

## 日本語慣用句コーパスの構築と慣用句曖昧性解消の試み

橋本 力<sup>†</sup> 河原 大輔<sup>††</sup>

<sup>†</sup> 山形大学大学院理工学研究科 〒992-8510 山形県米沢市城南 4-3-16

<sup>††</sup> 情報通信研究機構 〒619-0289 京都府相楽郡精華町光台 3-5

E-mail: <sup>†</sup>ch@yz.yamagata-u.ac.jp, <sup>††</sup>dk@nict.go.jp

**あらまし** 慣用句には、慣用句の意味だけでなく文字通りの意味も表しうるといった意味的曖昧性がある。我々は、慣用句の曖昧性解消技術の確立に向けて、日本語慣用句コーパスを構築している。本稿では、コーパスの現状とそのコーパスを用いた慣用句曖昧性解消実験の結果について報告する。我々の慣用句コーパスは曖昧性のある基本的な慣用句146句を対象としており、全体で113,460用例からなる。各用例は、それに含まれている慣用句相当文字列が慣用句として用いられているのか、あるいは文字通りの意味で用いられているのかを示すラベルが人手で付与されている。用例は全てWebコーパスから収集した。本コーパスは近日公開予定である。慣用句曖昧性解消実験では、機械学習に基づく標準的な単語曖昧性解消手法を直接的に適用した。正例と負例がともに50用例以上利用可能な93句を対象に実験したところ、英語慣用句の曖昧性解消に取り組んだ先行研究と同等かそれ以上と考えられる性能が得られた。

**キーワード** 日本語慣用句, コーパス, 曖昧性解消, 言語資源

## Construction of Japanese Idiom Corpus and its Application to Japanese Idiom Identification

Chikara HASHIMOTO<sup>†</sup> and Daisuke KAWAHARA<sup>††</sup>

<sup>†</sup> Graduate School of Science and Engineering, Yamagata University

<sup>††</sup> National Institute of Information and Communications Technology

E-mail: <sup>†</sup>ch@yz.yamagata-u.ac.jp, <sup>††</sup>dk@nict.go.jp

**Abstract** Some phrases can be interpreted as either idiomatically (figuratively) or literally in context, and the precise identification of idioms is indispensable for full-fledged NLP. To this end, we have been constructing a Japanese idiom corpus that we hope provides a solution. This paper reports on the current status of the corpus and the result of Japanese idiom identification experiment using the corpus. The corpus targets 146 ambiguous idioms, and consists of 113,460 sentences, each of which is annotated with a literal/idiom label. The sentences have all been collected from the Web. As for Japanese idiom identification, we adopted a word sense disambiguation method, and targeted those 93 idioms for which more than 50 sentences for both literal and idiomatic usages were available. As a result, our system showed a performance that seemed equally well or better than that reported earlier on English idiom identification.

**Key words** Japanese idiom, corpus, idiom identification, language resources

### 1. はじめに

我々は、日本語慣用句の検出と格解析のための言語資源を構築している。その言語資源は、辞書、検出・格解析器、用例集から構成され、フリーウェアとして公開する予定である [1]。

慣用句の検出・格解析技術は正確な言語理解に欠かせない。例えば Excite の翻訳サイトで「その話が彼女の胸を強く打った。」を英語に翻訳させてみると、“The story strongly hit her

chest.”と誤訳される。これは、「胸を打つ」を「感動させる」という意味の慣用句として検出し、「彼女」を経験者格、「その話」を対象格として解析できなかったためである。本研究の慣用句言語資源を機械翻訳器に組み込むことにより、上記の文を “She was very impressed by the story.” と翻訳することが可能となる。

本稿では、以下、このプロジェクトの大きな柱となる慣用句曖昧性解消に焦点を当てる。

慣用句の中には文字通りの意味にも解釈可能なものがあるが、この慣用句の意味と文字通りの意味との間の曖昧性が慣用句の正確な検出の障害となっている。我々は、慣用句曖昧性解消技術の確立の第一歩として、日本語慣用句コーパスを構築している。また、そのコーパスと機械学習による標準的な曖昧性解消法を用いて、慣用句曖昧性解消実験を行った。本稿ではこの2点について報告する。

