複合機能表現「という」の分類にみる MCN コーパスの方法論検証

叢悠悠（お茶の水女子大学理学部）
田中リベカ（お茶の水女子大学理学部）
中村絢子（お茶の水女子大学理学部）
酒向美帆（お茶の水女子大学理学部）
佐宗智子（お茶の水女子大学理学部）
清水蘭（お茶の水女子大学理学部）
劉月晴（お茶の水女子大学理学部）
川添愛（国立情報学研究所）
戸次大介（お茶の水女子大学院人間文化創成科学研究科／国立情報学研究所）

Methodology of the MCN Corpus in the Classification of a Functional Compound “toiu”

Yuyu So (Faculty of Science, Ochanomizu University)
Ribeka Tanaka (Faculty of Science, Ochanomizu University)
Ayako Nakamura (Faculty of Science, Ochanomizu University)
Miho Sako (Faculty of Science, Ochanomizu University)
Tomoko Saso (Faculty of Science, Ochanomizu University)
Ran Shimizu (Faculty of Science, Ochanomizu University)
Yuechin Ryu (Faculty of Science, Ochanomizu University)
Ai Kawazoe (National Institute of Informatics)
Daisuke Bekki (Graduate School of Humanities and Sciences, Ochanomizu University / National Institute of Informatics)

１．はじめに

自然言語で記述されるテキストには、書き手にとって真であることが確実な情報と、そうでない情報が混在する。例えば、「太郎が結婚した」という命題について、以下のような文が考えられる。

1. 太郎が結婚した。
2. 花子は太郎が結婚したと言う。
3. 噂によると、太郎が結婚したという。
4. 仮に、太郎が結婚したとする。

このうち、1.は事実だが、2. 3. 4.は事実として捉えてはならない。なぜなら、書き手にとって命題の真偽がはっきりとしておらず、書き手が命題に対して何らかの心的態度を持っているためである。このように、ある命題に対する書き手の認識や態度を表す言語表現をモダリティという。モダリティ表現には、「…という」「…かもしれない」といった様相表現、「…でない」のような否定表現、「～なら…」という条件表現などがある。人間はモダリティ表現から情報の確実性を判断しているのである。

また、上に挙げた例文は、Web 上で「太郎　結婚」というキーワード検索を行ったとき

にヒットする可能性のあるものとしてみることもできる。膨大なテキスト情報の中から確実性の高い情報を選り抜くためには、事実である 1.と、そうでない 2. 3. 4.を区別したいものである。すなわち、モダリティ表現に着目することが必要になる。

本研究では、MCN コーパス（川添ら（2011））のアノテーションガイドラインで使用している言語学的テストの改良を行っている。MCN コーパスは、モダリティ表現に意味アノテーションを付与した言語データである。具体的には、各表現の用法ごとの分類が示されたガイドラインを用いて、テキスト中の表現にラベル付けしたものである。言語学的テストとは、理論言語学の知識に基づいて作成されたテストで、文または文の一部の容認性や適切性を判定するものである。MCN コーパスのアノテーションにおいては、言語学的テストとして「ネガティブテスト」（田中ら（2012a, 2012b））を採用しており、各表現に対するネガティブテストを用意したガイドラインを作成している。本論文では、様相表現「（と）いう」「とする」に対する最新のガイドラインについて、その問題点を考察する。

本論文では以下、第 2 節で MCN コーパスのガイドラインで用いているネガティブテストの概要を述べる。第 3 節では、実際のアノテーション作業で、アノテータ間の意見が分かれやすかった表現について論じる。

２．MCN コーパスのガイドラインにおける言語学的テスト

MCN コーパスのアノテーションでは、アノテータの判断の不一致を避けるために、ネガティブテストを導入している。ネガティブテストは、「文中の表現を別の表現に置き換えたときに文として成立しない、あるいは意味が変化する場合、その用法としてアノテーション不可能（つまりそのカテゴリに分類されない）」という形式をとる。ここで、「置き換え不可能であればアノテーション対象ではない」としているのは、「置き換えが可能」という判断よりも、「置き換えは不可能」という判断の方が、アノテータ間での一致度が高いという傾向が見られるためである（田中ら（2012a, 2012b））。ネガティブテストで置き換え不可と判定された場合、その分類に属さないことが断定できるため、これを用いたアノテーション作業では、消去法で分類先を一つに特定することになる。一つの表現に対する分類先が一意に決定されることは、一貫性のあるコーパスを構築するにあたって重要である。消去法を行った結果として複数のカテゴリが残った場合は、それらのうち、本来の分類先でないカテゴリのテストが不適切であることを意味する。

