置換可能な表現に変えた場合，追加したり，削除する文字があるか，否か，あればその手順は常に一定か，例文で検証する。不自然であれば別の置換可能な語に変える。

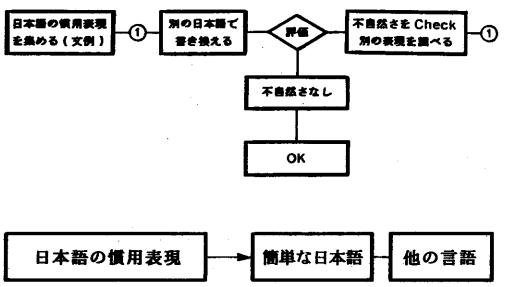


図1　慣用表現について

1. 登 録 No.　（SEQ. No.）
2. 慣用表現
3. 読　み
4. 分かち書き
5. 自立語の見出し
6. 文，句の区別
7. 品詞，又は品詞相当のもの
8. 活　用
9. 語　尾
10. 置換可能語
11. 置換にあたっての文字列操作 1（前）
12. 置換にあたっての文字列操作 2（後）
13. 訳　語
14. 共起語
15. その他区分

表2　慣用表現ファイルの項目

1. 登 録 No.
2. 例　　文
3. 慣用表現ファイルの対応No.
4. そ の 他

表3　慣用表現例文ファイルの項目

- **慣用表現例文ファイル**

  慣用表現を考えるにあたっては，慣用表現例文ファイルを考えなければならない。慣用表現例文ファイルの項目としては次のものが考えられる。

  ① 登 録 No.　② 慣用表現（見出し）　③ 例　文　④ そ の 他

  入力した例文のデータ数 5,974文である。

  （A…… 3,542文，　D…… 1,507文，　E…… 670文，　G…… 255文）

5. 慣用表現ファイルの整理と慣用表現の同定

慣用表現は幾つかの語が集まり一つの意味を表現するため，単語よりも長くなり，各種の表記のゆれが発生する。

例 1.

表記のゆれ	頻度		
地団駄を踏む	5	→ 地団太をふむ	10
地団太を踏む	1		
地団太をふむ	1		
地だんだを踏む	1		
じだんだを踏む	1		
じだんだをふむ	1		

例 2.

表記のゆれ	頻度		
嫌気がさす	1	→ 嫌気がさす	5
いや気がさす	2		
嫌気が差す	1		
イヤケがさす	1		

表 4　慣用表現の表記のゆれ

（資料Hより抽出した内容の一部）

また，"地団太をふんだ"というように語尾が活用するものがある。このため，語幹の抽出，活用の種類を見付けておかなければならない。

慣用表現を一般文の中から見付け出し，同定するためには表記のゆれのフィルタを通し処理しなければならない。この内容をまとめると次のようになる。

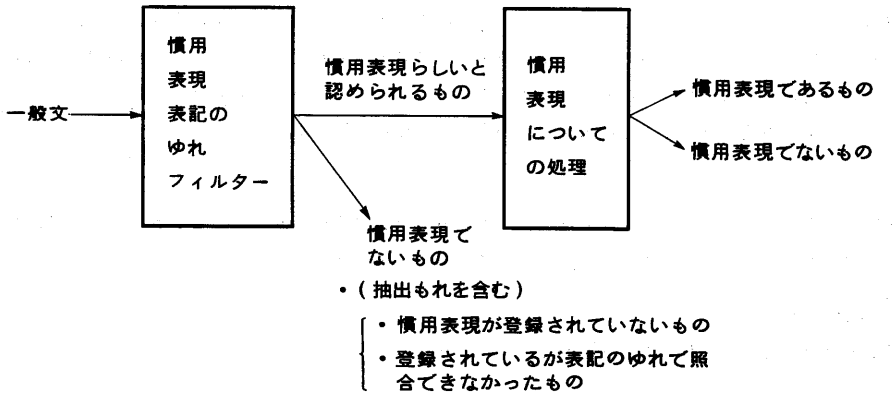


図 2　慣用表現の処理

慣用表現として同定するにあたっても文の内容を検討しなければならないものがある。これは慣用表現と他の語の共起関係によって判別することができるのか等，同定する方法を考えなければならない。