本稿の対象は慣用句の意味と文字通りの意味の間の曖昧性だけだが、1つの慣用句が複数の慣用句の意味を持つ場合もある。例えば「手を出す」という慣用句には、「殴る」や「盗む」、「女性と関係を持つ」など複数の意味がある。このような同一慣用句内の意味的曖昧性にも将来的には対処する必要がある。

また本稿では、より根元的な、「慣用句」の定義に関わる問題には触れない。本稿の「慣用句」は、慣用句文献〔2〕中の5文献〕に掲載されているものを指す。

以下、§2. で関連研究について概観した後、§3. で我々が対象とした慣用句について述べる。§4. では我々が構築したコーパスについて、§5. ではそれを用いた曖昧性解消実験について詳述する。最後に、§6. で結論と今後の展望を述べる。

## 2. 関連研究

### 2.1 慣用句コーパス

我々の知る限り、日本語慣用句コーパスの研究は〔3〕のみである。〔3〕では慣用句用例データベースの構築法を考察し、小規模な用例データベースを構築している。本研究のコーパスはより単純な構成となつてはいるが、より多くの慣用句を対象に、より多くの用例を収集した。

日本語以外では、〔4〕が TroFi Example Base という英語の慣用句や比喩表現のコーパスを構築し、公開している。<sup>〔注1〕</sup>

慣用句以外の類似研究として、〔5〕は、MUST という日本語の複合辞の用例データベースを構築し、公開している。<sup>〔注2〕</sup>

### 2.2 慣用句曖昧性解消

従来は、慣用句として使用される表現の収集（タイプの検出）に関するもの〔6〕〔7〕〔8〕などが大半を占めていた。

しかし近年、ある文脈中の句の意味が慣用句のものか文字通りのものかを判別する曖昧性解消（トークンの検出）の研究が徐々に盛り上がりを見せている〔9〕〔10〕〔4〕〔11〕など。〔9〕は、LSA を用いて、文脈中の慣用句の単語ベクトルと慣用句構成語の単語ベクトルを比較することで曖昧性解消（意味的構成性の程度の測定）を行う手法を提案している。〔10〕では、慣用句として用いられる場合と文字通りの意味で用いられる場合との間の文法的制約の差異に着目して、明示的な文法規則による曖昧性解消を試みた。〔11〕は、慣用句はごく少数の決められた形態で表出される傾向が強いという観察に基づいた unsupervised な手法を提案している。

これらの手法は、慣用句、あるいは複合語表現全般の特徴を活用したものである。一方、本研究では、過去の膨大な量の研

究に裏打された、成熟した技術といえる単語の曖昧性解消手法を、慣用句の曖昧性解消に直接的に適用した。〔4〕も単語の曖昧性解消手法を使用している。違いは、〔4〕の手法が (nearly) unsupervised であるのに対し、本稿のものは完全に supervised である点である。

## 3. 対象とする慣用句

次の手順で対象とする慣用句 146 句を選定した。

(1) 〔2〕の約 3,600 句の中からより基本的な句を抽出する。〔2〕では、小学生用辞典 2 つ、慣用句辞典 2 つ、慣用句研究文献 1 つの計 5 文献から慣用句を集めている。我々はより基本的な句として、〔2〕の 5 文献中、3 つ以上の文献に記載されているもの 926 句を抽出した。

(2) (1) の結果から、曖昧性があると認められる句を、複数人による判断結果を踏まえ、最終的に著者の一人が選定する。結局、146 句が曖昧性ありと認められた。

(2) に関して、慣用句の曖昧性の認定基準が問題となる。つまり、文脈を工夫することで文字通りに解釈できないこともない句、例えば「生き馬の目を抜く」などに対して、あるいは、もっぱら文字通りの意味で使われる句、例えば「くちばしが黄色い」などに対して曖昧性を認めるかという問題である。ほとんど片方の意味でしか用いられないものに無理に曖昧性を認めて本研究の対象としても、実用上の意義はほとんどないだろう。しかし、曖昧性の有無に関してどこで線を引くのが実用上有意義なのか、その判断が難しい。もちろん、大量の言語データから曖昧性の度合いを見積もることができればよいが、そのためには本研究で構築しようとしているリソースが必要になる。

