

問題解決知識を求める自然言語メッセージの分析

林 良彦

大阪大学大学院・言語文化研究科・言語情報科学講座

hayashi@lang.osaka-u.ac.jp

1 はじめに

インターネット上の掲示板やNewsgroupなどにおいては、問題解決に関する知識や情報(以下、知識情報と記す)の交換が活発に行われている。このような知識情報のアーカイブを情報源として適切に利用できれば、従来のキーワード検索による情報検索とは一線を画する、より高度な知識情報へのアクセス手段を実現できる可能性がある。アクセス手段の実現へ向けては、知識情報の交換の場においてやり取りされるメッセージの特徴や、そこで記述される問題状況や問題解決行為の類型を把握しておく必要があると考えられる。

本研究では、そのような取り組みの第一歩として、Web上の掲示板サイトに投稿された問題解決に関する知識情報を求める問い合わせのメッセージを対象とし、メッセージにおける問題状況記述の言語表現上の特徴を分析した。さらに、この分析結果を手がかりとし、メッセージの背景にある実際の問題状況の類型化を試みた。

2 知識情報へのアクセス手段

従来からの汎用的なキーワード情報検索を補完する情報アクセス手段として、製品・サービスなどの評判情報に特化した検索や、いわゆる質問応答(QA)技術をWeb検索上に実現した事実知識の検索などの検討・試行が盛んに行われている。

一方で、問題解決を指向した知識情報(ここでは、問題解決に有用な手段・道具などの情報やノウハウ・ティップス的な知識などを指す)に対するアクセス手段の実現も望まれており、ネット上のコミュニティにおける知識や経験を活用したサービス¹が一定の支持を得るようになってきている。

従来よりNewsgroupなどの場を用いてネットワーク上で草の根的に交換されてきた知識情報は、信頼性などの情報の質の問題はあるが、コミュニティメンバーのノウハウや経験に即した知見や知恵が多く含まれていることから、メーカーの公式サイトなどから得られる情報とは異なる利点がある。このため、コミュニティによるメッセージのアーカイブを情報源として利用することにより、自身が直面する問題の解決を指向するユーザの情報行動(information behavior)[3]をより具体的に支援できる可能性がある。すでに、Newsgroupをターゲットとする情報検索機能[4]が検討されているが、交換される知識情報の構造

化を行うことにより、FAQの生成支援や問題解決の状況に即したより高度な情報検索手段を実現できる可能性がある。

3 知識情報を求める問い合わせメッセージ

本研究では、数あるWeb上の知識情報交換のサイトの中から、パソコンのOS(Windows)に関して活発なやりとりが行われているサイトであるWindows.FAQ²を分析の対象として選択した。具体的には、同サイトにおいてWindows 2000に関する過去ログのコーナーに整理されているページ群からランダムに4つのhtmlページを抽出し、そこに収録されているメッセージを初期分析の対象とした。

当サイトにおける過去ログは、いわゆるスレッドごとにメッセージが整理されている。すなわち、スレッドを立ち上げたメッセージに回答のメッセージ群が続く形となっている。スレッドにおける応答メッセージを含めた検討は今後の課題とし、今回は、スレッドを立ち上げたメッセージ(初期メッセージと呼ぶ)のみを分析の対象とした。上記の4つのページにおいては、1873件のメッセージが含まれているが、上記の定義によりスレッドの起点として初期メッセージとして分類されるものは、332件であった³。

3.1 問い合わせメッセージの抽出

332件の初期メッセージの多くは、ゴール達成や問題解決の過程における問題状況(より正確には、それに対面するメッセージの書き手自身の心的状態)によりトリガされる情報探索行動の発現である問い合わせのメッセージであることが期待できる。実際、332件中の309件(93.1%)がこのような問い合わせメッセージであった。これらの309件の問い合わせメッセージにおいては、明示的な情報要求の表現を含むものが多いが、メッセージの書き手(以下、「書き手」または「ユーザ」と記す)のゴールや直面している問題状況のみが記述されているメッセージも少ない割合(11件; 3.6%)ではあるが存在する。メッセージが交換される場がWindowsに関係する問題解決知識の交換の場であることがコミュニティにより共有されているため、上記のような情報要求を陽に含まないメッセージも十分に知識情報を要求する機能を果たす。