例　鼻が高い
- あの人は鼻が高いが目は小さい。
- むづかしい問題を解いたので鼻が高い。

可能な判別方法や，慣用表現と語の共起関係を調べることが重要である。

6. 慣用表現ファイルとヒューマン・インターフェイス

慣用表現ファイルの項目としては機械処理を中心としたことを考えてきたが，このほかに人間が理解することや理解を促進する内容を追加しなければならない。

機械辞書
- 機械処理のための項目
- 人間が理解し，解決を深めより適切な翻訳等の作業が行えるための項目

Pre-edit，Post-editのために画面を通して会話をする必要がある。このために次のような項目を追加している。

(1) 意　　味 ： 慣用表現の意味を表わすもの
(2) 使 用 条 件 ： 慣用表現が使用できる環境
(3) 類　義　語 ： 慣用表現とほぼ同じ意味の語，一部置換語としても入力している
(4) 反義語，否定語 ： 慣用表現と反対の意味を持つ語，否定の意味を持つ語
(5) 語　　源 ： 語の出典，語の持っている歴史的用例や意味
(6) 誤った利用 ： 語の意味を誤って使用している場合などの説明
(7) キーワード ： 慣用表現中に含まれる自立語は既に抽出できるように考えているが，さらに関連する用語を入力する。

このような内容を，慣用表現ファイルの項目として追加することにより，さらに利用し易くすることができる。

7. 慣用表現ファイルの応用分野

二つの慣用表現ファイルを作ることにより，次のような応用分野が広がる。

(1) 機械翻訳への利用

慣用表現を一つの語と同じように扱うことにより慣用表現に使われている語の意味の多義性を減らすことができる。機械翻訳はマニュアルの翻訳とか，科学技術文献速報の翻訳といったものからの要求が強いが，今後は一般的な文章の翻訳に利用されるであろう。このようになると慣用表現の処理は重要な意味を持ってくる。
また，慣用表現例文ファイルは機械翻訳の例文として使用可能である。

(2) 研究の重複の削除

研究初期作業としてのデータを集める作業が省ける。研究の準備が簡単になる。多くの研究者が新しい分類区分を追加することにより，よりよいファイルになる。

(3) 日 本 語 教 育

外国人に対する日本語教育の教材として使うことができる。

(4) 国語等の研究者へデータを提供する。

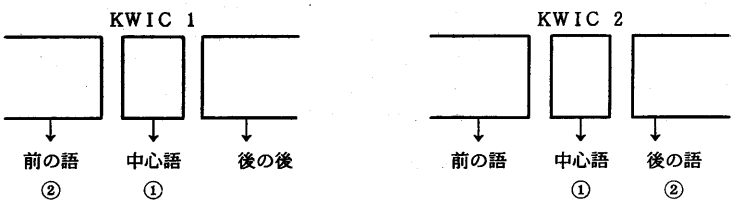
言語を研究している人々へ研究データを提供することが出来る。

8. 今後の検討課題

慣用表現について日本語からの研究であったが，他の言語，特に英語の面から検討してみる必要がある。英語にも慣用表現や共起関係の強い表現，一定の言い方等がある。これらをうまく取り出し，整理する必要がある。

英語の慣用表現を自動抽出するためには単語によるKWICを作成し，それをもとに考えるべきである。連続する2単語のKWICや連続する単語のKWICと頻度情報等により，慣用表現らしいものを簡単に抽出することができる。（冠詞a，an，theは省いて考える。）

KWICも中心語と直前の語で分類したもの，中心語と直後の語で分類したもの等を作成すると抽出しやすいものができる。中心語は1単語以上である。



①，②は分類の順序

慣用表現ファイルの内容を充実させることも重要であるが，慣用表現がどの程度一般文章の中に含まれているか，どの慣用表現が多く使われているか等について定量的に分析する必要がある。

さらに，次のようなことも考えなければならない。慣用表現は形態素解析でだけ利用されるのではなく，慣用表現を一つの意味の魂と考えると，単語と品詞による構文解析だけでは処理できず，次に意味処理の構文解析（Semantic Parser）が必要になるのでなかろうか？　この面の研究も今後活発になることを期待する。

9. お わ り に

慣用表現についての研究は開始したばかりであるが，少しづつデータを整備したいと考えている。今後の成果に期待していただきたい。

また，この研究の一部は文部省科研費課題番号（60302090）（代表者 吉田　将）と文部省科研費課題番号（61880005）（代表者 藤崎博也）と文部省科研費特定研究言語（代表者 長尾　真，A04班長 野村雅昭）によって行った。

