

テキストマイニングのための意味アノテーション

小泉敦子 森本康嗣 相菌敏子

(株)日立製作所 中央研究所

〒185-8601 東京都国分寺市東恋ヶ窪 1-280

{koizumi, y-morimo, aizono}@crl.hitachi.co.jp

営業日報、コールセンター応答履歴などのテキスト情報活用のためのアノテーションについて検討している。本報告では、営業日報の分析において「実際に起きたこと」と「起きなかったことや予定していること」の区別が重要であることに着目し、事象の生起に関するアノテーションを提案する。営業日報を対象とする予備評価を行った結果、生起／未生起のいずれのケースもある程度の頻度で生じていることからアノテーションの有効性を確認した。また、動詞や動作名詞のごく近傍の情報を手がかりとするルールだけで 30～50%の精度が得られたことから、実用レベルでの自動付与が可能との感触を得た。自動学習も含めて自動付与方式の改良を検討することと、他分野への適用性を確認することが課題である。

1. はじめに

インターネットの普及にともない、世界規模での情報共有が実現される一方で、情報が氾濫し、情報の収集・統合が困難になるという状況にある。このような中で、次世代 Web 技術として、W3C(World Wide Web Consortium) の Tim Berners-Lee が提唱する Semantic Web が注目されている。現在の Web 空間は、主に人間が理解するためのコンテンツの空間である。これに対して Semantic Web では、コンテンツにメタデータを与えることにより、機械処理可能な知識空間を構築することを目指している。テキストに与えるメタデータとしては、作者や作成日などのいわゆる書誌情動的なメタデータの他に、記述内容の意味を扱うためのメタデータが必要である。このためにはテキストにアノテーション(註釈)を与える必要があるが、「テキストの意味を扱うにはどのようなアノテーションをすべきか」は明らかにされていない。

テキストの意味を扱う技術のひとつに、大量のテキスト情報から有益な知識を抽出する手法・プロセスであるテキストマイニングがある[1]。テキストマイニングに関しては、データマイニングの領域から、また自然言語処理や情報検索に領域から、さまざまなアプローチでの研究が行われており、すでにいくつかの実用システムが出現し

ている。しかし、データマイニングに比べて使い方や効果の典型例がはっきりしない状況にあり、テキストマイニング技術として何が求められているかということもあまり明確にされていない。テキストマイニングにおいても「テキストの意味を扱うにはどのようなアノテーションをすべきか」ということが課題である。

本研究では、「テキストの意味を扱うにはどのようなアノテーションをすべきか」という問題に対して、テキストマイニングにおける具体的な課題を取り上げ、それを解決するためのアノテーションを検討するというアプローチを取っている。本報告では、営業日報の分析においては、「実際に起きたこと」と「起きなかったことや予定していること」の区別が重要であることに着目し、事象の生起に関するアノテーションを提案する。

2. テキストの意味を扱うためのアノテーション

テキストを機械処理しやすくするために言語学的アノテーションを付与する研究は、もともとコーパス言語学の分野でおこなわれて来た。また、近年では、セマンティック Web やテキストマイニングの研究において、テキストの意味を扱うためのアノテーションが提案されている。

(1) 言語的アノテーション

言語的アノテーションについては、テキスト一般を対象とした TEI (Text Encoding Initiative) (1987-)[2] のガイドライン、コーパスを対象とした EAGLES (The Expert Advisory Group on Language Engineering Standards)(1993-1999)のガイドライン、EAGLESガイドラインの一部として提案されている CES (Corpus Encoding Standard) [3], GDA(Global Document Annotation)[4]などが知られており、アノテーションのレベルやタグセットに関する標準化の提案がされている。いずれも、目的を特化しない汎用的なアノテーションを目指している。また、アノテーションの自動化は必ずしも重視していない。オーサリングツールを使って人手や半自動でアノテーションをしても良いという立場である。意味アノテーションには、大きくわけて(a) 意味関係(格関係など)、(b) 単語の属性(語義など)という2つのタイプがある[5]。

