

基本慣用句五種対照表の作成

佐 藤 理 史

名古屋大学大学院工学研究科

2 種類の小学生用国語辞典、小学生用慣用句辞典、および、2 種類の慣用句辞典、の計 5 種類の資料に、どのような慣用句が掲載されているかを調べ、その結果を整理して、「基本慣用句五種対照表」を作成した。本稿では、その概要と作成手順、および、その過程で明らかになった慣用句の見出し形の揺れについて述べる。

Compilation of a Comparative List of Basic Japanese Idioms from Five Sources

SATOSHI SATO

Graduate School of Engineering, Nagoya University

We have compiled a comparative list of basic Japanese idioms from five sources; one idiom dictionary and two Japanese dictionaries for children in elementary school and two standard idiom dictionaries. This paper describes the format of the comparative list, the compilation process, and an interesting phenomenon of unstable standard forms of idioms, which we found in the compilation process.

1. はじめに

我々は、母国語を話すとき、特に意識することなく、多くの慣用句を使いこなしている。しかし、いったいどのくらいの慣用句を使いこなしているのか、あるいは、厳密にはどのようなものを慣用句と呼ぶのか、といった疑問を突き付けられたならば、それに明確に答えることができる人はほとんどいないであろう。

日本語慣用句辞典¹⁾では、「慣用句の定義はいまだに決定的なものがない」と断った上で、おおよそ一般的な共通理解としては、

「単語の二つ以上の連結体であって、その結びつきが比較的固く、全体で決まった意味を持つ言葉」である

と述べている。大辞林第三版²⁾では、「慣用句」を次のように説明している。

- (1) 二語以上が結合し、その全体が一つの意味を表すようになって固定したもの。「道草を食う」「耳にたこができる」の類。慣用語。イディオム。
- (2) 二語以上が、決まった結びつきしかない表現。「間髪を入れず」「悦に入る」の類。慣用語。イディオム。

筆者が特に興味を持っているのは、次のことである。

- (1) 日本語の慣用句のうち、誰もが知っているような「基本慣用句」はいくつぐらいあるのか。そのリストに含まれるのは、どのような慣用句か。
- (2) もし、1 万語の基本語彙表を作成するとするならば、そこには、いくつぐらいの慣用句を含めるべきか。

筆者は、慣用句や複合辞といった複合形式は、日本語を構成する基本的な要素(プリミティブ)として捉えるべきものであり、基本語彙表に含めるべきものだと考えている。そのような考え方に基づいて、慣用句や複合辞を含んだ日本語の基本語彙表(基本要素表と呼ぶ方が適切かもしれない)を作成することを目指している。慣用句に関しては、そこに含めるべき基本慣用句を選定することが研究の一つの目標となる。

そのような基本慣用句の選定のための基礎資料として、五種類の資料にどのような慣用句が掲載されているかを調べ、その結果を整理した基本慣用句五種対照表(以下、五種対照表と呼ぶ)を作成した。本稿では、この五種対照表の概要と、その作成の手順、および、そこから学んだことについて報告する。

以下、まず 2 節で、使用した五種類の資料について説

明する。次に3節で、作成した五種対照表の構成について述べる。4節ではデータの入力について、5節では入力したデータからの五種対照表の編纂について述べる。6節では、慣用句の見出しが揺れる原因について整理する。7節では編纂作業を通して学んだことを整理し、8節で全体をまとめる。