また，データ整理の一部は香川大学 土屋信一教授と研究室のみなさんの協力を得て行った。感謝の意を表す。

文献 page No.	慣用表現見出	慣用表現例文
A 15	頭に来る	「真夜中のいたずら電話は全く頭に来る」
A 15	頭の上の蠅を追え	「他人のことより，まず自分の頭の上の蠅を追え」
A 15	頭の黒い鼠	「これは頭の黒い鼠がやったしわざだ」
A 15	頭を痛める	「不景気で資金繰りが思うようにいかず，頭を痛めている」
A 15	頭を抱える	「子供の結婚問題で頭を抱えている」
A 15	頭を掻く	「頭を掻いてばかりいないで，後の始末を考えろ」
A 15	頭を切り替える	「我々年寄もこの辺で頭を切り替えないと，若い人たちに取り残されてしまう」
A 16	頭を下げる	「頭を下げて頼む」
A 16	頭を下げる	「強がり言っていないで，素直に頭を下げたらどうだ」
A 16	頭を絞る	「いくら頭を絞っても，いい知恵が浮かばない」
A 16	頭をはねる	「彼は下請業者に払う代金の頭をはねて，競馬につぎこんでいたそうだ」
A 16	頭を拈る	「この機械を作る時にいちばん頭を拈ったのは，この自動制御の部分だ」
A 16	頭を冷やす	「自分のしたことを，頭を冷やしてよく考えなさい」
A 16	頭を丸める	「失敗したら頭を丸めるぐらいのつもりでやれ」

慣用表現例文ファイル

文献 page No.	慣用表現見出し	seg No.
A-043	嘘の皮が剝がされる	10912
A-043	嘘も方便	10913
I-038	嘘をつく	10914
A-043	嘘を固める	10915
A-043	嘘八百を並べる	10916
A-045	鰻の寝床	10917
A-049	瓜二つ	10918
F-068	運がいい	10920
H-064	運がつく	10921
F-068	運がよい	10923
H-064	運がわるい	10924
F-068	運が悪い	10926
H-081	運が開ける	10927
F-068	運が強い	10929
F-068	運が尽きる	10931

慣用表現見出しファイルの例

参　考　文　献

(1) 田中康仁　吉田　将　自然言語における知識データについて　情報処理学会第31回（昭和60年後期）全国大会

(2) 田中康仁　吉田　将　慣用表現について　情報処理学会第32回（昭和61年度前期）全国大会　ＩＳ－３

(3) 田中康仁　吉田　将　慣用表現について　－収集と整理－　情報処理学会第34回（昭和62年前期）全国大会　IX-1

(4) 田中康仁　吉田　将　自然言語の分析による知識データ　情報処理学会自然言語処理研究会　54-3　1986.3

(5) 村木新次郎　慣用句・機能動詞結合・自由な語結合　日本語学第４巻第１号　1985.1　明治書院

(6) 宮地　裕　"慣用句の意味と用法"　明治書院　1982.10

(7) 倉持・阪田　"必携慣用句辞典"　三省堂　1985.1

(8) 高木一彦　"慣用句研究のために"　教育国語 38　麦書房　1974.9

(9) 覆刻／文化庁国語シリーズ　語源・慣用語　教育出版株式会社　1975.4

(10) 白石大二編　国語慣用句辞典　東京堂出版　1984.3

(11) 白石大二編　擬声語／擬態語　慣用句辞典　東京堂出版　1982.4

(12) 白石大二編　国語慣用句大辞典　東京堂出版　1983.8

(13) 宮地　裕　日本語慣用句用例集　大阪大学文学部（限定出版物）　1985.3

(14) 伊藤菊子，小沢厚子他　慣用句の調査　国立国語研究所補助員研修慣用句班　1981.3

(15) 日本語処理技術に関する調査研究　61-C-535　(社) 日本電子工業振興協会

(16) 故事俗信　ことわざ大辞典　小学館

(17) Adam Makkai　IDIOM STRUCTURE IN ENGLISH　MOUTON　1972

(18) Wolfgang Hleischer　Phraseologie der deutschen Gegenwartssprache　VEB Bibliographisches Institut Leipzig　1982
（この資料はドイツ語で書かれている。）