結局、作業員 2 名（この 2 名を以後、グループ甲と呼ぶ）に曖昧性認定作業を予備的に行わせ、それらに基づいて、最終的に著者の一人が直観に基づいて曖昧性の認定を行った。

グループ甲とは異なる被験者 2 名（この 2 名を以後、グループ乙と呼ぶ）に曖昧性認定作業を行わせ、Kappa 統計量に基づいて、この 2 名の間の一致率を測定した。対象としたのは、上記 (1) で得た 926 句からサンプリングした 101 句である。これらを、「区別することが実用上有意義だと思われる曖昧性を有する慣用句」とそれ以外に分けてもらった。結果、一致率は 0.6576 となり、判断の安定性はまあまあと言える。

## 4. 慣用句コーパス

### 4.1 仕様

本研究の慣用句コーパスは慣用句の正確な検出技術、とりわけ、正確な曖昧性解消技術の開発のために構築されている。そのためコーパス中の各用例には、用例中の慣用句相当の文字列が慣用句として用いられているのか、あるいは文字通りの意味として用いられているのかを表すラベルが付与されている。以後、前者のタイプの用例を正例、後者のタイプを負例と呼ぶ。具体的には、一行が一用例の情報に対応し、各行は下記の 4 つの情報から成る。

ラベル 正例なら「i」、負例なら「f」

ID 各慣用句に与えられた 4 桁の数字。〔2〕に準拠

〔注1〕: <http://www.cs.sfu.ca/~anoop/students/jbirke/>

〔注2〕: <http://nlp.iit.tsukuba.ac.jp/must/>

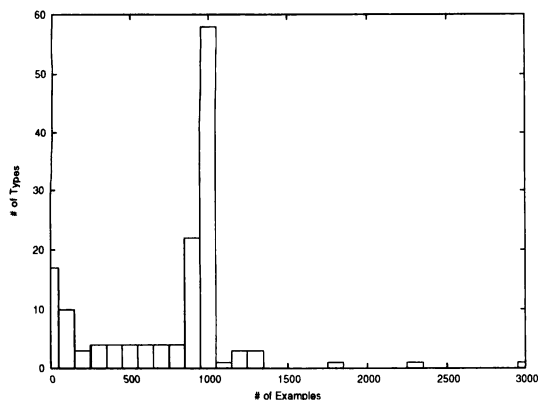


図1 慣用句コーパスの用例数の分布

Fig. 1 Distribution of the number of idiom examples

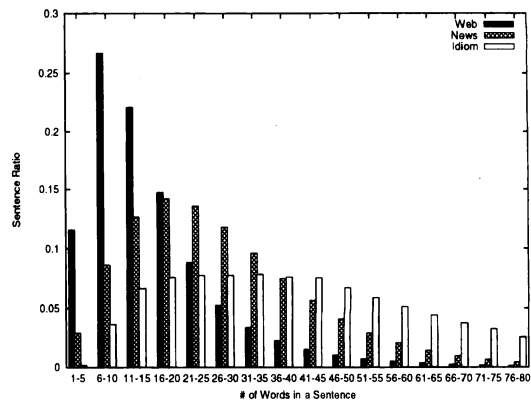


図2 文の長さの分布

Fig. 2 Distribution of sentence length

**表記** 対象とする慣用句の表記, [2] に準拠

**用例** 用例文そのもの

下記に「ごまをする」の負例の例を挙げる.

- L1417「ごまをする」すり鉢でごまをすり、ごま油を...

正例と負例の区別を一貫させるため、慣用句ごとに正例と負例の意味を明記したリストを作成した。両者の意味は、市販の複数の辞書を参考にしつつ、最終的に著者一名による判断で決められた。以後、このリストを意味参照リストと呼ぶ。例えば「骨を折る」の意味参照リストの記載は下記ようになる。

**正例** 努力する, 苦労する

**負例** 骨折する

用例数は、原則として、慣用句ごとに、正例負例併せて1,000用例になるまで収集した。1,000文に至らない場合はなるべく多く収集した。<sup>(注3)</sup> 用例は全て Web コーパス [12] から収集した。