2. 使用した五種類の資料

五種対照表で比較した資料を以下に示す。

- (1) 〈レインボー〉＝金田一春彦、金田一秀穂 監修。新レインボー小学国語辞典 改訂第3版。学研、2005。
小学生用の国語辞典の一つ。収録語彙3万5千語。成句等もすべて見出しとして立項されている。慣用句・ことわざ・四字熟語・故事成語に該当する項目は、そのことが明示されている。これに該当する項目数は、我々の調査では2585項目。
 - (2) 〈宮地〉＝宮地 裕 編。慣用句の意味と用法。明治書院、1982。
日本語の慣用句について書かれた数少ない専門書の一つ。巻末に、1280項目から成る常用慣用句一覧が示されている。
 - (3) 〈米川〉＝米川明彦、大谷伊都子 編。日本語慣用句辞典。東京堂出版、2005。
宮地に学んだ米川・大谷の両名によって編纂された新しい慣用句辞典。「慣用句」を比較的厳格に捉えている。公称1563句。
 - (4) 〈まんが〉＝金田一秀穂 監修。小学生のまんが慣用句辞典。学研、2005。
小学生用の小さな慣用句辞典。第1章に132句、第2章に204句、合わせて336句を収録している。
 - (5) 〈小学館〉＝金田一京助 編。小学館学習国語新辞典 全訂第二版。小学館、2006。
収録語彙1万9千語の小学生用国語辞典(低学年用)。成句等もすべて見出しとして立項されている。慣用句等の種別は明示されていない。
- これらの資料を選んだ理由は、次のとおりである。
- ・〈レインボー〉は、我々が調査した小学生用国語辞典の中で、唯一、慣用句等の種別が明示されていた辞典である。種別が明示的に示されているので、その種別に従って慣用句等を拾うことができる。
 - ・〈小学館〉は、我々が調査した小学生用国語辞典の中で、収録語彙が最も少ない辞典である。ここに掲載されている慣用句は、基本的な慣用句とみなすことができる。但し、この辞典には慣用句等の種別が明示されていないので、直接、この辞典から慣用句等を安定して拾うことはできない。故に、リストとの照合調査(与えられたリストに含まれる慣用句等がこの辞典に掲載されているかどうかを調べること)に用いる。
 - ・〈宮地〉は、数少ない慣用句の専門書であり、そこに

掲載されている常用慣用句一覧は、この種の表として、おそらく唯一のものである。

- ・〈米川〉は、〈宮地〉の後をうけた最新の慣用句辞典である。
- ・〈まんが〉は、小学生向けの慣用句辞典である。この辞典に掲載されている慣用句は、基本的な慣用句とみなすことができる。

結果的に、小学生用の国語辞典2冊、専門的な慣用句辞典2冊、小学生向けの慣用句辞典1冊という内訳となった。このうち、〈宮地〉を除く4冊は、2005年以降に出版された新しい資料である。

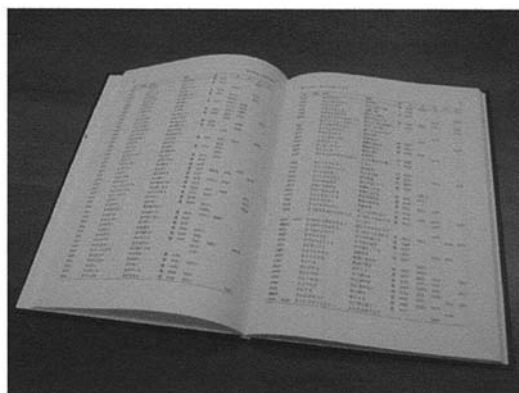


図1 作成した五種対照表

3. 五種対照表の構成

作成した五種対照表(図1)は、それぞれの慣用句等(項目と呼ぶ)に対する一連のデータ(レコードと呼ぶ)の集合として構成されている。最新版(v0.93 - 2007年1月18日)のレコード数は、3629である。

3.1 レコードの構成

1つのレコードは、次の10フィールドから構成されている。

- (1) ID (五種対照表 ID)
4桁の数字。
- (2) 参照 ID
この項目が別の項目の異形と考えられる場合に値を持つ。値は、その別項目のID。詳細は、5.2節を参照のこと。
- (3) 見出し
次節(3.2節)で述べる見出し表記規則に従って記述される文字列。
- (4) 代表表記
この項目の代表的な表記。五種類の資料のいずれかの表記が採用されている。代表表記決定規則は、5.1節で述べる。
- (5) 種別

表 1 五種対照表の一部

ID	参照	見出し	表記	種	レ	宮	米	ま	小
0001		ああいえばこういう	ああ言えはこう言う	慣	0001				
0002	0005	あいそがつきる	愛想が尽きる				0001		
0003	0007	あいそをつかす	愛想を尽かす				0002		
0004		あいそがいい	愛想がいい	慣	0002				
0005		あいそがつきる	愛想が尽きる			0002.2			
0006		あいそをいう	愛想を言う	慣	0003	0001			
0007		あいそをつかす	愛想を尽かす	慣	0004	0002.1			0001
0008		あいそをふりまく	愛想を振りまく	慣	0005				
0009		あいたくちがふさがらない	開いた口がふさがらない	慣	0006	0004	0003	0001	0002
0010		あいづちをうつ	相づちを打つ	慣	0007	0003	0004	0002	0003