MCN コーパスのアノテーションで使用しているガイドラインは、「言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4」（川添ら（2011））をもとにしている。これは、言語情報の確実性に関わる表現にアノテーションを付与し、機械による確実性判断の基盤となるコーパスを構築するために作成されたものである。もともとのガイドラインには、各言語表現について用法別のカテゴリが例文や統語環境などとともに示されている。しかし、これらの基準だけでは、ある表現がどのカテゴリに属するかを判断できない場合がある。例えば、以下の文中の「という」に対し、例文ベースのガイドラインを用いて、他人の認識を表す「（と）いう」としてアノテーション可能かを考える。

1. 太郎が責任をとるべきという人はどうかしている。
2. 太郎が結婚したという話だ。

ガイドラインの記述：

他人の認識 【(と) いう】
分類 : 他人の認識 (他人の報告する事柄や、命題の真偽に関する他人の判断を表す表現)
例 : 今年のインフルエンザの流行は全国的に遅れているという。

1.および 2.の「という」は、ともに名詞句を修飾しており、例文と異なる形をとっているようにみえるが、ガイドライン設計者は、1.は他人の認識としてアノテーション可能であり、2.は不可能であると意図している。両者の違いは、専門的な知識を有しない一般のアノテータが容易に見出せるものではない。

そこで、田中ら (2012a, 2012b) は、「(と) いう」を「との」と「(と) 述べる」のそれぞれに置き換えるテストを作成し、どちらに置き換え可能かで別々のカテゴリに分類するようにした。上の文にこの二つのテストをそれぞれ適用すると、以下のようになる。

1a. 太郎が責任をとるべきとの人はどうかしている。
1b. 太郎が責任をとるべきと述べる人はどうかしている。
2a. 太郎が結婚したとの話だ。
2b. 太郎が結婚したと述べる話だ。

1.の「という」は「と述べる」の置き換えは可能だが、「との」に置き換えると不自然な文になる。一方、2.は「との」に置き替え可能だが、「と述べる」に置き換えることはできない。このように、テストを用いると判断がしやすく、アノテータ間の一致度も高くなる場合が多い。

無論、複数のアノテータによるアノテーション結果が完全に一致するようなテストを作成することは難しい。例えば、「今回の募金は 10 万円を目標とする」という文に対し、「とする」を「にする」に置き換えたとき、容認できるか、不自然に感じるかは人それぞれである。テスト適用時に生じる変化は、個々人の言語感覚に問うものであり、容認の可否が分かれるのは避けられない。また、「太郎は花子を許さないという」という文に、テストとして「わざわざ」を挿入したとき、明らかにニュアンス上の変化が起こるが、それが意味に影響するか否かの判断はアノテータに委ねられる。このような置き換えや挿入による語感の変化がどれだけ大きいと「置き換え不可」となるかを明確に定義するのは、事実上不可能である。

それでもテストを採用しているのは、実際にテストを用いてアノテーションを行った際に、例文ベースのガイドラインを使用するのに比べてアノテータの判断が容易になり、一致度も向上する傾向があるからだ。また、テストを使用しない場合は、例文との類似性のみから判断するほかないが、そうして得られたアノテーション結果に確たる根拠は見出せない。テストを用いたアノテーションは、より信頼性のあるコーパスを得るのに必要な手法であると考えられる。

３．現在のガイドラインの問題

表 1 は、「(と) いう」「とする」のアノテーションに対する最新のガイドラインから一部を抜粋したものである。ガイドラインの完成度は、テストをもとにしてアノテーションを行った際に、分類先が一つに特定できたか、また、アノテータの判断がどれだけ一致しているかによって測られるが、現在のガイドラインは改良の途上にあり、問題点が多くある。本節では以下、アノテーション結果の不一致を招く要因となるもののうち、アノテータの