#### 4.2 構築方法

下記の手順で慣用句コーパスを構築した。

(1) Web コーパスから慣用句相当の文字列を含む文を(正例負例関係なく)収集する。具体的には、慣用句の構成語が全て正しい係り受け関係で出現している文を自動で収集した。その際、係り受け解析器 KNP<sup>(注4)</sup>を使用した。

(2) 収集された文を手で正例と負例に分ける。この判断は上記の意味参照リストに基づいてなされる。正例負例併せて1,000文になるまでアノテーションするが、長い文から優先的にアノテーションした。また、慣用句相当文字列として不的確な誤って自動収集された用例と、前後の文脈無しには正例負例の判断がつかない用例はアノテーション対象から除外した。

この作業は、グループ甲の2名がおおよそ230時間かけて完了した(460人時)。

#### 4.3 構築結果

用例の総数は113,460となった。図1に用例数(正例と負例の合計数)の分布を示す。68タイプに対して1,000用例以上構築できたが、100用例未満の句も17タイプあることが示され

ている。

一用例の平均単語数は46語と比較的長い。図2の「Idiom」は慣用句コーパスの文の長さ(単語数)の分布を示している。「Web」と「News」はそれぞれ、[12]に掲載されている。Webコーパス全体と新聞コーパスにおける文の長さの分布である。§4.2で述べたように、慣用句コーパスは長い文から優先的に1,000用例に至るまでアノテーションして構築しているため、他のコーパスと比べて長い文が多い。

図3に「胸を打つ」と「ごまをする」の最も長い正例と負例、最も短い正例と負例をそれぞれ挙げる。「胸を打つ」の最も長い負例は自動文区切りの結果が間違っているが、手作業修正等は今のところしていない。

正例負例分類作業の一致率(Kappa統計量)を計るべく、全113,460用例から1,421用例をサンプリングして、グループ乙の2名に同様の作業を行わせた。結果、グループ乙の2名の間の一致率は0.8519と高く、判断の揺れが少ないことが分かった。

## 5. 慣用句曖昧性解消

### 5.1 曖昧性解消手法

§4.のコーパスを用いて慣用句曖昧性解消実験を行った。本研究では、次のような機械学習に基づく手法を採用した。学習法はSVM [13] (2次の多項式カーネル、TinySVM<sup>(注5)</sup>)を使用し、使用した素性は下記の通りである。

**f1:** 前後3語(機能語も含む)の品詞

**f2:** 前後3語(機能語も含む)の表出形(形態素解析器JUMAN<sup>(注6)</sup>でいう「見出し」)

**f3:** 文中の全内容語の原形

**f4a:** 係り元形態素(慣用句の先頭文節に係る文節の中で一番最後に現れる文節における、最初の形態素)の原形

**f4b:** 係り元形態素の品詞

**f5a:** 係り先形態素(慣用句の最後の文節に係る文節の中で一番最初に現れる文節における、最初の形態素)の原形

(注3): 1,001文以上集めた慣用句も10句ある。

(注4): <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/knp.html>

(注5): <http://www.chasen.org/~taku/software/TinySVM/>

(注6): <http://nlp.kuee.kyoto-u.ac.jp/nl-resource/juman.html>

i 3193 胸を打つ このアラビア語版『武士道』を読んだアラブの大使や外交官から日本人の精神が良く分かったと言う感想を貰ったが、「切腹はわが国の中世にはじまって、武士がその罪をつくらない、過ちを謝し、恥をまぬがれ、友人につぐない、そして自分の誠実を証明する方法であった」という解説に、当時のエジプト大使館の情報参事官から「切腹など全く野蛮な行為と思っていたが、腹を切ったらさぞ痛からうに、それをあえて実行するサムライの精神の高さに胸を打たれた」という感想を貰ったことも懐かしい思い出である。

l 3193 胸を打つ と言うのも、自転車で横断歩道を渡ろうとして、左折した車とぶつかってこけてしまったんです その時に自転車のハンドルで胸を打ち、すぐ救急車を呼ばれて… 歩いて3分くらいの病院なのにサイレン鳴らされました(´ー´); 幸い大事には至らず、車に乗っていた方も誠意ある対応をして下さったんですが、その後警察で事情聴取や今後どうするかなどの説明を受け大変な一日でした!