(2) セマンティックWebのためのアノテーション

W3Cが提唱するセマンティックWebでは、オントロジとの対応付けを特徴とするアノテーションが提案されている。意味アノテーションに関しては、SHOE[6]、Annotea[7]などのプロジェクトがある。概念間の関係を記述したオントロジと対応付けることで、推論などの知的な処理を行えるようにするのが狙いである。メタデータ記述言語、オントロジ記述言語など、アノテーション記述の枠組みに関する標準化は進められているが、タグセットは使う人が目的に応じて設計するという立場である。

(3) 目的指向のアノテーション

テキストマイニングにおいては、単語、文節間の係り受け関係、意図(質問・要望など)などの情報をテキストから抽出し、頻度やその偏りからテキストの特徴を分析する[8][9]。アノテーションは自動化することが前提である。したがって、前述の言語的アノテーションのように詳細なアノテーションを付与するのではなく、自動化できる範囲で、目的を達成する上で必要なアノテーションのみを付与する。

上記(1)～(3)のいずれの立場も、テキストの意味を扱うためのアノテーションをめざしている。我々は、「テキ

ストの意味を扱うにはどのようなアノテーションをすべきか？」という共通の問題に対して、テキストマイニングにおける具体的な課題を取り上げ、それを解決するためのアノテーションを検討するというアプローチを取る。

3. 営業日報のマイニング

テキストマイニングの適用分野のひとつに、コールセンター応答履歴や営業日報などを対象としたCRM向けテキスト分析がある。本章では営業日報の分析を取り上げ、マイニングという観点で、何と何を区別しなければならないのかを考察する。

3.1. 営業日報の記述内容

営業日報の記述内容は、案件ID、活動日、訪問先、面会者などの定型データ部分とテキスト自由記述部分とに大きく分けることができる。

テキスト自由記述による報告内容からは、以下に関する情報を得ることができる。

- 営業プロセス： 提案活動(セミナー・展示会への案内、デモ実施など)、引き合い対応(ヒアリング、見積りなど)、受注/失注とその要因(価格、機能、保守体制・・・)など
- 製品： 自社製品の販売状況や評判、自社/他社の関連製品(前提製品など)、競合製品との比較
- 顧客： 社内状況、現在使用中の製品の状況、顧客ニーズ、顧客が現在取引のある他社
- 競合： 競合他社名、競合他社の営業活動・強み・弱み、受注社名(失注した場合)
- 協力関係： 販社・他事所等との協力関係

上記の情報は、社内の各部門において以下の目的で活用することができる。

- 営業部門： 営業活動、営業戦略立案
- 企画・設計部門： 商品開発・改善、リスク管理
- 経営部門： 経営課題発見と施策立案、施策実施結果の検証

3.2. マイニングの手法

3.2.1. マイニングの目的

営業日報のような細切れで大量のテキストに対して我々がしたいのは以下のことである。

定量分析

- テキストを読まずに定量的な傾向を知りたい
- テキストを分類し、偏りや時系列変化、相関関係を知りたい(定型データと同様の定量分析をしたい)

定性分析

- テキストを読んでキラリと光る情報を得たい
- 全部は読めないなので、まずは何が書いてあるか、トピックの偏りや関係を示す要約情報を得たい
- 読むべきものだけを効率良く見つけて読みたい
- 読むべきものは分類・整理してから読みたい(同じものを何度も読みたくないし、似ているものはまとめて読みたい)

3.2.2. マイニングの手順

そこで、我々が提案するテキスト分析手法では、以下の手順によって分析を行う。

(1) 状況把握: 業務の観点から設定した分析軸でテキストを分類し、かたよりや増減を定量的に分析することによって全体傾向を把握する。定量分析の方法としては、時系列分析や相関分析がある。

