課題研究報告書

理由や含意関係を対象とした質問応答システムにおける推論方式に関する調査研究

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

柳生 泰利

2014 年 9 月

課題研究報告書

理由や含意関係を対象とした質問応答システムにおける推論方式に関する調査研究

指導教員　白井清昭 准教授

審査委員主査　白井清昭 准教授
審査委員　東条　敏 教授
審査委員　池田　心 准教授

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

1210755 柳生 泰利

提出年月：2014 年 8 月

概要

質問応答システムはファクトイド型とノンファクトイド型に分類される。事実や事物、名称などを尋ねるファクトイド型質問応答システムは、検索システムとして広く利用されている。一方、より知的な質問応答を実現するために、理由や方法などの質問に対応するノンファクトイド型質問応答システムの研究も重要性を増している。

本課題研究の目的は、ノンファクトイド型質問応答システムの一つとして、理由や原因を問う質問に答える why 型質問応答システム、及び文章の多様な表現に対応するための含意関係認識に焦点を当て、その推論方式の研究の現状についてサーベイを行うことである。

why 型質問応答システムについては、情報源の文書から、理由・原因部分と帰結部分を抽出する方向で研究が進められてきた。文章抽出の方法として、手掛かり語などによる質問応答のルールベース型の手法、自動で網羅的にルールを生成するために抽象化された語彙のルールを機械学習で獲得する手法の研究がある。また、意味解析ネットワークによる手法や Web 検索エンジンを利用する手法、質問タイプに依らずに文章抽出を行う手法も研究されている。要素技術としては、why テキストセグメントの識別、意味的極性の利用、回答文と周辺文との関係利用などが研究されている。

含意関係認識は、文章間の類似度やアラインメント、構文構造の変換、形式論理、Natural Logic 等のアルゴリズム的アプローチと、知識ベースによるアプローチが並行して研究されてきた。

このような研究アプローチの事例について概括しながら、研究の現状について整理した。

目次

第1章　はじめに

1.1. 背景

情報化社会の発展に伴い、有用な情報を容易に取得できるようにすることの重要性が増々大きくなっている。情報取得を行うシステムの一つに質問応答システムがある。質問応答システムとは、自然言語による質問文を入力とし、その質問文に対する回答を知識源となる文書集合から検索し、ユーザに返すシステムである。図 1-1 は質問応答システムの一般的な処理の流れを表わす。

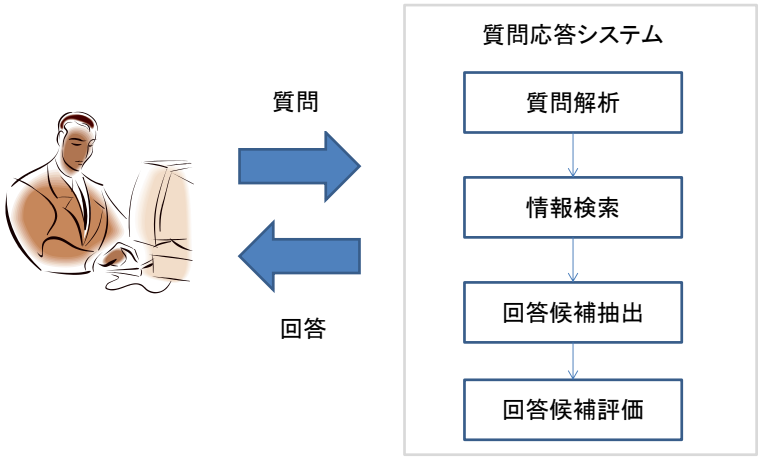

図 1-1: 質問応答システムの概要

まず、「質問解析」では質問文の形態素・構文・意味解析を行い、情報検索に必要なキーワードを得る。「情報検索」ではキーワード検索を行い、回答を含む文書の集合を得る。「回答候補抽出」では得られた文書から回答の候補をいくつか抽出する。最後に「回答候補評価」では、回答候補のスコア付けを行い、正答と思われる回答をユーザに提示する。

質問応答システムは、大きく以下のように分類される。

- ファクトイド型
  事実や事物、名称などを尋ねる質問文を扱う質問応答システムである。この場合、回答の多くは固有名詞になる。初期の質問応答システムに関する研究のほとんどはファクトイド型である。
- ノンファクトイド型
  理由や実現方法などのより知的な内容を尋ねる質問応答システムである。語の定義を問う質問を扱う定義型、原因や理由を問う質問を扱う why 型、方法を問う質問を扱う how 型などの種類がある。単純な事実を回答とするファクトイド型と比べて高度な言語処理技術を要する。

1.2. 本課題研究の目的

本課題研究の目的は、より知的な質問応答システムを実現するための各種推論方式についての調査を行うことである。特に、原因や理由を問う質問に答える why 型質問応答システムに焦点を当てる。また、文章の多様な表現に対応できるようにするために、含意関係認識も合せて調査の対象とする。これらについての研究事例を概括しながら、研究の現状についてサーベイを行う。現在実現できている最新の技術、また未解決として残されている課題をまとめ、知的な質問応答システムを精緻化するための基礎資料を作成することを目指す。

1.3. 課題研究報告書の構成

本課題研究報告書の構成は以下の通りである。2 章では why 型質問応答システムに関する研究動向の調査を報告する。3 章では、含意関係認識の技術が why 型質問応答システムの重要な要素技術であることを述べ、含意関係認識に関する主な研究事例を紹介する。最後に 4 章で報告書を総括する。

第2章　why 型質問応答システム

2.1. why 型質問応答システムへの取組み

2.1.1. 概要

why 型質問応答システムは、ノンファクトイド型質問応答システムの一つであり、原因や理由を尋ねる質問に応答するシステムである。例を以下に示す。

Q. 月の満ち欠けはなぜ起きるのですか？

A. 太陽光の月面での反射光が月の光として地球に届くので、太陽と地球の月の位置関係によって月の見え方が変わるためです。

NTCIR(NII Testbeds and Community for Information access Research) [3]は、質問応答システムの手法を評価するタスク QAC を実施している。2006 年に催されたタスク QAC-4 には、全 100 問の内、why 型質問は 3 割を占め、各チームがそれぞれ独自の手法で取り組んだ [2]。その後、why 型質問応答システムの研究は、各手法の改良や要素技術の開発などの取組みが行われている。

2.1.2. 評価方法

why 型質問応答システムの代表的な評価値について以下に述べる [2] [4] [26]。

(1) MRR

MRR(Mean Reciprocal Rank)は、ある質問に対して最上位正答順位の逆数を平均した値であり、式(1-1)のように定義される。ただし N は問題数、$rank_k$ は問題 k における正答の最高順位である。

$$\mathrm{MRR} = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{rank_k} \tag{1-1}$$

(2) 累積検索成功率

累積検索成功率は、回答候補ランキング上位が正答を含む割合である。全質問に対して、回答候補の一位以内に正解が含まれていた質問の割合は一位正解率（success@1)、十位以内に正解が含まれていた質問の割合は十位正解率（success@10)、のように呼ばれる。