慣用句・ことわざ・四字熟語・故事成語の別。この項目が〈レインボー〉に掲載されている場合に値を持つ。

- (6) 〈レインボー〉での掲載の有無
- (7) 〈宮地〉での掲載の有無
- (8) 〈米川〉での掲載の有無
- (9) 〈まんが〉での掲載の有無
- (10) 〈小学館〉での掲載の有無

五種対照表の一部を表 1 に示す。

3.2 見出し表記規則

五種対照表の見出しは、次の規則に従って記述されている。

規則 1 和語・漢語はひらがなで表記する。

規則 2 外来語はカタカナで表記する。

規則 3 ひらがな、カタカナ以外の文字は使用しない。

3.3 見出しの同一性と項目の立項

見出しの同一性と項目の立項は、次の規則に従う。

- (1) 見出しの同一性は、見出し表記規則に従って記述された見出し文字列の同一性によって判定する。
- (2) 異なる見出しに対しては、かならず別の項目を立てる。

なお、同一見出しに対しては、複数の項目を立てる必要がある可能性が残るため、同一見出しに対して項目をひとつしか立てないという条件は、規則化しない。

但し、現時点では、同一見出しに対して複数の項目を立てる必要があるものは見つからない。

3.4 各資料での掲載の有無

各資料での掲載の有無は、その資料にその項目が掲載されている場合は、その資料における、その項目の ID(資料別 ID) が記述されている。掲載されていない場合は、値を持たない。

4. 各資料に基づくデータの入力

データの入力は、すべて手作業で行なった。各資料に基づいて作成したデータのレコード形式とレコード数を表 2 にしめす。

4.1 〈レインボー〉データの入力

この辞典に掲載されている項目のうち、慣用句・ことわざ・四字熟語・故事成語と明示されている項目を拾い、次のような手順で電子化した。

- (1) 該当する項目に対して、辞書に記述されている次の 4 つの情報を入力した。

- (a) 見出し。
- (b) 見出しが赤字で記述されているか否か。重要語は赤字で印刷されている。
- (c) 表記。漢字表記がある場合に記述されている。記述されていない場合は、見出しをそのまま表記とした。
- (d) 種別。慣用句・ことわざ・四字熟語・故事成語のいずれか。

- (2) 一つの辞書項目に対して、1 レコードを作成した。但し、「いきもつかず(に)」のように、辞書の見出しが複数の見出しをまとめた形式であり、3.2 節で述べた見出し表記規則に違反する場合は、それらを展開して、複数のレコードを作成した。

- (3) 辞書の各項目に対して、辞書の出現順に 4 桁の ID を付与した。辞書項目から 1 レコードを作成した場合は、項目の ID をそのままレコードの ID とした。辞書項目から複数のレコードを作成した場合は、小数以下の枝番号を追加して区別した。これをレインボー ID と呼ぶ。

4.2 〈宮地〉データの入力

「慣用句の意味と用法」に掲載されている常用慣用句一覧(pp267-285)を次のように電子化した。

- (1) 常用慣用句一覧は、1280 項目から成る。それぞれの項目には、ひとつ、または、複数の慣用句(意味的に関連した慣用句群)が記述されている。
- (2) 各慣用句に対して明示的に記述されている情報は、表記のみである。「慣用句の意味と用法」の本文で取り上げられている慣用句は、太字で印刷されており、それ以外の慣用句は、通常の字体で印刷されている。
- (3) 常用慣用句一覧に含まれる全ての慣用句に対して、その表記を入力した。
- (4) 各慣用句に対して、五種対照表の見出し規則を満たす見出しを作業者が付与した。表記にルビが振っており、読みが明示されているものは、それに従った。それ以外のものは、作業者の判断によった。
- (5) 常用慣用句一覧の各項目に、常用慣用句一覧での出現順で、4 桁の数字の ID を付与した。項目が一

表 2 資料別に入力したデータのレコード形式とレコード数

資料	レコード数	フィールド				
		1	2	3	4	5
〈レインボー〉	2590	レインボー ID	見出し	表記	赤字?	種別
〈宮地〉	1546	宮地 ID	見出し	表記	太字?	
〈米川〉	1566	米川 ID	見出し	表記		
〈まんが〉	336	まんが ID	見出し	表記	何章?	
〈小学館〉	873	小学館 ID	見出し	表記		