i 3193 胸を打つ よくぞここまでキャラの立った登場人物が自分勝手に行動すればするほど笑いが起こり、かと思うと一点、熱い友情が展開されて胸を打つ。

l 3193 胸を打つ 顔を天に向けようとせぜ、自分の胸を打ち叩き「こんな罪人の私を憐れんで下さい」と祈る姿をイエス様は義と認め下さったのです。

i 1417 ごまをする ただし、1562年にグレシャムは1562年に、同一の額面価値で流通する素材価値を異にする2種類の貨幣が存在すると劣悪な貨幣が流通に残り、優良な貨幣は駆逐されるという「グレシャムの法則」を発表していることから、女衛がしたたかであったように、IT興行師もメーカーにごまをすり、政府の省庁にこびを売り、知性が無くても生命力だけで、目新しい言葉のつまみ食い、悪毒、凶々しく、虚勢で生き続けることだろう。

l 1417 ごまをする 煮た大豆をつぶすには、ミンチみたいな器具があればいいのですが、ない 場合は、ごまをするもので潰すか、もっと簡単な方法としては、ビニール 袋に大豆を入れ、封をしてタオルをかけてその上から瓶でこするようにする といいでしょ。

i 1417 ごまをする 上にごまをする小役人タイプ。

l 1417 ごまをする ごまをすり調味料とあえる

図3 「胸を打つ」と「ごまをする」の最も長い正例と負例と最も短い正例と負例  
 Fig.3 The longest and shortest examples (each for literal and idiomatic usages) of *mune-otutu* (chest-ACC hit) 'impress' and *goma-o suru* (sesame-ACC mince) 'smarm'

f5b: 係り先形態素の品詞

f6: 文中の全内容語の JUMAN カテゴリ

f7: 文中の全内容語の JUMAN ドメイン [14]

これらの素性の抽出処理には JUMAN と KNP を用いた。

この枠組みは、概ね、[15] に基づく。[15] はいくつかの学習法と素性を単語の曖昧性解消に適用し、その結果、SVM (WEKA<sup>(注7)</sup>) を使用。パラメータ等は WEKA のデフォルトのまま) と上記の f1 から f5 に概ね相当する素性の組み合わせが良好な結果を示したことを報告している。

f4 と f5 は、[15] の Syntactic Relations 素性を慣用句用に設計し直し、かつ、より単純にしたものである。例えば次の「胸を打つ」の用例の場合<sup>(注8)</sup>、f4 は「聴衆」の原形と品詞、f5 は「歌声」の原形と品詞になる。

- (聴衆/の)(胸/を)(打つ)(美しい)(歌声)

f6 と f7 は JUMAN が出力する情報の一つで、前者が単語の上位概念に、後者が単語の属する分野あるいはトピックに概ね対応する。例えば「聴衆」の場合、JUMAN カテゴリは「人」、JUMAN ドメインは「文化・芸術あるいはメディア」である。「歌声」なら、JUMAN カテゴリは「抽象物」、JUMAN ドメインは「文化・芸術あるいはレクリエーション」である。

## 5.2 実験条件

実験では、正例と負例がともに 50 用例以上利用可能な慣用句 93 タイプ<sup>(注9)</sup>を対象として、慣用句タイプごとに SVM モデ

ルの構築と評価実験を行った。

評価指標は [11] に倣い Accuracy を採用した。

$$Accuracy = \frac{\text{正しく曖昧性解消できた用例数}}{\text{全用例数}}$$

ベースラインは、慣用句コーパスにおいてより用例数が多い方に一律に解釈した場合とした。

$$Baseline Accuracy = \frac{\max(\text{正例数}, \text{負例数})}{\text{全用例数}}$$

これは慣用句ごとに決定される。例えば、慣用句コーパスにおいて、正例数が 600 で負例数が 400 なら、その慣用句の Baseline Accuracy は 60% ととなる。

[11] のベースラインも、本研究と同様、より用例数が多い方に一律に解釈した場合という方針だが、実際は全て慣用句と解釈した場合になっている。

タイプごとの Accuracy と Baseline Accuracy は、その用例集の学習用/評価用への分割の仕方を 10 通りに変化させて得た 10 の値の平均として求めた (10 分割交差検定のスタイル)。