判断に対する影響が顕著であった 4 つの問題を取り上げる。

表 1：最新のガイドライン（一部抜粋）

sem	表現	別表記	特徴	例文	テスト	統語環境	備考
3	いう		人あるいは物と、その名前を関連づける。「と」の前には原則として固有名詞であり、「インスリン」のような専門用語の一般名詞が現れることもある。 語用論的には、「という」の前の固有名詞あるいは専門用語の指示する対象が、話し手か聞き手の少なくとも一方にとって馴染みのないものであることを表す。	その人は山田という。 そのホルモンはインスリンという。 初めまして。私、山田といいます。 その人は名前を山田という。 そのホルモンは名前をインスリンという。 初めまして。私、名前を山田といいます。 日本には、富士山という山があります。 富士山という山は、どこにあるのですか？ 長崎の名物に、トルコライスというものがあります。	「呼ばれ（てい）る」に置き換えて意味が変化する場合はこのカテゴリではない。 （※主語が 1 人称の場合は違和感がある。） 「～という」の前に「名前を/名を～という」のように「名前を/名を」を補って意味が変化する場合はこのカテゴリではない。	[動作主(NP)]が[名前(NP)]という [動作主(NP)]が名前を/名を[名前(NP)]という	「という5」との区別がつきにくいが、「という」の前が固有名詞でも専門用語でもない場合は「という3」ではないと考えてよい。
2	とする		仮想的な状況を記述する。「想定する」「仮定する」に近い意味をもつ。	太郎が犯人だったとする。その場合、アリバイはどう説明するんだ？ 無人島に、一つだけ物を持っていけるとしよう。君は何を持っていく？ 運転中に視界が悪くなったとします。その場合はどうすればよいでしょうか。 来年三月までの収入の合計を 300 万円とする。その場合、税金はいくらになるか。 直線 AB 上の点を Q とする。	「とする」を「想定する」「仮定する」のいずれにも置き換え不可、あるいは置き換えて意味が変化する場合はこのカテゴリではない。	[S]とする [NP]を[NP]とする	

3.1 「いう 1」「いう 2」の区別について

本ガイドライン中の「いう 1」は、「言葉を発するという意図的な動作を表す」とある。また、動作主を特定の人物以外に「世間一般、人々、みんな」と設定し、「多くの場合は明示される」としている。一方、「いう 2」は、「伝聞の意味を持つ」とある。動作主に関しては、「明示されない」としている。また「～によると」とともに使われる場合が多いとされている。

しかし、動作主が「（と）いう」の直前にない場合、「いう 1」か「いう 2」かの決定が困難なことがあった。以下は、家庭訪問の実態を題材にした新聞記事からの引用である。

1. 家庭訪問は、明治初期に不就学児を登校させるよう親を説得する目的で始まった。「師範学校付属校のような中核校から周辺に広がっていったのでは」と佐藤教授はみている。また、家庭と学校の不干渉が徹底している欧米では、家庭訪問は基本的にないという。

表 2：「いう 1」および「いう 2」

表現	特徴	例文	テスト
いう1	言葉を発するという「意図的な動作」が意味の中心である。 また「NP が」という形の項として動作主を要求する（多くの場合節内に動作主が明示される）。 ※動作主が「世の人」「人々」「みんな」「誰か」である場合、「という2」と意味的に近くなるが、これは「という1」である。	太郎は「昨日渋谷で花子を見た」と言う。 花子はまだ怒っているようで、太郎を絶対に許さないという。 「太郎が責任をとるべき」と言う人は、どうかしている。 「叶わない夢はない」と人はいう。 花子は、太郎を天才だと言う。 その時警官が通りかかったことは、幸運だったというしかない。 花子が「おいしい」という店には行かない方がいいよ。	「話す」「主張する」「述べる」「表現する」「評価する」「判断する」のいずれにも置き換え不可、あるいは置き換えて意味が変化する場合はこのカテゴリではない。 「わざわざ」「口に出して」「あえて」「しつこく」のいずれを挿入しても意味が不自然になる場合はこのカテゴリではない。
いう2	伝聞の意味をもつ。「言葉を発する動作」よりも、むしろ「言説の存在」あるいは「言説が流布している状態」を表しているもの。 動作主が明示されない。 語用論的には、話者が間接的な言語情報として得たことを表す（直接経験して知っていることについては使わない）。 情報源を表す「～によると」と共起することが多い。情報源が明示されない場合、「世間一般」「人々」「専門機関あるいは公的機関の公式発表」である。	ニュースによると、インフルエンザが流行っているという。 警察の調べでは、男は以前から現場付近で目撃されていたという。 駅前の焼肉屋は、このあたりで一番おいしいという。 日本人の9割が何らかのストレスを抱えているという。 世界には自分と同じ顔の人間が7人はいるという。 私たちの普段の生活の中にも、空海が中国からもたらしたというものがあります。それは一体、何でしょう。 私たちの普段の生活の中にも、空海が中国からもたらしたというものがあります。それは一体、何でしょう。	「そう（だ）」と置き換えて違和感がある場合はこのカテゴリではない。 「いわれる」「いわれている」に置き換え不可、あるいは置き換えて意味が変化する（尊敬の意味になる等）場合はこのカテゴリではない。