(3) F 値

F 値は、Recall と Precision の調和平均として式(1-2)のように定義される。ただし Recall は正答数を質問数で割った値であり、Precision は正答数を出力数で割った値である。

$$\mathrm{F-value} = \frac{2 * Recall * Precision}{Recall + Precision} \tag{1-2}$$

2.2. ルールベース型の手法

諸岡らは、why 型の質問回答関係が成立する文の抽出パターンを発見するために、手掛かり語や単語の前後関係及び品詞情報に着目し、新聞記事と分析用の why 型質問回答セットの分析を行った[1][1]。分析の結果、手掛かり語として、「ため」「から」「ので」と「ことで」「ことにより」「ことによって」が得られた。これらの手掛かり語を基に、質問回答関係の成立する文の抽出パターンを作成した。しかし、抽出パターンを含むが質問回答関係の成立しないものも見られた。抽出パターン「～するため」を含み質問回答関係の成立する例と成立しない例として図 2-1 と図 2-2 を挙げている。

本件犯行の動機・目的は、教団に対する警察の強制捜査を阻止するため、首都中心部を大混乱に陥れるということにあった。

Q：なぜ本件犯行の動機・目的が首都中心部を大混乱に陥れるということにあったのですか。
A：教団に対する警察の強制捜査を阻止するため。

図 2-1: 質問回答関係の成立する例 [1]

訴えの提起前の時期を含め当事者が早期に証拠を収集するための手段を拡充すべきである。

Q：なぜ手段を拡充すべきであるのですか。
A：訴えの提起前の時期を含め当事者が早期に証拠を収集するため。

図 2-2: 質問回答関係の成立しない例 [1]

図 2-1 の場合は理由を回答しているため、why 型質問回答関係が成立しているが、図 2-2 の場合は目的を回答しているため、why 型質問回答関係は成立していない。

これらの例から、手掛かり語「ため」の抽出パターンと非抽出パターンを図 2-3、図 2-4 のようにまとめている。非抽出パターンとは、手掛かり語「ため」があっても why 型質問回答関係の成立する文として抽出しないことを意味する。

[1] この研究はその後の why 型質問応答システムの研究のベースラインとなっている。

動作 ＋ ため
サ変接続名詞 ＋ ため
ため ＋ 助詞「に」
ため ＋ 記号「 、，。，a-z，A-Z」
ため ＋ 助動詞 「だ」

図 2-3: 手掛かり語「ため」の抽出パターン [1]

代名詞 ＋ 助詞「の」 ＋ ため
念 ＋ 助詞「の」 ＋ ため
動詞 ＋ ため ＋ 助詞「の」
サ変接続名詞 ＋ ため ＋ 助詞「の」

図 2-4: 手掛かり語「ため」の非抽出パターン [1]

分析の結果、「ため」「から」などの他に、手掛かり語として「理由」「原因」「背景」が得られた。これらの同義語も手掛かり語となると考え、シソーラスから表 2-1 に挙げた手掛かり語のリストを得ている。

表 2-1: 手掛かり語の同義語類 [1]

理由	一因	根因	訴因
要素	遠因	罪因	動因
背景	外因	死因	導因
動機	禍因	主因	道因
事由	画因	従因	内因
根拠	起因	勝因	敗因
引き金	基因	心因	病因
ゆえん	近因	真因	副因
ゆえ	偶因	成因	福因
きっかけ	原因	善因	誘因
悪因	業因	素因	要因

ただし、「ため」などと同じく質問応答関係の成立しない例として、図 2-5 の例外パターンを挙げている。

学校側が思うほど前向きな理由での退学ではないということだろう。
営団は「原因は運輸省が調べている」といい、運輸省は「遺族対応は営団の仕事」という。
磐梯山の噴火口を背景に、中瀬沼など裏磐梯の自然を一望できる。

図 2-5: 質問応答関係の成立しない例 [1]

上記の考察を踏まえて、図 2-6 に示すパターンを「原因」「理由」を手掛かり語とする抽出パターンとしている。

助詞 「に, が, を」 ＋ 手掛かり語
格助詞 「という」 ＋ 手掛かり語

図 2-6: 「原因」「理由」を手掛かり語とする抽出パターン [1]

次に、回答範囲と質問範囲を特定するための手法を分析した。手掛かり語の分析に用いた 100 記事から質問回答関係を持つ文を抽出し、質問と回答を人手で同定し、どのような句や節などが質問範囲・回答範囲となるかを調べた。その結果、質問範囲、回答範囲を特定する以下の手法を提案している。

- 手掛かり語が「ため」「から」「ことで」「ことにより」などのとき、手掛かり語以前が回答範囲となり、手掛かり語以降が質問範囲となる。
- 手掛かり語が「原因」「理由」などのとき、手掛かり語の前にありかつ一番近い読点から手掛かり語直前の助詞の前までが回答候補となり、手掛かり語以降が質問範囲となる。

図 2-7、図 2-8 に回答範囲を特定した例を示す。

終了後に「ユーゴスラビアが現在の危機を克服し、国際的な孤立から脱却し、民主化の道に進むため、ロシアは貢献する用意がある」と声明を発表した。

図 2-7: 回答範囲の特定の例(手掛かり語「ため」) [1]

１９９８年6月、米科学誌「ブレティン・オブ・ジ・アトミック・サイエンティスツ」は、印パ核開発など**を理由**に人類破滅までの時間を示す「終末時計」を１４分前から９分前に５分早めた。

太字：手掛かり語、下線：回答範囲

図 2-8: 回答範囲の特定の例(手掛かり語「理由」) [1]

「だから」「なぜなら」のように原因-結果の関係を表わす接続詞が文頭にある場合、これらの接続詞も質問回答関係を抽出するための手掛かり語となる。ただし、質問回答関係が成立する範囲が手掛かり語の出現する１文だけでなく、直前の文も含めた複文となる。文をまたいで質問範囲、回答範囲を特定することを範囲の拡張と呼ぶ。接続詞を手掛かり語としたときの質問範囲、回答範囲の特定手法を以下に示す。

- 順接の接続詞「ゆえに、それゆえ、これゆえ、そのため、このため、だから」の場合、前の文を回答範囲として拡張を行う。
- 説明の接続詞「なぜなら」の場合、前の文を質問範囲として拡張を行う。

図 2-9、図 2-10 はそれぞれ回答範囲、質問範囲の拡張の例である。

「パキスタン政府は、イスラム諸国の中でもサウジアラビアなど親米派諸国と手をつないでいくという選択をした。**だから**、パキスタンの核技術は中東の原理主義者らには流れていない」と語る。

太字：手掛かり語、下線：回答範囲、点線下線：質問範囲

図 2-9: 文頭が「だから」のときの回答範囲拡張例 [1]

９９年３月から約２カ月半にわたるＮＡＴＯのユーゴ空爆で、連邦軍の不満は一層募る結果になった。**なぜなら**、「コソボはセルビアの発祥地」との掛け声でトマホーク・ミサイルの標的にされる兵士たちはもはや、セルビア民族主義を掲げた政権の号令を信じなくなっていたからだ。

太字：手掛かり語、下線：回答範囲、点線下線：質問範囲

図 2-10: 文頭が「なぜなら」のときの質問範囲拡張例 [1]