つの慣用句しか含まない場合は、項目の ID をそのまま慣用句の ID とした。項目が複数の慣用句を含む場合は、小数以下の枝番号を追加して区別した。これを宮地 ID と呼ぶ。

4.3 〈米川〉データの入力

「日本語慣用句辞典」に収録されている慣用句をデータ化した。

- (1) この辞書の見出しをすべて拾い、データ化した。見出しは表記で示されており、ルビにより読みも示されている。この 2 つの情報を、前者を表記、後者を見出しとして入力した。
- (2) 一部、旧字等の理由で入力できなかった漢字は、ひらがなで代用した。
- (3) 入力した慣用句を辞書順にソートし、4 桁の ID を付与した。これを米川 ID と呼ぶ。

4.4 〈まんが〉データの入力

「小学生のまんが慣用句辞典」に収録されている慣用句を目次に基づいてデータ化した。

- (1) この辞書には、第 1 章に 132 句、第 2 章に 204 句、合わせて 336 句が収録されている。第 1 章の方が、より基本的な慣用句という位置付けである。
- (2) 目次では、各項目に対して、表記が明記されると同時に、ルビという形で読みが明記されている。
- (3) 各慣用句に対して、見出し(読み)、表記、何章に記述されているか、の 3 つのデータを入力した。
- (4) 入力した慣用句を辞書順にソートし、4 桁の ID を付与した。これをまんが ID と呼ぶ。

4.5 〈小学館〉データの入力

上記の 4 種類の資料から収集した項目が、この辞書に掲載されているかどうかをデータ化した。

- (1) 収集した項目の見出しと同一の項目が辞書に掲載されている場合は、辞書に記述されている表記を入力した。
- (2) 収集した項目の見出しと同一の項目は辞書に掲載されていないが、その異形と考えられるものが掲載されている場合は、それを新たな項目として拾い、見出しと表記を入力した。
- (3) こうして得られたデータを辞書順にソートし、4 桁の数字の ID を付与した。これを小学館 ID と呼ぶ。

5. 五種対照表の編纂

五種の資料に基づいて入力したデータを、見出しをキーにして結合した後、代表表記を決定し、新たに ID(五種

対照表 ID) を付与した。最後に、以下で説明する参照 ID と呼ぶフィールドを追加した。こうして完成したのが、五種対照表である。

5.1 代表表記の選定規則

代表表記は、五種類の資料に次の優先順序を定め、その項目が掲載されている資料のうち、優先順序の最も高い資料で採用されている表記を、代表表記として選定した。

〈レインボー〉>〈小学館〉>〈まんが〉
>〈米川〉>〈宮地〉

5.2 参照 ID の付与

五種対照表を作成する過程で、我々は、それぞれの辞書において、慣用句の見出しが「揺れている」ことに気付いた。すでに、3.2 節で述べたように、五種対照表では、ひらがな・カタカナで表記される文字列を見出しとして採用した。つまり、ここで観察される揺れは、いわゆる「表記の揺れ」ではなく、「一つの慣用句が異なる形の見出しとして立項される」ことを意味する。

この現象に対処するために、次のことを行なった。

- (1) 一つの慣用句に対して、代表項目(主見出し)を一つ定める。
- (2) 代表項目として採用しない項目(つまり、他の項目の異形と考えるもの)に対して、対応する代表項目へのリンクを付与する。
- (3) 具体的には、各レコードに参照 ID と呼ぶフィールドを追加し、そのレコードが、別レコードの項目の異形と考えられる場合に、その別レコードの五種対照表 ID を記述した。

なお、上記の参照 ID 付与に必要な判断は、編者が行なった。

最終的に、3629 レコードのうち、3397 レコードを代表項目として選択し、残りの 232 レコードに対して参照 ID を付与した。

5.3 掲載資料数の集計

五種対照表のそれぞれの項目(レコード)が、いくつの資料に掲載されているかを数えた。また、見出しのゆれを考慮した場合についても、同様の調査を行なった。それらを集計した結果を表 3 に示す。

6. 見出しが揺れる原因

ここでは、見出しが揺れる現象の原因を整理する。それらは、大きく、狭義の異形と標準化の程度の 2 つに分類することができる。

表 3 レコード別・慣用句別の資料掲載数の集計

資料数	揺れを考慮しない場合		揺れを考慮した場合	
	レコード数	累計	慣用句数	累計
5	121	121	133	133
4	323	444	360	493
3	485	929	491	984
2	859	1788	829	1813
1	1841	3629	1584	3397