実験対象の慣用句全体での Accuracy と Baseline Accuracy も求めた。これは、慣用句タイプごとの Accuracy を合計し、慣用句のタイプ数 (93) で割ったもの (macro-averaged) である。

また、Relative Error Reduction (RER) を次の式に基づいて算出した。

$$RER = \frac{\text{ベースラインのエラー率} - \text{システムのエラー率}}{\text{ベースラインのエラー率}}$$

慣用句全体の RER は全体の Accuracy と Baseline Accuracy から上の式で直接算出した。

(注7) : <http://www.cs.waikato.ac.nz/ml/weka/>

(注8) : 「(...)」が文節を表し、「/」が形態素区切りを表す。

(注9) : 素性抽出処理の失敗で利用不可能になった用例もある。そのため実験に使用された用例数は、慣用句コーパスの実際の用例数より若干少ない。

表1 タイプごとの実験結果: Accuracy (%Acc), Baseline Accuracy (%Base), Relative Error Reduction (%RER), 用例数  
 Table 1 Individual results: Accuracy (%Acc), Baseline Accuracy (%Base), Relative Error Reduction (%RER), and # of examples

タイプ	%Accu	%Base	%RER	正例	負例	タイプ	%Accu	%Base	%RER	正例	負例
青筋を立てる	86.89	83.48	20.66	293	58	手を回す	94.07	68.78	81.00	310	683
あぐらをかく	91.99	61.76	79.04	617	382	峠を越す	89.45	72.31	61.89	692	265
足が付く	81.37	72.43	32.42	188	494	泥を塗る	92.99	79.27	66.20	803	210
足が出る	91.56	79.27	59.30	170	650	波に乗る	93.20	86.60	49.28	892	138
足元を見る	85.99	63.14	61.99	627	366	熱が冷める	92.03	89.70	22.65	923	106
足を洗う	93.38	68.04	79.28	677	318	熱を上げる	93.99	92.30	21.97	935	78
足を伸ばす	94.60	78.90	74.39	774	207	熱を入れる	91.39	85.32	41.34	773	133
頭が痛い	83.88	61.00	58.67	172	269	根を下ろす	93.48	85.43	55.23	844	144
頭を抱える	91.52	86.32	37.98	852	135	根を張る	87.83	60.20	69.43	593	392
頭をもたげる	93.57	82.54	63.20	832	176	バスに乗り遅れる	92.18	76.45	66.78	207	672
脂が乗る	93.36	83.43	59.91	202	1017	バトンを渡す	81.92	65.21	48.03	476	254
油を売る	88.55	86.70	13.91	515	79	鼻息が荒い	76.00	52.71	49.26	292	262
油を絞る	85.89	67.14	57.07	70	143	鼻が高い	82.51	50.67	64.55	663	681
網を張る	80.93	70.17	36.07	377	887	鼻を折る	74.08	58.68	37.26	69	98
息が詰まる	79.05	70.56	28.85	707	295	鼻を鳴らす	81.08	55.46	57.52	554	445
一から十まで	92.55	91.90	8.13	771	68	腹を割る	96.53	95.64	20.47	1293	59
色を失う	84.06	72.88	41.24	272	731	歯を食い縛る	69.07	64.82	12.08	199	108
腕が上がる	84.75	54.19	66.70	511	432	人を食う	88.24	75.15	52.66	747	247
尾を引く	93.80	87.70	49.59	877	123	火花を散らす	89.72	75.81	57.51	752	240
顔を出す	88.10	85.60	17.36	856	144	筆を入れる	82.69	76.53	26.26	238	73
肩を並べる	92.93	88.70	37.48	879	112	船をこぐ	78.17	50.75	55.69	170	165
角が取れる	79.14	57.36	51.07	374	278	骨が折れる	93.55	61.23	83.36	608	385
唇をかむ	78.27	71.09	24.81	605	246	骨を埋める	90.05	82.83	42.07	767	159
口を切る	85.28	51.64	69.56	221	236	骨を折る	92.67	61.95	80.73	379	617
口をとがらせる	87.91	86.22	12.22	676	108	幕が開く	86.10	54.67	69.35	539	447
首が回らない	86.57	66.23	60.23	651	332	右から左	90.14	73.85	62.31	796	2248
首を切る	90.74	52.37	80.56	520	473	水と油	83.40	55.66	62.57	1053	839
首をひねる	93.50	92.11	17.56	934	80	水に流す	87.49	67.54	61.46	674	324
事によると	97.39	66.96	92.10	231	114	身に付ける	95.30	91.30	45.98	913	87
ごまをする	86.67	50.27	73.19	91	92	耳が痛い	89.60	61.45	73.02	384	612
背を向ける	85.00	68.80	51.92	688	312	耳に入れる	89.79	73.10	62.03	723	266
血が通う	82.38	50.74	64.23	431	444	実を結ぶ	96.16	89.05	64.98	878	108
力を入れる	96.81	91.41	62.83	915	86	胸が痛む	95.66	93.34	34.83	925	66
宙に浮く	88.48	59.80	71.36	398	592	胸が膨らむ	94.42	55.45	87.46	343	427
土が付く	82.69	72.99	35.90	74	200	胸を打つ	95.92	91.03	54.56	893	88
手が届く	87.27	80.96	33.14	812	191	目が覚める	94.20	93.89	5.02	61	938
手がない	93.39	86.28	51.82	874	139	芽が出る	92.43	55.23	83.09	432	533
手が離れる	90.96	54.46	80.15	454	543	目がない	96.07	91.89	51.53	906	80
手に乗る	93.05	59.96	82.63	410	614	メスを入れる	96.39	85.64	74.83	853	143
手を入れる	93.01	52.65	85.23	527	474	目に入る	89.65	83.07	38.87	834	170
手を打つ	96.17	94.96	23.93	942	50	目を覆う	90.57	85.97	32.75	858	140
手を掛ける	92.04	70.45	73.08	294	701	目を覚ます	87.62	83.52	24.87	165	836
手を切る	90.73	56.08	78.91	563	441	目をつぶる	90.55	69.47	69.05	685	301
手を取る	93.03	88.89	37.23	112	896	目を細くする	77.76	55.56	49.97	116	145
手を握る	95.65	91.59	48.22	85	926	指をくわえる	95.88	92.45	45.42	919	75
手を延ばす	92.78	89.36	32.12	108	907	弓を引く	95.92	88.46	64.63	138	1058
手を広げる	89.50	69.98	65.04	697	299						