(2) 課題設定: 定量分析結果に基づいて優先的に注目すべき分類カテゴリを決める。分類カテゴリ内のテキストを定性的に分析し、解決すべき課題を設定する。さらに課題を細分化し、それにあわせて分類軸も細分化する。定性分析の方法としては、分類カテゴリ内の文書から特徴的な語句を抽出し、(a)概観する、(b)関連語を見る、(c)テキストを読む、といった方法がある。

(3) 仮説発見: 1つの分類項目の中で、あるいは複数の分類項目にわたってテキストの内容を定性的に分析し、問題の真の原因、解決策についての仮説を立てる。

(4) 仮説検証: 解決策の妥当性を裏付ける事例を収集する。

3.3. マイニングためのアノテーション

このために、テキストに対して、分析軸に基づくアノテーションを付与する。営業日報分析では、例えば、製品、顧客、自社、競合、協力者という業務的観点からアノテーションを付与する。

(1) モノに関するアノテーション

- 顧客: 企業名
- 製品: 製品名, カテゴリ
- 自社: 担当者, 部署
- 競合: 競合会社名, 競合製品名
- 協力関係: 販社・他部署等

(2) 事象に関するアノテーション

係り受け関係(命題)とモダリティの情報により、テキスト中に含まれる事象を表現する。これらを上記のモノに関するアノテーションと組み合わせることで、営業活動、顧客状況、製品の評判などを分析することができる。

(a) 係り受け関係

テキスト内の係り受け関係を「命題」として抽出する。例えば、以下のような述語に関する係り受け関係が得られる。

営業プロセス: 「提案」「引き合い」「ヒアリング」「デモ」「見積」「提示」「値引き」「内示」「受注」など

顧客: 「導入」「決定」「使用」「要望」「発注」など

製品: 「変更」「保守」「構成」「展開」「連携」「サポート」「稼動」「使いやすい」「高価だ」など

競合: 「提案」「引き合い」「デモ」「見積」「提示」「値引き」など

協力関係: 「依頼」「報告」「担当」「手配」「相談」など

(b) モダリティ

文の表す意味は、命題とモダリティ(話者の心的態度)という2つの側面から成り立っている[10]。モダリティには、「～したい」、「～してください」、「～しなくてはなら

ない」、「～しますか？」などがある。テキストマイニングにおいて、モダリティの区別は重要である。例えば、製品に対する顧客の評価として、「安い」と「安くして欲しい」は区別する必要がある。

事象に関するアノテーションの記述内容の例を以下に示す。

(例)「新製品を提案して欲しい」
→(新製品－提案する, 要望)

4. 事象の生起に関するアノテーション

4.1. アノテーションの目的と概要

テキストマイニングにおいては、実際に起きた事柄だけを数えたいという要求がある。例えば、「製品別のデモ件数を調べたい」というとき、「製品A－デモ」という係り受け関係の出現件数を数えると、「製品Aデモ実施の予定」のような、実際にデモをしていないケースがノイズとして入ってしまい、正確さに欠ける。同様に、実際に起きた事柄だけを検索して読みたいという要求がある。製品Aのデモをした報告を探して読みたいのに、読んでみたら「製品Aデモ実施の予定」では期待はずれである。営業日報のマイニングでは、「生起／未生起」の区別、すなわち、「した」と「するつもり」の区別が重要である。さらに、「事象の生起について書き手が知らない」というケースも含めて「生起／未生起／不明(書き手が知らない)」の区別をすることにより、以下のような要求に応えることが可能となる。

- 「ハードディスクを静かにして欲しい」も「ハードディスクが静かでない」もともに、「ハードディスクが静かであること」が「未生起」として扱いたい。
- 「プリンターが故障した。」と「プリンターが故障した場合は、どこに問い合わせればいいのか？」を区別したい。
- 顧客先で発生した障害について検索したり数えたいが、「障害が発生した場合の・・・」は除外したい。
- 顧客がどんな製品／業務プロセスの導入を検討しているかを知りたいが、「導入の検討を提案」「導入を検討しているか？」は除外したい。