例文 1.の「という」は、動作主が明示されておらず、「いう 1」と「いう 2」両方のテストが適用可能であるため、どちらか一方に分類することは難しい。ここで注目すべき点は、「欧米では、家庭訪問は基本的にない」という言葉を佐藤教授が実際に発したのか、あるいは話者が他の情報源から伝聞したものなのかというところである。この前後の文を参考にすれば、どちらか特定できる可能性もあるが、実際のアノテーション時におけるアノテータの負担を考慮すると、一つの命題を判断するために広範囲の文章を参照するのはなるべく避けたいものである。

次に、1.の最後の一文を次のように換えてみる。

2. 家庭訪問は基本的に「ない」<u>という</u>。

こちらは、かぎ括弧をつけたことによって、一見佐藤教授が発した言葉のように感じられる。しかし、このかぎ括弧は話者が強調のためにつけたとも考えられ、その場合は佐藤教授から直接聞いた言葉でない可能性がある。かぎ括弧がせりふを表すものであるか、強調のためにつけられたものであるかは、文章全体を読んでもなかなかわかるものではない。新聞等では、既出の人物の発した言葉が、主語を伴わず、かぎ括弧に括られた形で出現することが多々あるが、その場合にこのような問題に直面してしまう。

また、ガイドラインにおいては、「いう 1」の動作主として「世間一般、人々、みんな」等が挙げられている。しかし、これらが動作主となっている場合は、主語を省略する傾向がある。特に、新聞等においてはそれが顕著で、「(と) いう」の前にある命題が「動作主を明示していないが、世間一般で言われていること」であるケースが少なくない。

3. よく「朝食を摂る子供は成績が良い」<u>という</u>。それは本当なのだろうか?

かぎ括弧の中は、世間一般でよく言われているという点においては、不特定多数の人物が意図的に発している言葉である。しかし、この文には動作主が明示されておらず、話者が伝聞したことのようにもとれるため、「いう 2」としてもアノテーションできてしまう。実際、「いう 1」と「いう 2」のテストを適用すると、「いう 1」のテスト「よく～と話す」よりも、「いう 2」のテスト「よく～といわれる」の方が自然である。

これらの問題を根本的に解決する方策として、省略されている動作主を補うことができるか (補ったことによって文が不自然にならないか)、といったテストを追加することが考えられるが、文脈に応じて適切な動作主を補うのは、多くのアノテータにとって容易でないことが推測される。

３.２　「(と) いう」「とする」の命題 / 名詞句の判断

「(と) いう」と「とする」のどちらにも共通して、直前が命題か名詞句かの判断が必要となるカテゴリがある。例えば、「3 は奇数である」「インフルエンザが流行している」は命題であり、「私の赤いドレス」「野球をすること」は名詞句である。

ここで、三平方の定理「直角三角形の斜辺の二乗は他の二辺の二乗の和に等しい」を考える。「三平方の定理」は紛れもなく名詞句である。一方、直角三角形の斜辺を c、他の二辺をそれぞれ a, b とおくと、この定理は「$a^2 + b^2$ は c^2 に等しい」と表せるが、これは命題である。しかし、同じ等式を意味する「$a^2 + b^2 = c^2$」が命題であるか、名詞句であるかの判断は困難である。そのため、「$a^2 + b^2 = c^2$ という式」の「という」に対して、「いう 5」「いう 7」の二つのカテゴリが候補となってしまう。

別の例として、「自分を含めて客が 4 人というライヴに行ったことがある」という文を考える。「という」の前の部分は、一見名詞句のように見えるが、これは「自分を含めて客が 4 人である」から「である」が省略された形となっており、命題であるとされる。

このように、数式の形になっているものや、語尾の「である」が省略されているものは、統語環境を判定できず、分類の決定時に混乱を招く恐れがある。現在のガイドラインにおいては、命題と名詞句に対する定義が不十分であるため、今後より幅広い表現に対応できるよう改善する必要がある。