### 5.3 実験結果

表1に、対象とした93タイプ各々の実験結果を示す。

対象とした慣用句全体の結果は表2の通りである。

これらの結果から、全体のベースラインが73.60%と比較的

表 2 全体の実験結果

%Accu	%Base	%RER
89.19	73.60	59.07

表 3 [11] の実験結果

Table 3 Results reported in [11]

	%Accu	%RER
Baseline [11]	61.9	-
Unsupervised [11]	72.4	27.6
Supervised [9]	76.2	37.5

高いこと、にも関わらずシステムの Accuracy が 89.19% とさらに高いことが分かった。表 1 の個々の結果を見ても、ベースラインを下回るものはない。「腹を割る」は慣用句として使われることが多く、ベースラインは 95.64% とかなり高いが、システムの Accuracy は 96.53% とそれを上回っている。

[11] では、[11] におけるベースライン（全て正例と判断した場合）と、[11] の提案した Unsupervised な手法、そして、[9] の提案した Supervised な手法の macro-averaged accuracy を比較している（表 3）。[11] の対象は英語の慣用句であり、実験条件も異なるので、単純には比較できないが、本研究のシステムは基本的に単語の曖昧性解消手法の流用でありながら、その性能は [11] や [9] のそれと同等かそれ以上であると考えられる。

## 6. おわりに

本稿では、我々が構築している日本語慣用句コーパスについて述べた。また、それをを用いた慣用句曖昧性解消実験の結果について報告した。

コーパスに関して、まだ用例数が足りないものがある。それらについては、表記を変更して Web コーパスから収集しなおす予定である。例えば、現在「身に付ける」という表記で用例を収集している慣用句に対して、表記を「身につける」に変更して収集しなおす。