- 自社製品や競合製品に対する顧客の評価を知りたいが、「使いやすいか？」などは除外したい。
- 顧客がどんな製品を使っているかを知りたいが、「〇〇は使用できない」などは除外したい。

そこで、製品、顧客、自社、競合、協力者という観点で、何についてどんなことが起きたかを検索したり数えたりすることができるようにすることを目的に、事象の生起に関するアノテーションを提案する。

営業日報の記述内容を時間の観点から見ると、以下のことが言える。

- 営業日報は、過去のある時点に書かれたものである。
- 報告内容は、執筆時を基点とした過去／現在／未来の事象に関するものである。書き手は、事象の生起について「生起したと思っている」「生起しないと思っている」「知らない」のいずれかである。
- 「執筆時を基点とした過去の事象」の中には、顧客や関係者との会話も含まれる。会話の内容は、会話時を基点とした過去／現在／未来の事象に関するものである。

したがって、営業日報に記述された事象の生起に関して我々が知ることができるのは、「過去のある時点における命題の真偽(に関する書き手の判断)」と言うことができる。そこで、事象の生起については、「生起／未生起／不明」の3通りで記述する。

事象の生起／未生起に関するアノテーションの記述内容の例を以下に示す。

- 「新製品を提案した」→生起
- 「新製品を提案中」→生起
- 「新製品を提案予定」→生起
- 「新製品を提案して欲しい」→未生起
- 「新製品の提案を希望。」→未生起
- 「新製品の提案を実施。」→生起
- 「新製品を提案したら興味を示した」→生起
- 「新製品を提案したら興味を示すかもしれない」→未生起

4.2. アノテーション自動付与方式

本章では、事象の生起に関するアノテーションの自動付与方式について述べる。

4.2.1. アノテーションの手がかりとなる情報

否定・テンス・アスペクトは、事象の生起のありようを表すものである。モダリティは、事象の生起に関する話者の何らかの認識に基づく心的態度である。したがって、これらの表現を手がかりに事象の生起を判断することができる。以下にその例を示す。

(1) 否定

➤ 否定の「ない」は、事象が生起していないことを表す。(例:「製品Aを提案しなかった。」)

(2) テンス

➤ 過去の「た」は、事象が生起したことを表す。(例:「製品Aを提案した。」)

➤ 現在形は、事象が「生起している」という認識を表す場合と、「将来において生起する」という意志や予想を表す場合がある。(例:「製品Aは製品Bと競合する。」「今週中にデモを実施する。」)

(3) アスペクト

➤ 完了／結果状態の「た」「いる」「ある」は、事象が「生起した」という認識を表す。(例:「見積は提出しである。」)

(4) モダリティ

➤ モダリティには、行為の実行に関するモダリティ(意志、勧誘、依頼など)、叙述のモダリティ(評価・認識・説明)、疑問のモダリティ(質問・確認要求)がある[10][11]。

➤ 行為の実行に関するモダリティは、話し手や聞き手が行為を実行することを促したり制限したりするものであるから、「事象がまだ生起していない」ことが前提である。(例:「～を提案したい」)

➤ 評価のモダリティは、事象が「まだ生起していない」ことが前提である。評価のモダリティ表現には、「～すべきだ」(必要)、「～してもいい」(許可)、「～しなくていい」(不要)、「～してはならない」(不許可)、

「～したらいい」(願望)などがある。(例:「～を提案すべきだ」)