また、文中に読点を後ろに伴う主語が含まれるとき、その主語を回答範囲ではなく質問範囲とする。図 2-11 にその例を挙げる。

ＮＡＴＯ加盟国は、兵士の危険度が高いことと、過去の戦いで戦死者を出してきたことから、地上軍の派遣には慎重な構えを変えていない。

Q：なぜＮＡＴＯ加盟国は地上軍の派遣には慎重な構えを変えていないのですか。
A：兵士の危険度が高いことと、過去の戦いで戦死者を出してきたことから。

太字：読点を伴う主語、下線：質問範囲

図 2-11: 読点を伴う主語の例 [1]

ただし、読点を伴う主語が常に質問範囲に含まれるわけではなく、例外もある。そこで、後ろに読点「、」を伴う主語に注目して分析を行ったところ、読点を伴う主語を含む30個の文のうち28文がその主語を質問範囲とする(文頭から主語の後ろの読点までの範囲を質問範囲に加える)ことが妥当であることを確認できた。また、読点を伴う主語の末尾に出現する助詞は、「は」「が」「も」の3個であった。

さらに、逆接節と回答範囲の位置関係によって冗長部分の除去が必要なときもある。図 2-12、図 2-13、図 2-14にその例を挙げる。これらの例では逆接節の前もしくは後を質問範囲、回答範囲から除去している。

~~国連安保理決議があればイタリアは公海での船舶検査を行える~~ **~~が、~~**当時のユーゴ経済封鎖は「決議抜き」のため、イタリアの対応が注目を集めていた。

下線：回答範囲，~~打ち消し線~~：冗長部分，**太字**：逆接助詞「が」

図 2-12: 回答範囲の前に逆接節がある場合 [1]

コソボからユーゴ部隊が撤退し平和維持部隊を派遣する過程で、３万～５万人の組織に成長した解放軍が厄介な存在になるのは間違いない。このため「武装解除」を求めている**~~が、厳密な意味で武装放棄を実現させるのは現実には無理だ。~~**

下線：回答範囲，~~打ち消し線~~：冗長部分，**太字**：逆接助詞「が」

図 2-13: 原因を表わす文の直後に逆接節がある場合 [1]

地勢的に重要なことから米国が支持している**が、~~ロシアが反発しているため、今後の対露交渉が注目される~~**。

下線：回答範囲，~~打ち消し線~~：冗長部分，**太字**：逆接助詞「が」

図 2-14: 逆接節内に回答範囲がある場合 [1]

諸岡らは、上記のようなルールにしたがって回答範囲と質問範囲を特定し、それぞれ回答候補，質問候補と呼んでいる。1 つの質問に対し複数の回答候補が得られたとき、回答候補の中から質問に対応した回答を選択する必要がある。このために質問と質問候補の類似度を計算する。質問（q）と質問候補（qc）の類似度として、式(2-1)による計算方法を提案している。

$$\mathrm{R(q,qc)} = \frac{\sum_{k=1}^{N} q_k qc_k}{\sqrt{\sum_{k=1}^{N} q_k^2}\sqrt{\sum_{k=1}^{N} qc_k^2}} \qquad (2\text{-}1)$$

N = 質問文と質問候補に出現する自立語の総数

$$q_k = \begin{cases} \text{単語が q に出現している時　単語の出現回数} \\ \text{単語が q に出現していない時　0} \end{cases}$$

$$qc_k = \begin{cases} \text{単語が qc に出現している時　単語の出現回数} \\ \text{単語が qc に出現していない時　0} \end{cases}$$

質問との類似度が最も高い質問候補を求め、それに対応する回答候補を最終的な回答として出力する。

次に RE:Why という why 型質問応答システムについて述べる。渋沢らは、回答位置特定アルゴリズムと回答度スコアリングに基づく質問応答システムを提案した[5]。まず、ある事柄の結果・事実を表わす文を事実文、その原因・理由を表わす文を理由文と呼び、事実文と理由文の位置関係を次の 3 つに分類した。

(1) 理由文が事実文の前方にある場合
(2) 理由文が事実文の後方にある場合
(3) 理由文が事実文と同一の文となっている場合

そして、事実文に現れる特徴語（理由語，前方指示語，後方指示語）の組合せと理由文の位置関係を表 2-2 のようにまとめている。この表に従い、特徴語の出現の有無によって理由文の位置を決定する手法を回答位置特定アルゴリズムとする。

表 2-2: 事実文中の特徴語と理由文の位置の関係

Case	理由語	前方指示語	後方指示語	理由文の位置
Case1	○	○	×	事実文の前方
Case2	○	×	○	事実文の後方
Case3	×	×	×	事実文の後方
Case4	○	×	×	事実文と同一

ただし Case3 は、上記の組合せに加え、事実文の後方に理由語が出現することを条件とする。

渋沢らが用いた疑問語、理由語、前方指示語、後方指示語などの特徴語、及び後述する回答度スコアリングに用いる理由語、疑問語の重みを表 2-3 に示す。なお、疑問語は回答位置特定アルゴリズムでは使われず、回答度スコアリングのみに使われることに注意していただきたい。

表 2-3: RE:Why における特徴語と重みの一覧 [5]

特徴語の種類	特徴語の一覧
疑問語	（なぜ，10），（何故，10），（どうして，10），（？，5）
前方指示語 かつ理由語	（だから，41），（ですから，41），（したがって，37），（ゆえに，37），（それゆえ，41），（すると，27），（そうすると，27），（それで，23），（そこで，23）
理由語	（用言基本形＋から，41），（用言基本形＋からこそ，45），（ことから，33），（ところから，33），（用言基本形＋ので，41）、（形容動詞語幹＋なので，41），（名詞＋なので，41），（ため，20），（為，20），（用言基本形＋ため，37），（形容動詞語幹＋なため，37），（名詞＋のため，37），（もので，23），（おかげ，20），（用言基本形＋おかげ，41），（形容動詞語幹＋なおかげ，41），（名詞＋のせい，41），（用言基本形＋からは，27），（用言基本形＋からには，27），（名詞＋によって，31），（名詞＋により，31），（名詞＋による，31），（こととて，23），（ことで，23），（ゆえ，20），（用言基本形＋ゆえ，41），（名詞＋につき，31），（理由，20），（原因，20），（由縁，20），（所以，20），（ゆえん，20），（訳，20），（わけ，20），（由，20），（論拠，20），（証拠，20），（根拠，20），（きっかけ，20），（もと，20），（事情，20），（結果，20），（目的，20），（目当て，20），（狙い，20），（ねらい，20），（動機，20），（一説，20），（説，20），（諸説，20），（定説，20），（通説，20），（俗説，20），（所説，20），（新説，20），（自説，20），（持説，20），（旧説，20），（学説，20），（目途，20），（目処，20），（めど，20），（用途，20），（目標，20），（名目，20），（方針，20），（対象，20），（次第，20），（事由，20），（使途，20）
前方指示語	それ，これ，以上，その，この，そんな，こんな，そういう，こういう，そういった，こういった
後方指示語	以下，次，以後，以降，後述，示す，挙げる，挙がる，述べる