6.1 狭義の異形

慣用句は、句としての異形を持ちうる。これは、音韻的異形、機能語の交替、内容語の交替の3つに分類できる。

6.1.1 音韻的異形

数は少ない。

例：愛想(あいそ／あいそう)を尽かす

泣き面にはち／泣きっ面にはち

6.1.2 機能語の交替

通常の句に見られる機能語の交替現象と同様の現象が、一部の慣用句においても見られる。

(1) 助詞の交替

例：かけがえがない／かけがえのない、
目から鼻に抜ける／目から鼻へ抜ける、
水と油／水に油、
声を限りと／声を限りに

(2) 「も」の有無

例：足の踏み場がない／足の踏み場もない、
風上に置けない／風上にも置けない、
数限りない／数限りもない

(3) その他の機能語の交替

例：穴があつたら入りたい／穴があればはいりたい

(4) その他

例：手ぐすねを引く／手ぐすね引く、
目を皿のようにする／目を皿にする、
いやがおうでも／いやでもおうでも

6.1.3 内容語の交替

慣用句は比較的安定していると考えられているが、一部の慣用句において、慣用句を構成する内容語が、同義語・類語と交替する現象が観察される。

(1) 体言の交替

例：芋の子を洗うよう／芋を洗うよう、
ありのはい出るすき間もない／ありのはい出るすきもない、
上を下への大騒ぎ／上を下への大騒動／上を下への騒ぎ

(2) 用言の交替

例：大きい顔をする／大きな顔をする、
骨を埋(うず)める／骨を埋(う)める、
胸をときめかす／胸をときめかせる、
取るに足らない／取るに足りない、
目にも見せてくれる／目にも物見せる

6.2 標準化の程度

「血を分けた息子」のように、タ形の「血を分けた」が

標準的な使用法であるとき、

- (1) これをそのまま見出し形とする
- (2) 「分けた」を「分ける」に標準化して、「血を分ける」という形で立項する

という選択肢がある。

また、「お金を湯水のように使った」のような用例から、

- (1) 標準的な使用法である「湯水のように使う」を見出しとして採用する
- (2) 「湯水のように使用する／散財する」という表現もありえるので、「湯水のように」を見出しとして採用する

という選択肢がある。

このように、どこまでの範囲を慣用句と見なし、かつ、その見出し形をどのように設定するか(どこまで標準化するか)ということにおいて、辞書編纂者の自由度が存在する。これが、見出しの揺れとして現れる。