- 認識のモダリティは、「事象が生起した／していない」という言明に対して、断定・推量・可能性・必然性・推定・伝聞などを表明する。
- 事象の生起を表す表現(「た」「ている」など)の後に推量・推定・伝聞のモダリティ表現が付くと、「事象が生起した」との言明を支持していると解釈される。(例:「～を提案しただろう」「～を提案したようだ」「～を提案したそうだ」)
- ただし、推量については、「もし～していれば～しただろう」のように、事象が未生起であったことを前提に、「条件が違えば事象が生起しただろう」という仮定法の表現もある。
- 事象の生起を表す表現(「た」「ている」など)の後に可能性・必然性のモダリティ表現が付くと、「事象が生起した」との言明を支持するものの不確かだと解釈される。(例:「～を提案したかもしれない」「～を提案したはずだ」)
- 説明のモダリティは、「事象が生起した／していない」という言明に対して、それが説明であることを表す。
- 事象の生起を表す表現(「た」「ている」など)の後に説明のモダリティ表現が付くと、「事象が生起した」との言明を支持していると解釈される。(例:「提案したのだ」)
- 疑問のモダリティは、「事象が生起した／していない」ということについての話者の疑問を表明する。
- 事象の生起を表す表現(「た」「ている」など)の後に質問のモダリティ表現が付くと、事象が「生起したかどうか、話者は知らない」と解釈される。(例:「提案しましたか?」)
- 事象の生起を表す表現(「た」「ている」など)の後に確認要求のモダリティ表現が付くと、「事象が生起したと話者は信じている」と解釈される。(例:「提案しましたよね?」)

以上述べたように、否定・テンス・アスペクト・モダリティ表現に基づいて事象の生起に関するアノテーションをすることができる。

4.2.2. ルールの記述

以下の方針でルールを記述することにした。

(1) 内容語による否定・テンス・アスペクト・モダリティ表現の収集

否定・テンス・アスペクト・モダリティ表現としては、前述のような助動詞等の機能語を使った表現が良く知られている。しかし、営業日報等の報告書では、動詞や名詞などの内容語を使った表現が多く使われる。以下にその例を示す。

(例) 「製品 A の提案あり。」「製品Aを提案予定。」「製品Aの提案を希望。」「見積りは提出済み」「見積りの提出を依頼」「発注の可能性あり」「製品Aを提案することを希望。」

「～したい」のような機能語を使った表現は、日本語表現一般に共通する比較的閉じたセットと考えられる。これに対して、上記のような内容語を使った否定・テンス・アスペクト・モダリティ表現は、多くのバリエーションがあると考えられる。また、コミュニティごとに特徴的な文体があると考えられる。これらについては、営業日報データから表現を収集し、ルールを記述することにした。

(2) 手がかりを探る範囲

複文、名詞句表現など、手がかりが遠くにあるケースがある。

(例) 「FS納入に向け、要件の詳細検討及び基本設計作業を実施中。」
「製品Aの提案については、担当の〇〇が検討中。」

動詞や動作名詞の近傍の情報だけを使ったルールでもある程度の効果が得られるのではないかと考え、今回のルール記述では対象とする述語の文節とその後ろ2文節の範囲でルールを記述することにした。したがって、上記のようなものは正しく扱うことができない。

(3) 曖昧性のある表現

「～を提出する。」「～を提出。」のような現在形や動作名詞で文が終わっているものは、「生起／未生起」に関

して曖昧性がある(人間が読めば、動詞の特徴や文脈から判断できるケースが多い)。営業日報では図 5.1 に示すような箇条書きが多く、このような表現が多いが、先行する「今後の予定」「経過の報告」などの表現から判断できることもある。今回のルール記述では、狭い範囲の手がかりを使ったルールのみを適用することにしたので、曖昧性のある表現については「判断できないもの」として扱うことにする。