表中、各語句の後ろにある数値は重みを表わす。

理由語のうち名詞は、角川類語新辞典 [6]の「学芸-論理-理由」「学芸-学術-論説」「学芸-論理-目的」の項に記載されている単語が用いられた。名詞以外の理由語、一部の疑問語については文献 [7] [8]から原因・理由を表わす表現として記載されているものが用いられた。それ以外の疑問語，前方指示語，後方指示語については体系的に述べられている文献がないことから、Web ページの文章から各特徴語にあたる文章表現を人手で調

査、収集したものが用いられた。理由語と疑問語の重みについては、文献 [7] [8]を参考に経験的に設定された。

一方、回答度スコアリングでは、why 型質問応答システムが出力した事実文と理由文の組を回答らしさの観点から評価したものを回答度スコアSと呼び、式(2-2)のように定義する。

$$S = S_f + S_r \tag{2-2}$$

ここで、S_fは事実文から算出されるスコア、S_rは抽出された理由文から算出されるスコアであり、それぞれ式(2-3),(2-4)のように定義する。

$$S_f = \sum_{q \in Q_0} w(q) + \sum_{r \in R_0} w(r) \tag{2-3}$$

$$S_r = \sum_{k=1}^{n} \left[\sum_{q \in Q_k} \frac{w(q)}{k} + \sum_{r \in R_k} \frac{w(r)}{k} \right] \tag{2-4}$$

ここで、Q_0は事実文に含まれる疑問語の集合、R_0は事実文に含まれる理由語の集合、Q_kは事実文からk文離れた理由文に含まれる疑問語の集合、R_kは事実文からk文離れた理由文に含まれる理由語の集合を表わす。また、w(q)、w(r)はそれぞれ表 1-3 に示した疑問語 q、理由語rの重みを表わす。nは理由文を構成する文の数である。

次に RE:Why の評価について述べる。まず、回答位置特定アルゴリズムを評価する。ここではテキストセグメンテーションはすべて成功していると想定し、テキストセグメンテーションは人手により行われた。評価セットとして、特定のジャンルに偏らない 10 問の why 型質問を用意し、質問文中のキーワードを検索語として Web 検索し、1,041 個の事実文を含むテキストセグメントを得た。これらのテキストセグメントのうち質問の正しい解答を含む正解セットは 258 個であり、残りの 783 個の不正解セットは、テキストセグメント内の事実文が why 型質問の内容を表わしていなかったり、理由文が含まれなかったりするものであった。回答位置特定アルゴリズムによって出力した回答は 415 であり、そのうち、正解セットに合致したものは 130 個であった。適合率は 31.3%（130/415）、再現率は 50.4%（130/258）であった。また、Case 1 から Case4 のそれぞれにおける適合率と再現率を表 2-4 に示す。なお、適合率及び再現率の分子は回答位置特定アルゴリズムが出力した回答の中で正解セットと一致した回答の数、適合率の分母は回答位置特定アルゴリズムが出力した回答の数、再現率の分母は正解セットに含まれるテキストセグメントの個数を表わす。

表 2-4: 回答位置特定アルゴリズムの適合率と再現率 [5]

分類	適合率	再現率
Case1	13.8%（12/87）	34.3%（12/35）
Case2	41.2%（7/17）	70.0%（7/10）
Case3	34.5%（59/171）	51.8%（59/114）
Case4	37.1%（52/140）	52.5%（52/99）
総計	31.3%（130/415）	50.4%（130/258）

評価セットの 1,041 個のテキストセグメントの内、正しく正解または不正解を判定されたものは計 705 件（130 + 575）であった。残りの 336 件の内訳は、正解セットに対する誤判定 128 件、不正解セットに対する誤判定 208 件であった。正解セットに対する誤判定の原因は、表 2-5 に示すように 4 種類に分類された。

表 2-5: 正解セットに対する誤判定を導く原因 [5]

原因の内容	件数（件）	割合（%）
(1)特徴語の検出漏れ	33	25.8
(2)想定と違う役割の特徴語	46	35.9
(3)特徴語が明示されない	41	32.0
(4)理由語によって抽出が終了	8	6.3
合計	128	100.0

また、不正解セットに対する誤判定の原因は、表 2-6 に示すように 2 種類に分類された。

表 2-6: 不正解セットに対する誤判定を導く原因 [5]

原因の内容	件数（件）	割合（%）
(5)事実文が質問の内容ではない	104	50.0
(6)理由文が事実文と関係ない	104	50.0
合計	208	100.0

渋沢らは、表 2-6 に示した原因を課題と捉え、その対策を表 2-7 のようにまとめている。

表 2-7: RE:Why の課題と対策 [5]

(1)	課題	特徴語の検出漏れ
	対策	特徴語の追加によりある程度解決可能だが、かえって正解の検出精度が低下する可能性もあるため、追加する特徴語の吟味が必要である。
(2)	課題	想定と違う役割の特徴語
	対策	指示語が指し示す対象を正しく解釈できるようにするために、特徴語の単純なマッチングのみでなく、照応関係の解析が必要である。
(3)	課題	特徴語が明示されない
	対策	前方指示語、後方指示語という明示的な照応詞が用いられる照応関係だけでなく、前の文からの自立語の繰返しや、「ゼロ代名詞」[9] と呼ばれる照応詞の省略にも対応した照応関係の検出が必要である。そのためには、より多くの照応表現に対応する必要があり、意味に踏み込んだ自然言語処理が必要である。
(4)	課題	理由語によって抽出が終了
	対策	例えば、文章の見出しとして体言止めの形式で理由語が使われている場合など、事実文と理由文が同一であると判定した場合（Case4 と判定した場合)、事実文の周辺の文を解析することなく回答を抽出して処理を終了してしまう。そのため、実際には事実文の周辺に理由文が書かれていたとしても、回答文が不十分な形で出力されてしまう。対策として、特徴語の有無だけでなく、構文解析や HTML タグ情報の利用によって、事実文が見出しであるかどうかを判定するルールの導入などが考えられる。
(5)	課題	事実文が質問の内容ではない
	対策	検索語の有無だけでなく、意味に踏み込んだ言語処理が必要であり、難しい問題である。
(6)	課題	理由文が事実文と関係ない
	対策	(5)と同様、意味に踏み込んだ言語処理が必要であり、難しい問題である。

さらに、渋沢らは、20 文の why 型質問セットを用い、RE:Why によって出力された回答を評価する実験も行った。その結果、MRR(top-10)は 0.552、success@10 は 37.1%となった。MRR が比較的高いが、これは質問セットが「なぜ空は青いのですか？」といった一般によく質問される質問を多く含んでいるため、Web における FAQ サイトなどから正しい回答を抽出できたためではないかという指摘もある[2]。