² <http://homepage2.nifty.com/winfaq/>

³ これより、1スレッドあたりの平均メッセージ数は、5.6件となるが、Newsgroupを対象とした検索に関する最近の研究[4]において報告されている値と比べて、約2倍の件数となっている。

¹ OK Web (<http://okweb.jp/>) など

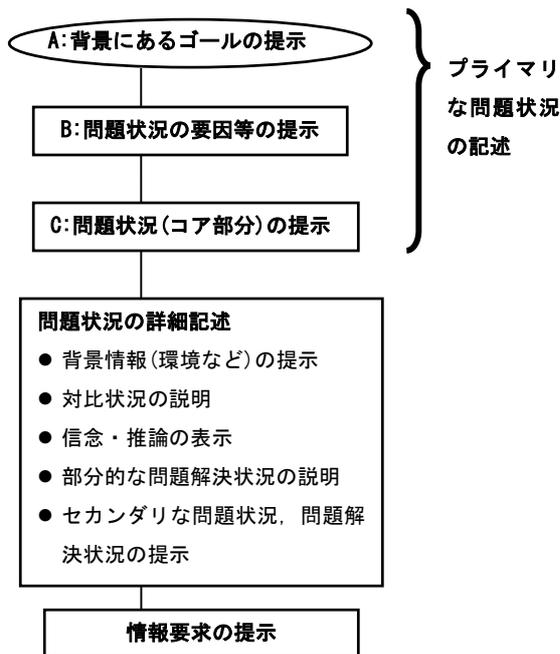


図1: メッセージの大局的構造

3.2 問い合わせメッセージの大局的構造

一般に、問い合わせメッセージに現れる問題状況の記述は、ユーザが直面している問題状況の複雑さ・深刻さやユーザの書き手としての特性に応じて、任意の長さ・複雑さとなりうる。今回の分析では、問い合わせメッセージの大局的な構成要素として、おおむね図1に示すようなものを想定する。

図2に問い合わせメッセージの典型例の一つを示す。破線下線部は『ゴールの提示(A要素)』、実線下線部は『要因等の提示(B要素)』、二重線下線部は『問題状況(コア部分)の提示(C要素)』を示す。図1におけるいくつかの構成要素は、実際には記述されない場合がある。また、各要素が必ずしも図1の順に具体化されるわけではない。実際この例においても、『ゴールの提示』より前に『背景情報(環境)の提示』が行われている。また、プライマリな問題状況に対する『詳細記述』はなされていない。

『情報要求の提示』に関しては、図2の例における波線下線部のように具体的な情報を求める形になっているものが多いが、「何かヒントをお願いします!」といった実態的な意味を持たないものも多く見られる。これらの情報要求の最大の機能は、問題状況に関連した何か有用な情報を求めていることを強調することであると考えられる。このため本分析では、問い合わせメッセージにおけるもっとも重要な要素として『プライマリな問題状況の説明』の部分に注目する。

4 問題状況を構成する要素の表現上の特徴

本節では、図1における『プライマリな問題状況の記述』を構

表題とおり自宅にある2000proから遠隔地にある2000サーバーにインターネット経由VPNで接続し、共有資源をメンテナンスしています。
この度、新規ユーザーを登録したいと思い、ActiveDirectoryユーザーとコンピュータをクライアント側で起動してドメインに接続を試みたのですが、“サーバーが利用可能でない”とエラーではじかれてしまいます。
サーバー側になにか設定が足りないのでしょうか?