このようにまだ完全に仕上がったとは言えないが、これだけの規模で日本語慣用句を扱ったコーパスは、我々の知る限り、これが最初である。

コーパスは近日公開予定である。

慣用句曖昧性解消に関して、本稿の実験では完全に supervised な枠組みのもと行った。しかし、我々は扱う慣用句数を徐々に増やす予定であり、scalability の確保のため、今後、unsupervised で高精度な手法を開発しなくてはならない。[4] の unsupervised な手法は WordNet に依存しており、日本語慣用句の曖昧性解消には向かない。一方、[11] は特定の言語資源に依存しない手法であり、今後我々のタスクへの適用を計画している。今後、unsupervised あるいは semi-supervised な手法の開発と評価においても、本研究のコーパスが大きな役割を果たすだろう。

### 謝辞

慣用句リスト [2] を提供して下さった名古屋大学大学院 佐藤

藤理史先生に感謝申し上げます。

また、数多くの有益なコメントをして下さった、京都大学情報学研究所-NTT コミュニケーション科学基礎研究所共同研究ユニットのメンバーの方々と情報通信研究機構のフランシス・ボンドさんに感謝申し上げます。

本研究は、日本学術振興会科学研究費補助金若手研究 (B) 「日本語慣用句の検出と格解析のための言語資源の構築」(課題番号 19700141, 代表者 橋本力) の援助を得てなされました。

## 文 献

- [1] 橋本, 河原: “慣用句の検出と格解析のための言語資源の構築”, 言語処理学会第 14 回年次大会, pp. 1148-1151 (2008).
- [2] 佐藤 (編): “基本慣用句五種対照表”, 名古屋大学大学院工学研究科 佐藤理史研究室 (2007).
- [3] 尾嶋, 佐藤, 宇津呂: “日本語慣用句用例データベースの構築法”, 言語処理学会第 12 回年次大会, pp. 456-459 (2006).
- [4] J. Birke and A. Sarkar: “A clustering approach for the nearly unsupervised recognition of nonliteral language”, Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2006), pp. 329-336 (2006).
- [5] 土屋, 宇津呂, 松吉, 佐藤, 中川: “日本語複合辞用例データベースの作成と分析”, 情報処理学会論文誌, 6, pp. 1728-1741 (2006).
- [6] 首藤, 吉村, 武内, 津田: “日本語の慣用表現について—語の非標準的用法からのアプローチ—”, 情報処理学会研究会報告, 1988-NL-66-1, pp. 1-7 (1988).
- [7] T. Baldwin, C. Bannard, T. Tanaka and D. Widdows: “An empirical model of multiword expression decomposability”, Proceedings of the ACL 2003 workshop on Multiword expressions, pp. 89-96 (2003).
- [8] A. Fazly and S. Stevenson: “Automatically constructing a lexicon of verb phrase idiomatic combinations”, Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006), pp. 337-344 (2006).
- [9] G. Katz and E. Giesbrecht: “Automatic identification of non-compositional multi-word expressions using latent semantic analysis”, Proceedings of the Workshop, COLING/ACL 2006, Multiword Expressions: Identifying and Exploiting Underlying Properties, pp. 12-19 (2006).
- [10] C. Hashimoto, S. Sato and T. Utsuro: “Detecting Japanese idioms with a linguistically rich dictionary”, Language Resources and Evaluation, 40, 3-4, pp. 243-252 (2006).
- [11] P. Cook, A. Fazly and S. Stevenson: “Pulling their weight: Exploiting syntactic forms for the automatic identification of idiomatic expressions in context”, Proceedings of the ACL 2007 Workshop on A Broader Perspective on Multiword Expressions, pp. 41-48 (2007).
- [12] 河原, 黒橋: “高性能計算環境を用いた web からの大規模格フレーム構築”, 情報処理学会 自然言語処理研究会 171-12, pp. 67-73 (2006).
- [13] V. Vapnik: “The Nature of Statistical Learning Theory”, Springer-Verlag, New York (1995).
- [14] C. Hashimoto and S. Kurohashi: “Construction of Domain Dictionary for Fundamental Vocabulary”, ACL '07 Poster, pp. 137-140 (2007).
- [15] Y. K. Lee and H. T. Ng: “An empirical evaluation of knowledge sources and learning algorithms for word sense disambiguation”, EMNLP '02: Proceedings of the ACL-02 conference on Empirical methods in natural language processing, pp. 41-48 (2002).