渋沢らは今後の課題として、以下を挙げている。

1)網羅性の問題

why 型質問は正解が必ずしも 1 つとは限らないため、様々な回答をできるだけ多く抽出すること。

2)冗長性の問題

同じ回答が複数の文書で述べられることがあるため冗長性を排除すること。

3)因果関係の連鎖性の問題

why 型質問と回答のような因果関係は連鎖的に連なっているため、どこまでを why 型質問の回答とみなすかを判断すること。

2.3. 原因表現獲得ルールの自動抽出手法

why 型質問応答システムでは原因を表わす文章が回答となることが多いため、テキストから原因を表わす表現を自動的に検出することが重要な要素技術となる。しかし、乾ら [10]が指摘するように、原因を表わす表現は多様であり、かつ明示的な特徴のない表現も多い。そのような明示的な特徴を持たない原因表現を検出するためのルールを人手で見つけることには限界がある [11]。そのため、東中らは、why 型質問回答パターンのルールを手作業でなく自動抽出する手法を提案した [12] [13]。

原因表現特徴量の自動抽出

東中らは、原因表現を自動抽出するパターンとして、ATS(abstracted text span)パターン、BACT(a Boosting Algorithm for Classification of Trees) [15]パターンを導入した。いずれのパターンも EDR コーパス [16]を元に生成している。

（１）ATS による原因表現パターン

EDR コーパスには図 2-15 に示すような意味情報が付与されている。

```
[[main 10:混雑:0f1e00]
 [object 14:空:0ff656]
 [cause [[main 8:航空便:3bd65b]
        [modifier [[main 6:運: 1e85e6]
                  [object [[main I#1:帰省客:”=Z帰省するための客”]
                          [modifier 1:お盆:0e800f]]]]]]]
```

図 2-15: EDR コーパスにおける意味表現の例

図 2-15 の例では、「お盆の帰省客を運ぶ航空便」と「混雑」の間に"cause"という関係があることが分かる。また、EDR コーパスでは次のようにそれぞれの単語に番号が振られている。

「/1:お盆/2:の/3:帰省/4:客/5:を/6:運/7:ぶ/8:航空便/9:で/10:混雑/11:する/12:真夏/13:の

/14:空/15:で/16:、/17:ヒヤリと/18:する/19:出来事/20:が/21:起き/22:た/23:。/I#1://3:帰省/4:客//」

なお、「帰省客」には I#1 という単語番号とは異なる複合語番号が振られており、これは単語番号 3 と 4 によって構成されることを表わしている。ここで、“cause”を構成する要素の単語番号を元に、それらを包含する単語を抜き出すと、単語番号 1～8 の単語群が相当し、「お盆の帰省客を運ぶ航空便」という文字列が得られる。

ただし、機能語を主要素とする表現パターンを獲得するためには、文節を区切りとした表現の獲得が望ましい。そこで、同じ文を構文・依存構造解析器 CaboCha [14]などのツールによって係り受け解析を行うと次のような文節区切りが得られる。

「お盆の/帰省客を/運ぶ/航空便で/混雑する/真夏の/空で、/ヒヤリと/する/出来事が/起きた。」(/は文節区切りを表わす)

「お盆の帰省客を運ぶ航空便」に対応する文節を抽出すると、「お盆の/帰省客を/運ぶ/航空便で」という文節列が原因を表わす表現として得られる。

このようにして得られた表現に対して、形態素解析を行うことによって、より抽象度の高い表現に変換する。具体的には、機能語（助詞、助動詞、非自立の名詞・動詞・形容詞、動詞―接尾）以外を、一つ以上の任意の自立語にマッチするアスタリスク記号（*）で置き換える。例えば、「お盆の帰省客を運ぶ航空便で」は「*の*を*で」のように変換される。

以上の処理によって得られた原因表現パターンが ATS パターンである。EDR コーパスから得られた ATS パターンの内、出現頻度の大きい上位 20 パターンを表 2-8 に示す。

表 2-8: ATS 手法により得られた原因表現パターン（[12] [13]より作成）

No.	原因表現パターン	出現頻度
1	で	985
2	の*で	659
3	に	328
4	の	233
5	が	164
6	による	160
7	の*に	158
8	は	135
9	によって	131
10	の*によって	88
11	の*が	76
12	から	71
13	の*の*で	64
14	の*は	60
15	を	60
16	の*から	53
17	の*による	53
18	により	51
19	ため	49
20	が*で	41

（２）BACT による原因表現パターン

EDR コーパスの全ての文を“cause”を意味情報に持つ文と持たない文に分け、“cause”を持つ文に特徴的に現れる表現を統計的分析によって獲得する。このような統計的分析を行うツールとして BACT [15]が用いられた。BACT は、boosting をベースにして木構造を分類できる機械学習アルゴリズムである。boosting は、弱学習器という精度の低い単純な分類器の出力を複数組み合わせることにより、より精度の高い分類ができるように学習する機械学習手法である。判定誤りの多い弱学習器には小さい重みを、判定誤りの少ない分類器には大きい重みを与える。また、BACT における弱学習器は、最初は木構造中の単純な性質の有無を用いるが、次第に大きな構造（部分木）の有無を利用する。

BACT では、“cause”を含む文の木構造にはラベルとして「+1」を付与し、“cause”を含まない文の木構造にはラベルとして「-1」を付与する。ラベル付けされた複数の木構造を入力とし、どのような部分木の存在が木構造全体を「+1」または「-1」とラベル

付けされるために寄与しているかを boosting に基づき数値化する。

BACT を用いる場合、その前処理として、下記の処理(a)～処理(c)を行い、文を木構造に変換する。図 2-16 は「X は詐欺で逮捕された」を例とした木構造への変換の流れを示している。

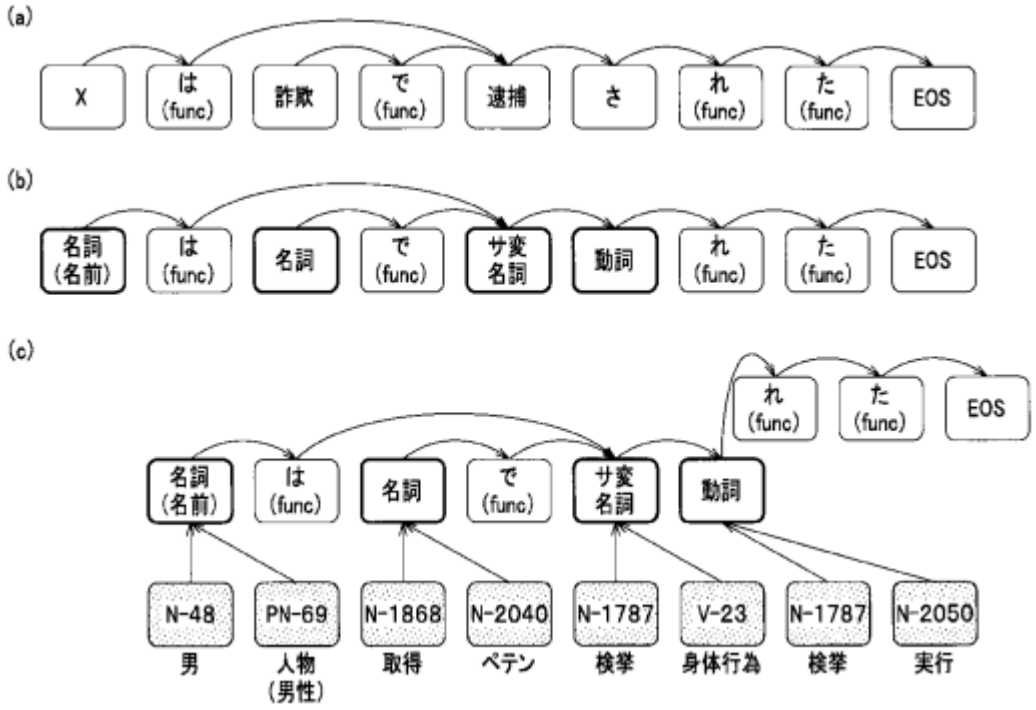

図 2-16:文の木構造への変換 [13]