ここでは、このような現象を、活用形・語形に関わるものと、部分／全体と考えられるものの2種類に分けて整理する。

6.2.1 活用形・語形の標準化に関わるもの

(1) 活用形

例：血を分けた／血を分ける、
間が抜けた／間が抜ける、
念には念を入れよ／念には念を入れる、
年甲斐もない／年甲斐もなく

(2) 「ない／ぬ／ず」

例：押しも押されもしない／押しも押されもせぬ、
目にもとまらない／目にも留まらぬ、
骨身を惜しまない／骨身を惜しまず

(3) 「に」の有無

例：息もつかず／息もつかずに、
手に取るよう／手に取るように、
否応無し／否応無しに、
藪から棒／藪から棒に

(4) 「よう」の有無

例：雲をつかむ／雲をつかむよう、
顔から火が出る／顔から火が出るよう、
猫の額／猫の額よう

(5) 末尾の助詞の有無

例：右から左／右から左へ、
のどから手が出る／のどから手が出るほど

6.2.2 部分／全体

(1) 体言の有無

例：蚊の鳴くよう／蚊の鳴くような声、
身を切られる／身を切られる思い、
願ってもない／願ってもないこと

(2) 用言の有無

例：穴のあくほど／穴のあくほど見る、
大きな顔／大きな顔をする、
首を長くする／首を長くして待つ、

そばづえ／そばづえを食う、
減らず口／減らず口をたたく、
湯水のように／湯水のように使う、
地獄で仏／地獄で仏に会ったよう

(3) その他

例：あごで使う／あごで人を使う、
右に出る／右に出る者がいない、
あとは野となれ／あとは野となれ山となれ

7. 考 察

五種対照表の作成を通して学んだことは、次のことである。

(1) 慣用句の判定は難しい。

慣用句をそれ以外の句と区別することが難しいことは、あらかじめ予想していた。概念的に慣用句を定義することには困難が付きまとうし、ことわざ・故事成句といった成句と区別することもそれほど容易ではない。コロケーションと慣用句の間に明確な境界線を引くことは、おそらく不可能であろう。五種対照表の作成では、この問題を回避するために、資料に慣用句等の種別が明示されている〈レインボー〉と、慣用句辞典とうたっている〈宮地〉、〈米川〉、〈まんが〉から慣用句等を拾う方法をとった。〈レインボー〉からは、意図的に、ことわざ・四字熟語・故事成句も拾ったが、〈レインボー〉では「ことわざ」と明記されているものが、比較的厳密に「慣用句」を捉えている〈宮地〉や〈米川〉に掲載されている例もあった。これらの観察より、慣用句を他の句と厳格に区別しようとするよりも、すこし広く「慣用句等の成句」と捉え、それらを収集していく方向の方がよいとの考えに至っている。

(2) 想像以上に見出しが揺れている。

五種対照表の作成の過程で一番驚いたことは、予想以上に見出しが揺れているという事実である。数で言えば、3397の慣用句等のうち、213(6.3%)で見出しが揺れていた。原因は、すでに6節で分析したとおりであるが、表現が固定化されているはずの慣用句が、辞書の見出しというレベルで揺れるという事実は、新たな発見であった。この事実から次のことが推測できる。

- 辞書編纂者の誰もが参照し、拠り所とするような、標準的な慣用句辞典は存在しない。
- 慣用句の見出し形を決定するための確立した手続きは存在しない。

(3) 同一慣用句の判断は難しい

見出しが揺れていることも関連するが、同一慣用句の異形と考えるべきか、あるいは、別の慣用句と考えるべきかの判断は、しばしば難しいものとなる。いわゆる自他の対応は、基本的に別慣用句とみなしたが、「詰め腹を切らせる／切られる」のようなものは判断に迷う。また、否定の有無の違いは基本的に同じ慣用句とみなしたが、「合点がいかない／いく」のような否定形の方をよく使うものの一部は、別慣用句と判断したものもある。同

一慣用句かどうかを判定するための、いくつかの原則を立てることはできるが、最終的には、個別に言語直観に従って判断する必要があると考える。

(4) 作成したリストは、もれはないとは言いきれない。

基本的な慣用句が掲載されていると考えられる四種類の資料から慣用句等を拾ったので、日本語の基本的な慣用句と考えられるもののほとんどは、このリストに収録できているのではないかと考えている。しかしながら、これまで述べていたとおり、何を慣用句と考えるのか、どこまでを同一慣用句と考えるのか、といった判断は揺れるため、いくつかの基本的な慣用句がこのリストから洩れているという可能性は否定できない。

(5) 語彙表を作成するとき、慣用句等の割合は、全体の5%~8%ぐらいか。

この五種対照表に基づいて、日本語の基本語彙に含めるべき基本慣用句の選定を行なう予定であるが、そこでの出発点は、いったいいくつぐらいを目処に選定していけばよいのかを決めることである。収録語彙3万5千語の〈レインボー〉から拾った慣用句等の数は2585であり、これは7.4%に当たる。また、収録語彙1万9千語の〈小学館〉で照合できた慣用句等の数は873であり、これは4.6%に当たる。〈小学館〉からは、すべての慣用句等を拾ったわけではないことを加味すると、語彙数に対する慣用句等の割合は、5%~8%程度と見積ることができる。この見積りを語彙表の設計に採用するならば、1万語の語彙表には、500~800程度の慣用句等を含めることになる。

8. おわりに

本稿では、基本慣用句五種対照表の概要とその作成手順、および、そこから明らかになった見出しの揺れについて述べた。今後、この五種対照表に基づいて、日本語の基本語彙に含めるべき基本慣用句の選定を行なう予定である。

謝辞 本研究は、次の研究費の支援を受けて実施した：受託研究「円滑なコミュニケーションのための言語処理基盤に関する基礎研究」（2006年度、独立行政法人 情報通信研究機構より受託）

参 考 文 献

- 1) 米川明彦, 大谷伊都子 (編): 日本語慣用句辞典, 東京堂出版 (2005).
- 2) 松村明 (編): 大辞林第三版, 三省堂 (2006).