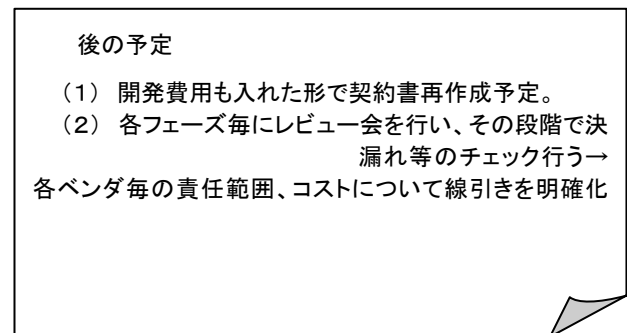


図 5.1 箇条書きの例

5. 予備評価

5.1. 評価の目的

本章での評価の目的は、以下の2点である。

(1) 定量的な有効性の確認

たとえ、生起／未生起を区別することに意味があるとしても、実際の分析において有効であるとは限らない。例えば、全ての語がほとんどの場合、生起と考えられるような場合には、生起／未生起の区別は定量的にはほとんど影響を与えない。また、ほとんどの語が同一の割合で生起／未生起と判断されるような場合には、相対的な頻度には変化がないため、有効性は大きく減じられる。そこで、実データを参照し、以下の点を確認する。

- 生起／未生起のいずれのケースもある程度の頻度で生じていること。
- 語によって生起／未生起の割合が異なること。

(2) 自動化のフィージビリティの確認

前章までに述べた観点に基づいて簡単なルールを作成し、どの程度の精度・カバレッジで自動抽出が可能であるかを確認する。評価には、弊社の営業日報データを

用いた。文数は 403,492 文である。

表 5.1 生起／未生起の割合

	生起	未生起	不明
導入	17	66	24
デモ	40	31	25
発注	9	53	45
発生	50	53	6
納入	35	54	21

5.2. 評価手順

5.2.1. ルールの作成

コーパスを分析用コーパスとテスト用コーパスに分割し、分析用コーパスから動詞、動作名詞を抽出して、その近傍に出現する単語を抽出した。分析用コーパスとテスト用コーパスのサイズはそれぞれ以下の通りである。

- 分析用コーパス(文):123,282
- テスト用コーパス(文):280,210

次に、分析用コーパスから抽出された近傍の単語を頻度順に集計し、人手によって精査して、ルールを作成した。

5.2.2. 評価対象語の選定

営業日報において重要であると考えられる以下の動詞を評価対象語とした。

- 導入(顧客の活動を示す)
- デモ(営業プロセスでの活動を示す)
- 発注(営業プロセスでの活動を示す)
- 発生(製品での事象を示す)
- 納入(営業プロセスでの活動を示す)

5.2.3. ルールによる自動付与と人手による評価

テスト用コーパス中に出現した評価対象語について、作成したルールによって、生起／未生起の自動判別を行い、結果を人手で評価する。これにより、生起／未生起の割合と、自動付与精度を確認する。なお、目視による評価を行うため、自動付与結果から、各評価対象語毎に 100 個をランダムに抽出し、評価対象とした。

5.3. 評価の結果

5.3.1. 生起／未生起の割合

人手で評価した生起／未生起の割合を表 5.1 に示す。表から分かるように、生起／未生起のそれぞれがある程度の頻度で生じており、また、語によって頻度の割合が大きく異なることから、生起／未生起を区別することは有用であると考えられる。

5.3.2. 自動付与精度

自動付与精度の評価結果を表 5.2 に示す。

表 5.2 自動付与精度

	生起を求める			未生起を求める		
	再現率	適合率	F 値	再現率	適合率	F 値
導入	47.1	66.7	55.2	25.8	100.0	41.0
デモ	45.0	78.3	57.1	9.7	100.0	17.7
発注	55.6	31.3	40.1	18.9	83.3	30.8
発生	54.0	64.3	58.7	11.3	100.0	20.3
納入	37.1	76.5	50.0	35.2	100.0	52.1
平均	47.8	63.4	52.2	20.2	96.7	32.4

F 値で 30～50%程度の精度で抽出が可能となっている。ごく近傍の情報を手がかりとするルールだけでもある程度の精度を得ることができたことから、実用レベルでの自動付与の見通しを得た。今回の予備評価の精度では、件数をなるべく正確に数えるという目的には不十分だが、適合する事例を早く見つけて読むという目的では効果がある。