図2: 問い合わせメッセージの例

成するA, B, Cの3つの要素に見られる言語表現上の特徴を分析する。

A: 背景にあるゴールの提示

この部分では、願望(「～たい」など)や意志(動詞意志形、「～うと思い」)などのモダリティ表現[7]を用いて、背景にあるユーザにおける未達成のゴールが提示される⁴。

言語表現的には、A要素の記述を従属節、B要素の記述を主節にとる複文の形式が多く観察される。すなわち、A要素として提示される未達成のゴールを実現するプランの一要素として、B要素に記述されるユーザのアクションが提示される形式である。このため、従属節においては、順接の接続表現が用いられる。一方で、主節においてC要素の記述を導くような複文の形式も比較的多く観察された。この場合には、ゴールの達成を阻むものとしての問題状況が提示されるため、逆接の接続表現が用いられることが多い。

A-a: 順接的な接続表現(連用中止や原因理由の接続表現を含む)

A-b: 逆接の接続表現

B: 問題状況の要因等の提示

この部分においては、C要素で提示される問題状況のコア部分に対して、要因や契機となった(とユーザが信じている)情報が提示される。多くのものは、ユーザによる手段・アクションの記述である。B要素の記述を従属節とし、C要素の記述を主節とする複文を構成する形が多い。従属節における接続表現上の特徴は以下のように分類できる。

B-a: 時間的な前後関係を表す表現(「～して以来」など)

B-b: 条件表現(「～すると」、「～したら」など)

B-c: 逆接表現(「～したが」、「～しても」、「～したのに」など)

B-d: 原因・理由を表す接続表現(「～ので」など)

B-aは、「XPにアップグレードして以来～」のように、その効果が比較的長期に及ぶ手段・アクションの記述と共起して状

⁴ ただし、問い合わせメッセージにおいては、ゴールは陽に提示されないことしばしばである。このような場合も、ユーザにおいては、「普通に問題なくパソコンを使いたい」といった何らかのゴールが存在すると考えられるが、本分析では考慮の対象外とする。

況の変化を表現する。これに対し、B-b, c, d は、より短期的な効果を持つアクションの記述と共起する。

B-b の条件表現は、手段・アクションと問題状況の因果的関係を表現するために用いられる。B-c の逆接表現の使用は、書き手のプランに存在し、従ってその実行結果による効果を期待しているアクションが、望ましくない実行結果に終わったことを表現する。特に、何らかの理由でアクションの実行に強いコミットがない・できない場合などは、「結果を気にしながらアクションを行う」[5]意味を持つ「～してみる」⁵との共起がみられる。対照的に、B-d の原因・理由の接続表現が用いられるのは、書き手が自身のとったアクションが問題状況の要因であったことを認識している場合である。特に、そのアクションがミスや誤りであったことを自覚している場合は、「アイコンを消去してしまったので」のように、遺憾の意のモダリティ表現である「～してしまう」がそのマーカーとして用いられる。

C: 問題状況(コア部分)の提示

この部分に関する記述の表現上の典型的なパターンは、以下のように分類できる。ただし、これらが重複して現れる場合もある(例:「システムが起動できなくなってしまう」)。

- C-a: 望ましくない事態を示すモダリティ表現
- C-b: 対象ドメインにおけるトラブル状態を示す表現
- C-c: 可能表現の否定形
- C-d: 方法・手段等が不明なことを表明する表現

C-a においては、受益表現の否定形である「～してくれない」や「～してしまう」といった、話し手にとって望ましくない出来事・状況を示す表現[5]が問題状況を示すマーカーとして典型的に用いられる。

C-b においては、対象ドメインにおけるトラブル状態を示す表現(例:「フリーズする」)が C-a のようなマーカーなしに用いられる。また、対象ドメインの特徴として、「～というエラーが表示される」というような形でシステムのエラーメッセージが引用される場合も多い。なお、書き手自身がトラブル状態を発見したという記述(例:「ファイルの破損が見つかりました」)もこの分類と考える。

C-c においては、可能表現(特に「状況の可能」[7]と呼ばれる意味)の否定形が典型的に使用される。さらに、主動詞の部分に着目すると、「ダブルクリックできない」のように、プリミティブなアクションの実行不能性を示すものと、「ファイルを削除できない」のように、(サブ)ゴールに相当するより高位のアクションの実行不能性を示すものとに区別できる⁶。