表 3：「いう 5」および「いう 7」

表現	特徴	例文	テスト
いう5	「という」の前の NP として「今日」「お前」のような直示的表現や、「東京」のような固有名詞、「コーヒー」のような一般名詞が現れることが可能。（前の NP の意味が後ろの NP の意味に含まれていることを表す） 前方の NP の意味が後方の NP の意味に含まれていることが常識的に明らかな場合…強調の効果 後方の NP が「妻」や「生きがい」などのロール概念を表す場合…前方の NP のどの側面に着目するかを限定する効果 前後の NP 間に意味的な包含関係があることが明らかでない場合…それらの間に包含関係があるとする（話者の）主張を強調する効果	今日という日を忘れないようにしよう。 お前という人間がわからなくなった。 長年住んでいるが、東京という町には親しみがわいてこない。 コーヒーという飲み物は実に奥が深いね。 犬や猫の目には、人間という動物がどのように映るのだろう。 ボランティアという生きがいに出会ってから、毎日が楽しくなりました。 相手の女性も、私という妻の存在を知らないわけがありません。 私は、家族という重荷を背負って生きていくのには向いていない。 経済成長という病（*本の名前） アメリカという記憶（*本の名前） 人間というものは、よほどのことがない限り考えを変えようとしない。 それが男というものだと割り切るしかない。 まったく、限度というものを知らないんだから。	（補助テスト*） 「NP という NP」の形で、 (1)前方 NP が固有名詞/専門用語でない場合…という5 (2)固有名詞/専門用語の場合… ①前方 NP の指示対象が話し手・聞き手の少なくとも一方にとって馴染みのないものである→という3 ②双方にとってなじみのあるものである場合→という5
いう7	特に意味的な内容はなく、関係節的な特徴を持つ[命題(S)]と係り先の[NP]との間のつなぎとしての役割のみを持つ。	子供が高校生や大学生という世帯は、全世帯の中でも特に出費が目立つ。 僕もデビューする前は、一年間収入が全くないという時期を過ごしたこともあります。 この人となら結婚してもうまくいくだろう、という人がなかなか現れないのよね。	「（名詞 or 状詞）＋という NP」の場合：「（名詞 or 状詞）＋（の or な）NP」に置き換え不可の場合はこのカテゴリではない。 「（動詞 or 形容詞）連体形＋という NP」の場合：「（動詞 or 形容詞）連体形＋NP」に置き換え不可の場合はこのカテゴリではない。

3.3 「いう 6」の抽象名詞リストについて

本ガイドラインでは、「[命題] という [名詞句]」という形のものを、名詞句が抽象名詞であるかどうかで、それぞれ「いう 6」「いう 7」に分類している。文中の名詞が抽象名詞であるかを判断する際には、表 4 の抽象名詞リストを参照している。しかし、実際のアノテーション作業において、文中の名詞がリストにないが、リスト中のほかの抽象名詞と近い意味を持つ場合、判断が困難であった。例えば、「教室にエアコンを設置してほしいという要望があった」の「要望」はリストに挙げられていないが、それに類似した表現「要求」「声」は含まれている。このように、あらかじめリストの形で提示できる抽象名詞の数は限られてしまい、現実世界の表現すべてに対応するのは原理的に不可能である。

表 4：抽象名詞リスト

事象	事実、真実、事態、事件、こと、出来事、事情、状況、状態、症状、人事、例、事例、判例、現象、問題
ことば	言葉、格言、名言、せりふ、文、文句、文言、遺言、言い方
情報媒体	情報、ニュース、話、記事、報告、知らせ、便り、メール、電話、口コミ、噂、報道、記録、声、音、声明、手紙
言語行為（発話内行為）	命令、忠告、約束、説明、発言、発表、指示、主張、提案、提言、要求、決定、指摘、質問、答え
思考行為	意図、理解、認識、反省、考え、意見、見解、結論、仮定、前提
具体的行為	行動、行為、作業、仕事、習慣
感情	感情、気持ち、意識、感じ、不安、希望、不満、欲望、恐れ、寂しさ、幸せ、喜び、悩み、心配、懸念、疑問、空しさ、自信
概念	概念、思想、主義、知識、理屈、理由、目的、説、学説、理論、法則、印象
モダリティ	可能性、見込み、危険
表出	態度、そぶり、ふり、表情
内容	内容、あらすじ、（話の）流れ、シナリオ
手順	作戦、手順、順序、順番、手続き、計画、企て、予定、プロジェクト
性質	性質、特性、側面、一面、点、利点、長所、短所、特徴、観点

「いう 6」のテスト：
「という」の後の NP が命題（あるいは命題の集合）を意味する抽象名詞（句）でない場合はこのカテゴリではない。
（※ただし、「というもの」の場合は、「もの」と同一指示関係を持つ名詞の種類が抽象名詞（句）であるかを考える）