5.3.3. ルールの記述に関する考察

今回の予備評価では、動詞・動作名詞を含む文節とその後ろ 2 文節という狭い範囲でルールを記述した。うまく扱えなかった事例から、以下のことがわかった。

(1) 表現のパリエーション

営業日報では、助動詞などの機能語ではなく動詞などの内容語を使った時制・アスペクト・モダリティの表現が多い。パリエーションがどの程度網羅できるかについて

て見通しを立てる必要がある。

(例)「発注に到っていない」

(2) 手がかりが遠くにあるもの

複文や名詞句表現など、手がかりが遠くにあるケースが予想以上に多かった。係り先を見る正攻法では、係り受け解析の精度に依存する。また、この方法では、「納入が2月末に前倒しになる予定」のようなケースの扱いが難しい。深い解析をせずに表層的なパターンで扱う方法を考える必要がある。

(例)

「FS納入に向け、要件の詳細検討及び基本設計作業を実施中。」

「障害が発生し、急遽〇〇に連絡し、電話で指示を仰ぎつつ作業を続行。」

「必要なものについては納入し年内にサービスインする。」

「PDCサーバを導入し、既設PDCをファイルサーバに置きかえる予定。」

「本番機の納入が2月末に前倒しになる予定。」

6. おわりに

本研究では、「営業日報分析において何と何を区別したいか」という観点からテキストマイニングの課題を考察し、事象の「生起／未生起／不明」の区別が重要であることを明らかにした。これをもとに、事象に関するアノテーションとして、従来から提案されている係り受け関係、モダリティ[10][11]に加えて、事象の生起に関するアノテーションを付与することを提案した。モダリティ表現を手がかりに、よりシンプルで本質的な情報である「話者が事実をどう認識しているか」を抽出する点に本研究の特徴がある。

有効性および自動付与の可能性に関する予備評価を行い、(a) 生起／未生起のいずれのケースもある程度の頻度で生じていること、(b) 語によって生起／未生起の割合が異なること、の2点を確認した。今回の評価では、ごく近傍の情報を手がかりとするルールだけでもある程度の精度は得られていることから、実用レベルでの自動付与が可能であるとの感触を得た。

自動学習の検討も含めて、アノテーション自動付与方式の改良が課題である。また、事象の生起に関する情報の使用目的(欲しい情報をピンポイントで得たいのか、件数をなるべく正確に数えたいのかなど)をより具体化し、再現率と適合率の優先付けや、「不明(事象が生起したかどうかを話者が知らない)」の扱いについて検討する必要がある。モダリティ表現の少ない文書(たとえばマニュアルなど)も含めて、異なるタイプの文書にも適用実験を行い、適用性を確認する必要がある。

参考文献

- [1] Sullivan, Dan: Document Warehousing and Text Mining; John Wiley & Sons, Inc. 2001
- [2] <http://www.tei-c.org/>
- [3] <http://www.cs.vassar.edu/CES/>
- [4] <http://i-content.org/gda/>
- [5] Aguado de Cea, Guadalupe: RDF(S)/XML Linguistic Annotation of Semantic Web Pages; COLING2002 Workshop. 2002
- [6] Heflin, J., Hendler, J., and Luke, S. SHOE: A Knowledge Representation Language for Internet Applications. Technical Report CS-TR-4078 (UMIACS TR-99-71), Dept. of Computer Science, University of Maryland at College Park. 1999.
- [7] <http://www.w3.org/2001/Annotea/>
- [8] 那須川哲哉: コールセンターにおけるテキストマイニング; 人工知能学会誌 Vol.16-2(2001)
- [9] 齊藤孝広: 文書記述内容のメタデータ化について; 情報処理学会研究会報告, 2003-F1-70(2003)
- [10] 日本語記述文法研究会編(代表仁田義雄): 現代日本語文法4 第8部 モダリティ; くろしお出版(2003)
- [11] 宮崎和人ほか: 新日本語文法選書4 モダリティ; くろしお出版(2002)