C-d において頻出する表現パターンは、「～する方法がわからない」、「～にはどうすればいいのでしょうか」のような書き

手の知識が欠如していることを自身が認識しているという心的状態に関わるものである。これらはほとんどの場合、メッセージ中の情報要求の提示に直接的(あるいは密接)に関係している。このような知識欠如状態は、プランニングにおける知識前提条件(knowledge precondition [1])が満たされない状態に相当する。この状況においては、ゴール達成のためのプランを適切に構成するための知識を得る手段として、情報探索(information seeking)行動がトリガされる([1],[2])。

上記の分類において、C-d で表現されるのは書き手の心的状態であるのに対し、C-a, b, c では、ユーザの心的態度が言語表現に反映されるにしろ、客観的な事実に基づく問題状況が示される。このため、C-a, b, c については、「～(ように)になる/～くなる」に代表される状況の変化(変化の構文[7])を示す表現や、問題状況が定常的であるか(例:「いつも～してしまう」)、不定期に発生する(例:「たまに～になる」)ものかを示す表現がしばしば付随する。さらに、おそらくは記述に客観的な推定のニュアンスを持たせるために、「～ようなのです」のような複合的なモダリティ表現(推定(ようだ)+説明(のだ))が用いられることがある。

5 問題状況の類型化

前節に示した A, B, C の各要素における言語表現上の特徴とその分布に基づいて、対象のドメインにおける問題状況の類型化を試みる。()内は、対象メッセージにおける頻度である。

α: プラン立案に関わるもの

理性的な主体においては、ゴール達成へ向けてプランを立案し、それを実行することが行われる。もし、前述の知識前提条件が満たされずにプランが立案できていなければ、ゴール達成へ向かうことはできない。このため、知識前提条件が満たされないという状況を問題状況の一類型と考える。

- Knowledge precondition failure (50件): ゴール A を達成するために必要な知識が欠如していて、プランを立案ができない状況(例:「2人で同じメールアドレスを使うには、どうしたらいいでしょうか」)。

β: 手段・アクションの実行に関わるもの

一方、プランが存在する場合においては、プラン実行のある段階でプランを構成する手段・アクションが「実行不能」であったり、実行が「失敗」したりすれば、プランを継続することはできず、ゴールを達成することはできない。

手段・アクションの実行に関わる問題状況は、下記のように分類できる。なお、分析データにおいては、背景にある上位ゴールとの関係を含めて手段・アクションの実行に関する状況が述べられる場合も多いが、上位ゴールが明示されていない場

⁵ 「～してみたところ～」のような逆接ではない形式により、C 要素において結果状態である問題状況が導かれる場合もある。

⁶ ただし、「ダブルクリックするには」「マウスを持っている指先を動かす」必要があるといったように、実際には違いは相対的である

⁷ 「A するために B したいが方法が分からない」のケースは、A は B よりも高位のゴールであるとみなし、ゴール B に対する知識前提条件の問題であると考えられる。

合も多い(脚注4参照)。以下では、明示的・非明示的なゴールの存在を排除し、手段・アクションと問題状況(コア部分)の関係のみを類型化する。

● **Action precondition failure (6件)**: プラン実行の過程において、プラン中の手段・アクションを実行するための前提条件が満たされず、これを実行できない状況 (例:「XPの起動を早くしようと思い、BootVisをダウンロードしようとしたところ、MSのページにBootVisがありません」)。

● **Ineffective action (44件)**: プラン実行の過程において、実行が行われた手段・アクションが意図した効果をもたらさず、プランの実行・継続ができなくなる状況 (例:「ダイアルアップ接続を行おうとしています、ユーザ認証で拒否されます」、「検索を実行してもファイルが見つかりません」)。

● **Failed action (49件)**:

◆ プラン実行の過程において、手段・アクションを実行したところ意図・予期せぬ結果が起こり、プランの実行・継続ができなくなる状況 (14件、例:「フォルダを開こうとダブルクリックすると検索結果が表示されてしまいます」)。

◆ 特に上記において、意図しない結果が『新たな別の問題状況』であるケース (35件、例:「このソフトを実行すると、システムがハングしてしまう」)。このケースの多くの場合、背景にあるゴールの未達成も問題であるが、生じた問題状況にユーザの関心は移動していると考えられる。