処理(a)

対象となる文を構文・依存構造解析器 CaboCha により解析し、単語をノードとする木構造を生成する。

処理(b)

機能語（助詞、助動詞）以外の単語については、それぞれの品詞情報で置き換える。品詞情報のノードを品詞ノードと呼ぶ。

処理(c)

対象となる文を morph [17]と JTAG [18]により解析する。morph は日英翻訳システム ALT J/E に付属する言語解析器であり、文に含まれる各単語に意味属性を付与することができる。JTAG も同様に、文に含まれる各単語に対して用言意味カテゴリ、固有名詞カテゴリを付与することができる。この解析で得られた情報を用いて、品詞ノードに対応する意味属性・用言意味カテゴリ・固有名詞カテゴリを子ノードとして追加する。

なお、処理(b)は機能語以外の単語品詞による置き換え、処理(c)は意味属性・用言意味カテゴリ・固有名詞カテゴリの付与を行っているが、これらは木の汎用性を高めること

が目的である。EDR コーパスの文を汎用的な木構造に変換した後、“cause”を含む木構造に頻出する部分木を BACT で求め、これを原因表現の抽出パターンとして獲得する。EDR コーパスから BACT により得られた原因表現パターンの内、BACT が付与する重みの上位 20 パターンを表 2-9 に示す。

表 2-9: BACT により得られた原因表現パターン（[12] [13]]より作成]）

No.	原因表現パターン	重み
1	で 名詞-一般 N-1398 ）の	0.0062
2	により	0.0031
3	,　より　に	0.0028
4	によって	0.0028
5	ので	0.0026
6	による	0.0021
7	ため	0.0020
8	ある で	0.0020
9	。	0.0018
10	N-2455	0.0015
11	ため　この	0.0013
12	N-1265	0.0013
13	で 名詞-サ変接続	0.0013
14	N-2115	0.0012
15	形容詞 で	0.0012
16	N-2558	0.0012
17	, で の	0.0011
18	, から	0.0011
19	, で	0.0010
20	動詞 を-,は 名詞-サ変接続	0.0010

表 2-9 において、例えば、「で 名詞-一般 N-1398 ）の」は、図 2-17 に示す部分木を表わす。これは「～の（N-1398 という意味属性を持つ名詞）で～」という表現に対応する。

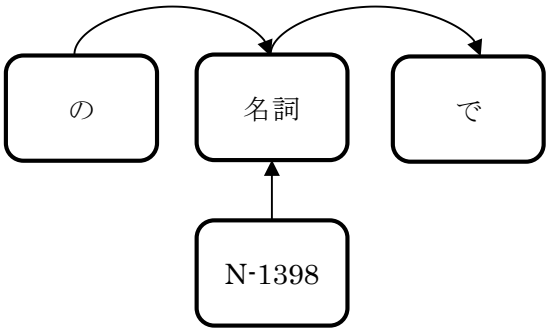

図 2-17: 得られた原因表現パターンの木構造（例）

なお、N、V、PN から始まるものはそれぞれ意味属性、用言意味カテゴリ、固有名詞カテゴリを表わす。表 2-9 においては、N-1391(疑い)、N-2455(理由)、N-1265(驚き)、N-2115(動揺)、N2558(忙)は意味属性を表わす。表 2-9 には現れていないが、V、PN から始まるものの例として、V-31(感情動作)、PN-87(公共機関名)などがある。

回答候補の特徴量

東中らは、回答候補を抽出するために用いる特徴量として以下を用いている。

(1) 原因表現特徴量（Causal Expression Features）

・AUTO-ATS-Causal Expression Features

EDR コーパスから自動獲得された 402 個の ATS パターン

・AUTO-BACT-Causal Expression Features

EDR コーパスから自動獲得された 669 個の BACT パターン

・MAN-Causal Expression Feature

人手で作成する手法 [19]により作成したパターン

(2) 内容類似度特徴量（Content Similarity Features）

・質問文と回答候補を単語の出現頻度ベクトルで表現したときの両者のコサイン類似度

・回答候補のランキング順位の逆数（ランキング順位は後述する文書獲得エンジンによる）

・同じ単語または同義語が、回答候補と質問文のそれぞれに存在すること。

(3) 原因結果関係特徴量（Causal Relation Feature）

EDR 概念辞書から、原因（cause）と結果（effect）を表わす単語のペアを抽出し、原因を表わす語が回答文にあること、結果を表わす語が質問文にあることを判定材料とした。原因語、結果語のいずれについても EDR 概念辞書から得られる同義語も展開し、355,641 組の原因―結果ペアを作成した。

質問回答システムの全体処理と評価

東中らの why 型質問応答システム NAZEQA は、前述の特徴量を用いて以下のように動作する。

1. 質問は、質問タイプが「理由」（REASON）である質問を抽出するために、ルールベース型の質問解析コンポーネントによって解析される。
2. 文書獲得エンジンは、DIDF（Decayed IDF）[20]という IDF に類似の測定手法によって、1998 年から 2001 年までの毎日新聞から、n-best 文書を抽出する。n として 20 を選択した。n 文書中の全ての文章／パラグラフは、回答候補として抽出される。回答候補として文章またはパラグラフのどちらを使うかは設定可能である。
3. 特徴抽出コンポーネントは、各回答候補から前述の(1)原因表現特徴量、(2)内容類似度特徴量、(3)原因結果関係特徴量を抽出する。
4. QA コーパスによって学習した SVM ランク付け器によって、抽出された特徴に基づき、回答候補をランク付けする。
5. 上位 N 件の回答候補が回答としてユーザに示される。

東中らは、システムの評価として、1,000 文の why 型質問セットを用いて実験し、文単位の MRR(top-10)は 0.244、段落単位の MRR(top-10)は 0.349 という結果を得た。

2.4. 意味解析ネットワークによる手法

原田らは、語の意味と語間の深層格を同定する意味解析を基礎とし、意味グラフベースで質問文と知識文を照合して類似度の高い知識文から回答抽出を行う質問応答システム Metis を開発した [21] [22]。Metis では、質問文と知識文の内容を照合するために、形態素解析と係り受け解析に加え、原田らの開発した Sage [23]によって意味解析と照応解析を行い、その結果を図 2-18 に示すような意味グラフの形式で出力する。意味グラフでは、各語に EDR 電子化辞書 [16]中の語義（6 桁の 16 進数）が割り振られ、質問文と知識文は、共通部分グラフにおけるノード対の類似度とアークグラフ対の類似度から求められるグラフ類似度によって照合が行われる。