「いう 7」のテスト：
「（名詞 or 状詞）＋という NP」の場合：
「（名詞 or 状詞）＋（の or な）NP」に置き換え不可の場合はこのカテゴリではない。
「（動詞 or 形容詞）連体形＋という NP」の場合：
「（動詞 or 形容詞）連体形＋NP」に置き換え不可の場合はこのカテゴリではない。
（※「ということ」の場合は、「こと」が抽象名詞句のいずれかに解釈可能な場合が多い）

３.４ 「として 3」の慣用表現について

3.3 項と類似した問題が「として 3」でも生じる。このカテゴリは、「結果として」「時として」などといった慣用表現に特化したもので、例文の項目に代表的な表現がいくつか提示されており、例文に含まれていない表現については、慣用表現かどうかの判断がアノテータに委ねられている。しかし、慣用表現かどうかについての認識には個人差がある。例えば、「感じとして」という表現について、「慣用表現である」という意見と、「単なる言い回しである」という考えに分かれる傾向があった。そもそも、慣用表現と言い回しをどう区別するのかも自明ではない。

表 5 :「として 1」および「として 3」

表現	特徴	例文	テスト
として 1	「～という位置づけで」とほぼ同義。 「～という立場で」、「～という役割で」、「～という名目で」などと言い換えると自然な場合がある。	山田氏を課長として採用する予定。 大人として恥ずかしくないのか。 賞金として二十万円が贈られます。 働きがいのある会社として注目されている企業。	「[NP を][NP として]」という形で出現している場合:「[NP として][NP を]」の順序に入れ替えることができない、入れ替えると意味が変化する場合はこのカテゴリではない。 [NP として]を省略することができない、省略すると意味が変化する場合はこのカテゴリではない。 [NP として]単独で出現した場合:「～という位置づけで」「～という立場で」「～という役割(役職)で」「～という名目で」のいずれにも置き換え不可、あるいは置き換えて意味が変化する場合はこのカテゴリではない。
として 3	慣用的な表現	原則として、部外者の立ち入りを禁ずる。 結果として、その年の合格者はたったの五人だった。 人生には時として、何をしてもうまくいかないことがある。 誰一人として理解してくれない。 一日として忘れたことがない。 遅々として進まない。 彼の行方は、杳として知れない。	

一見、単なる言い回しとの区別がしにくい慣用表現であるが、単独のカテゴリを設けるのは妥当なのだろうか。以下の文を考える。

1. それは原則としてしっかり押さえておく必要がある。
2. 原則として部外者の立ち入りを禁ずる。

1.の「として」は、位置づけの働きを持つので、「として 1」に分類される。一方、2.の「原則として」は慣用表現である。こちらは副詞的な役割であり、文中から除いたときに、程度上の変化はあっても、重要な情報の欠落は生じない。このことから、慣用表現の「として」は、ほかのカテゴリと明白に異なる機能を持つため、特別なカテゴリを設ける必要があると筆者らは考えている。

文中の表現が慣用表現であるか否かの判断は、一般のアノテータには困難である。そのため筆者らは、慣用表現を列挙したリストをガイドラインで提示し、「リストに含まれない表現は慣用表現ではない」というテストを今後追加する予定である。このようなリストの作成にあたって、3.3 項で述べたような問題が危惧されるが、慣用表現は抽象名詞に比べると数が限られている。よって、リストにはないが慣用表現であると判断した場合はテスト設計者にフィードバックし、その過程を通してリストの内容が収束していくことが期待で

きると考えている。

4. おわりに

本稿では、MCN コーパスにおけるアノテーションの問題点を考察した。今後、より一貫性のあるアノテーション結果が得られるよう、テストの改良を行う方針である。

文献

田中リベカ、小池恵里子、戸次大介、川添愛（2012a）「言語学テストに基づく意味アノテーションのガイドライン設計―確実性判断に関わる表現を中心に」言語処理学会第 18 回年次大会発表論文集, pp.401-404.

田中リベカ、川添愛、戸次大介（2012b）「MCN コーパス：言語学的テストに基づくモダリティ・アノテーションの理論と実証」国立国語研究所第 2 回コーパス日本語学ワークショップ予稿集, pp135-144.

川添愛、齊藤学、片岡喜代子、崔栄殊、戸次大介（2011）「言語情報の確実性に影響する表現およびそのスコープのためのアノテーションガイドライン Ver.2.4」Technical Report of Department of Information Science, Ochanomizu University, OCHA-IS 10-4.