γ: **トラブル状態に関わるもの**

「ドライブを認識してくれない」、「システムがフリーズしてしまう」、「動作が極端に遅い」といった、対象ドメインにおけるトラブル状態が起こるといふ問題類型である(手段・アクション実行の結果としてトラブル状態が生じるものは、βのいずれかの項目へ分類)。全118件中、変化の構文やB要素における「～して以来」などの表現により陽にマークされるものが28件(例:「アップグレードして以来、頻繁にフリーズするようになった」)に対し、無標識のものが90件にのぼった。後者の場合は、メッセージにおける『詳細化』の部分で、何らかの関連のある情報が提示されている可能性が高く、その分析は今後の課題である。

δ: **実行不能な手段・アクションに関わるもの**

「CD-ROMドライブが使えない」のように、できるはずのことができない、あるいは、できなくなったという問題類型である。手段・アクション実行の結果として生じる問題である場合は、「Ineffective action」へ分類しているが、これらの手段・アクションの中には、ゴールとみなすべきものも含まれている可能性がある。全38件中、状況・環境の変化が変化の構文などの表現上でマークされているものが11件(例:「バージョンアップしてから、メッセンジャーに接続できなくなった」)、無標識であるものが27件であった。

ε: **その他**

上記以外には、ユーザ自身が誤ったアクションを実行してしま

うというものがある(4件。ただし、結果状態が明示されている場合は「Failed action」へ分類)。また、「OL2002を起動しておかないとメールの自動取り込みができない」の例に示されるように、問題解決のあり方に対するユーザの心的態度に起因する、ある意味で拡張された問題状況も観察された(3件)。

今回は分析することができなかったが、メッセージにおいては、問題状況の原因に関して、ユーザの現実認識や信念、推論結果といった心的態度が多く表明されており、これらを含めた分析・構造化の方法を深める必要がある。

6 おわりに

本稿では、Web上の掲示板サイトに投稿された問い合わせメッセージを対象に問題状況の記述における表現上の特徴を分析し、メッセージの背景にある問題状況の類型化を試みた。その結果、プランの立案不能、手段の実行不能・失敗、トラブル状態の発現、アクションの実行不能状態という4つの上位の類型を導いた。この類型化は、メッセージにより交換される知識情報の構造化の基礎となるが、今後は機械学習による方法論[6]等を適用することにより、入力メッセージにおける構成要素を同定・分類するための処理機構の実現を検討していく。

また、今回は少量のデータに対する内省的な分析を中心に検討を進めたが、今後はより大量のデータを用いたクラスタリングなどのデータ駆動的なアプローチを組み合わせることにより、類型化を洗練させたいと考えている。さらには、『詳細化』の部分に報告されている、ユーザによってトライされた問題解決過程の表現と構造化の手法を、ユーザの心的態度を含めた形で検討していきたい。

参考文献

- [1] Lochbaum, K. E.: The Use of Knowledge Preconditions in Language Processing. In *Proc. of IJCAI-95*, pp.1260-1265 (1995).
- [2] Marchionini, G.: *Information Seeking in Electronic Environments*, Cambridge University Press (1995).
- [3] Wilson, T., D.: Human Information Behavior. *Journal of Informing Science*, Vol.3, No.2, pp.49-56 (2000).
- [4] Xi, W., Lind, J. and Brill, E.: Learning Effective Ranking Functions for Newsgroup Search. In *Proc. of SIGIR 2004*, pp.394-401 (2004).
- [5] 金水: 文脈的結果状態に基づく日本語助動詞の意味記述。影山・岸本(編)『日本語の分析と言語類型』柴谷方良教授還暦記念論文集, pp. 47-56, くろしお出版 (2004).
- [6] 工藤, 新保: 自然言語処理におけるマイニング技術の応用. 情報処理, Vol.46, No.1, pp.41-45 (2005).
- [7] 益岡, 田窪: 基礎日本語文法 -改訂版-. くろしお出版 (1992).