図 2-18 の場合、2 文間における類似共通部分グラフは、4 つのノード対と 3 つのアーク対から成る。これらのノード対ならびにアーク対の類似度を基に質問文と知識文の類似性を測る。質問文との類似度が高い知識文を検索し、これを回答として出力する。

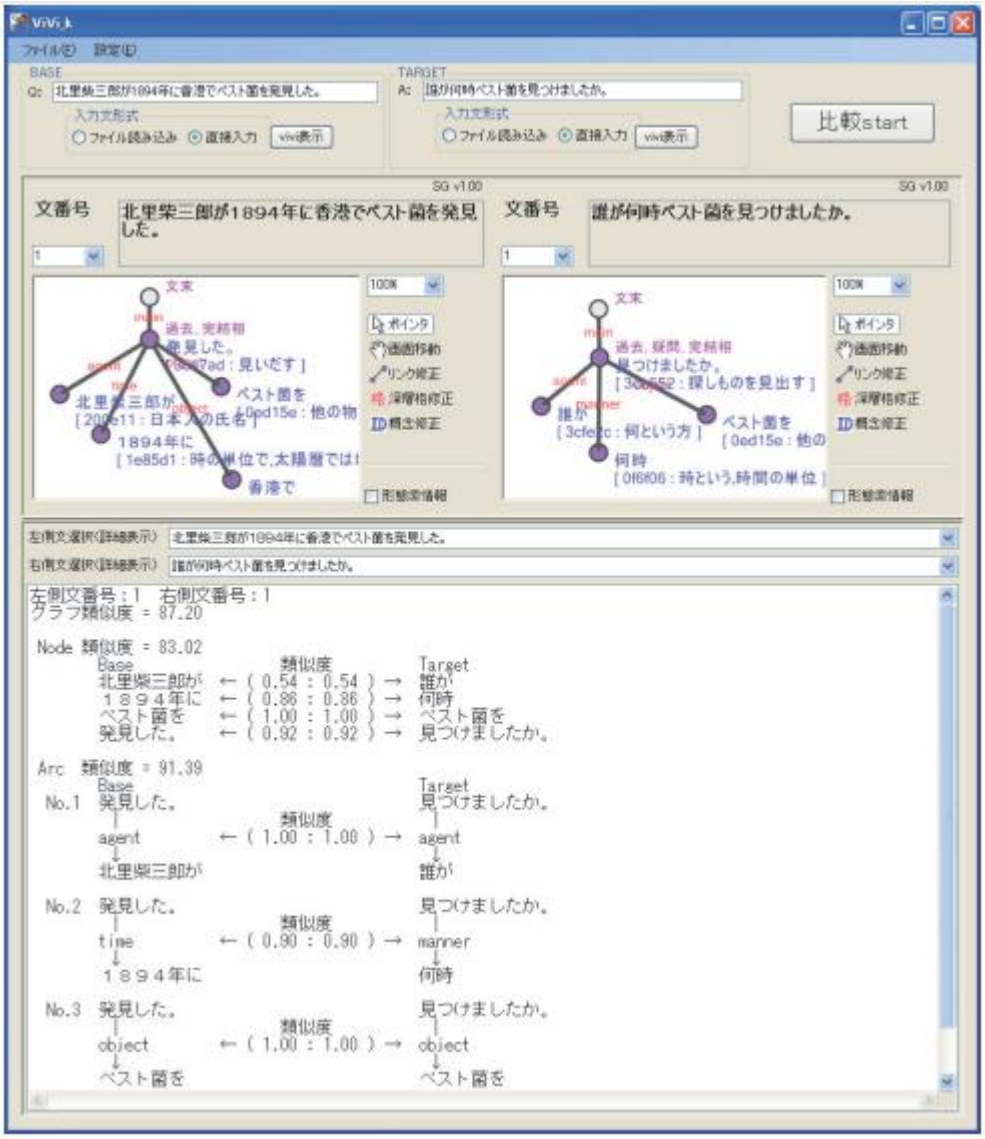

図 2-18: 質問文と知識文との意味的対応 [22]

Metis のシステム構成概要を図 2-19 に示す。

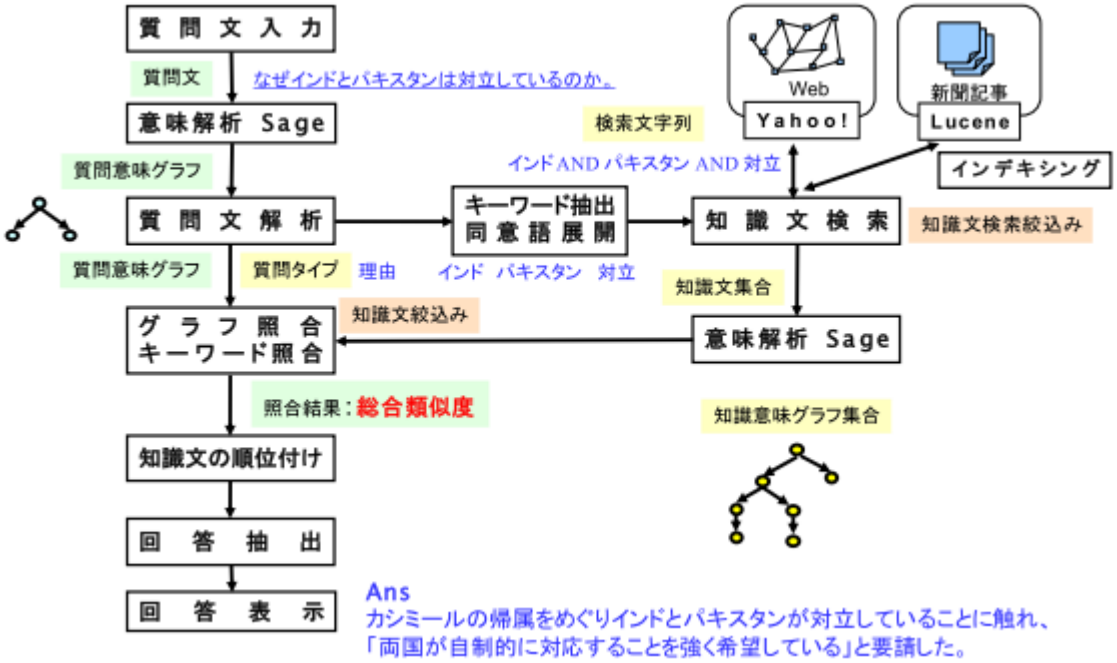

図 2-19: Metis システム構成図 [24]

Metis の処理の流れを以下に述べる。

1. 質問タイプの分類

ファクトイド型の質問を 12 タイプに、ノンファクトイド型の質問を why, how, definition の 3 タイプに分類する。

2. 検索キーワードの抽出

検索キーワードは質問グラフのノード（文節）単位で抽出を行う。検索キーワードは、Normal キーワードと、質問の核となる Must キーワードに分けて抽出する。知識ベース検索の際には Must キーワードを必ず含めるようにして、求める知識と関係ない不要な知識が多く検索されることを避ける。

3. データベース検索による知識文の取得

知識文のデータベースとして Web と新聞記事を用いる。Web 検索には検索エンジンとして Google[2], Yahoo![3]等を利用し、新聞記事の検索には検索エンジンとして Lucene[4]を利用する。また、知識文の検索精度を高めるために、質問文から抽出したキーワードに深層格（語の役割）を含めたクエリを用いて検索を行う。また、検索インデックスを作成する際に、知識文に含まれる単語と深層格の組もインデックスに加える。その手続きは以下の通りである。

R1）知識文の意味グラフにおけるノードの出力辺の深層格を、登録する語の後ろに「+

[2] https://www.google.co.jp/
[3] http://www.yahoo.co.jp/
[4] http://lucene.apache.org/

深層格」として登録する。

R2) 知識文の意味グラフにおけるノードの入力辺の深層格を、登録する語の前に「深層格+」として登録する。

例えば、図 2-20 に示すように、「1979 年に、米中が国交を正常化した」という知識グラフに対しては、正常化+time, 正常化+agent, 正常化+object, time+1979 年, agent+米中, object+国交の 6 つが登録される。このように深層格を付与してインデックスを登録することにより、この文が「正常化」の「時」「主体」「相手」を知識に持つことが分かる。

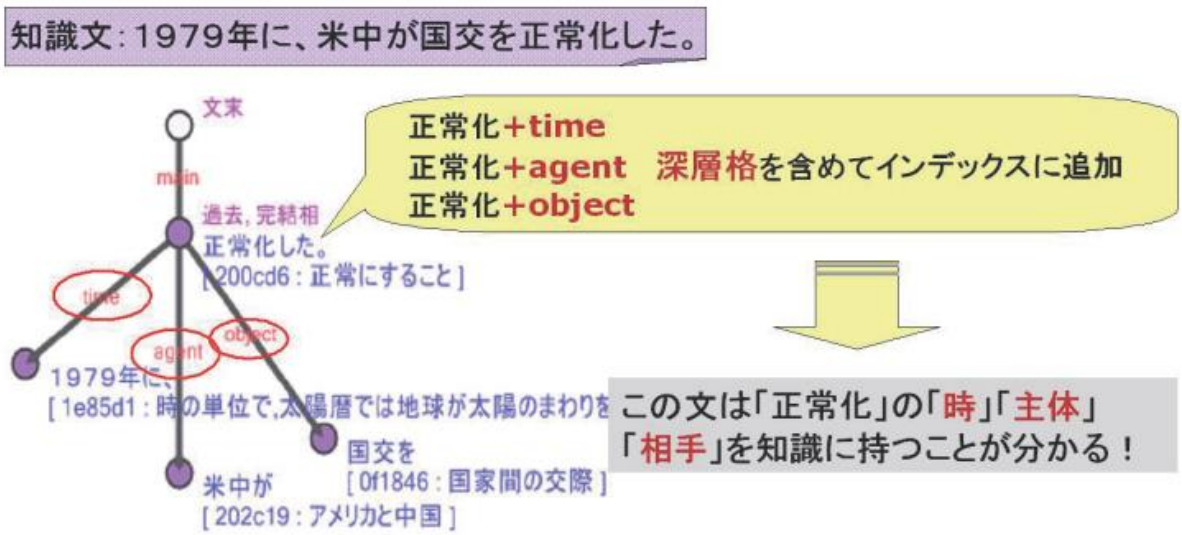

図 2-20: 語と深層格をペアにしたインデックス作成 [22]

4. 質問グラフと知識グラフの照合

ここでは質問文を解析して得られた意味グラフを「質問グラフ」、知識文の意味グラフを「知識グラフ」と呼ぶ。まず、質問グラフのノードとそれに対応する知識グラフのノード間の概念類似度（これをノード類似度という）を計算する。また、質問グラフノード間のアークとその両端のノードに対応する知識グラフノード間のアークの類似度(これをアーク類似度という)を計算する。

ノード類似度はそれぞれのノードが持つ概念の類似度とし、２つの概念c_1とc_2の概念類似度は EDR 辞書の概念体系の構造上での共通上位概念$\mathrm{c}(c_1, c_2)$の深さにもとづいて式(2-5)によって求める。

$$\text{概念類似度} = \frac{2\times \mathrm{d}\left(c(c_1,c_2)\right)}{d(c_1)+d(c_2)} \qquad (2\text{-}5)$$

ここで、$\mathrm{d}(c)$は概念cの深さを表わす。一方、アーク類似度はそれぞれのアークが持つ深層格が表 2-10 に示す深層格の類似グループのどれに共に属しているかによって求める。

同じグループに属さないアーク間の類似度は 0 とする。

表 2-10: 深層格グループ [22]

グループ名	属する深層格名	アーク類似度
動作の主体	agent,o-agent,a-object, object,scene	0.90
時系列	time,time-from,time-to, duration,sequence,reverse, cooccurence,manner	0.90
動作の対象	object,goal,implement, material,source,o-agent, basis,beneficiary	0.85
修飾表現	a-object,modifier,possessor, manner	0.90
理由・原因	cause,reason,manner	0.80
動作の目標	goal,beneficiary,purpose manner	0.85
場所	place,goal,from-to,location, scene,source,manner	0.90

ノード及びアークの類似度から式(2-6)(2-7)(2-8)でグラフ類似度を求める。

$$\text{グラフ類似度} = \text{ノードグラフ類似度} + \text{アークグラフ類似度} \tag{2-6}$$

$$\text{ノードグラフ類似度} = \frac{\Sigma(\text{ノード類似度} \times \text{ムード得点})}{\text{質問グラフのノード数}} \times 50 \tag{2-7}$$

$$\text{アークグラフ類似度} = \frac{\Sigma(\text{アーク類似度} \times \text{ムード得点})}{\text{質問グラフのアーク数}} \times 50 \tag{2-8}$$

ここで、ムード得点とは、ノードが持つ［断定」や「疑問」や「過去」といったモダリティを比較して決められる得点である。例えば、「発見した」と「発見していない」というノードを比較する場合、共に「発見」という概念であるため概念類似度は高いが、実際の意味としては全く逆になっている。このような場合に 1 以下のムード得点を掛けることによりノード類似度を低くし、人間の感性に合った類似度を与えようとしている。

5. why 型質問の回答抽出

質問文とグラフ類似度の高い上位数件の知識文から回答を抽出する。why 型質問の場合、質問グラフ中の主述語ノードを主題ノードとし、これと対応関係にある知識グラフの主述語ノードと「reason」,「cause」,「manner」,「sequence」,「location」,「sequence」という理由を表現する深層格で結ばれている根拠ノードを根とする知識グラフ内の部分木の集合を抽出し、これを回答とする。例として、「なぜインドとパキスタンは対立しているのか」という why 型質問に対する回答抽出を図 2-21 に示す。この例では主述語ノードは「対立している。」であり、これと sequence の関係にある「近く、」ならびに cause の関係にある「めぐって」を根とする部分木を取り出している。最終的にはこれを自然言語文に復元して回答を得る。

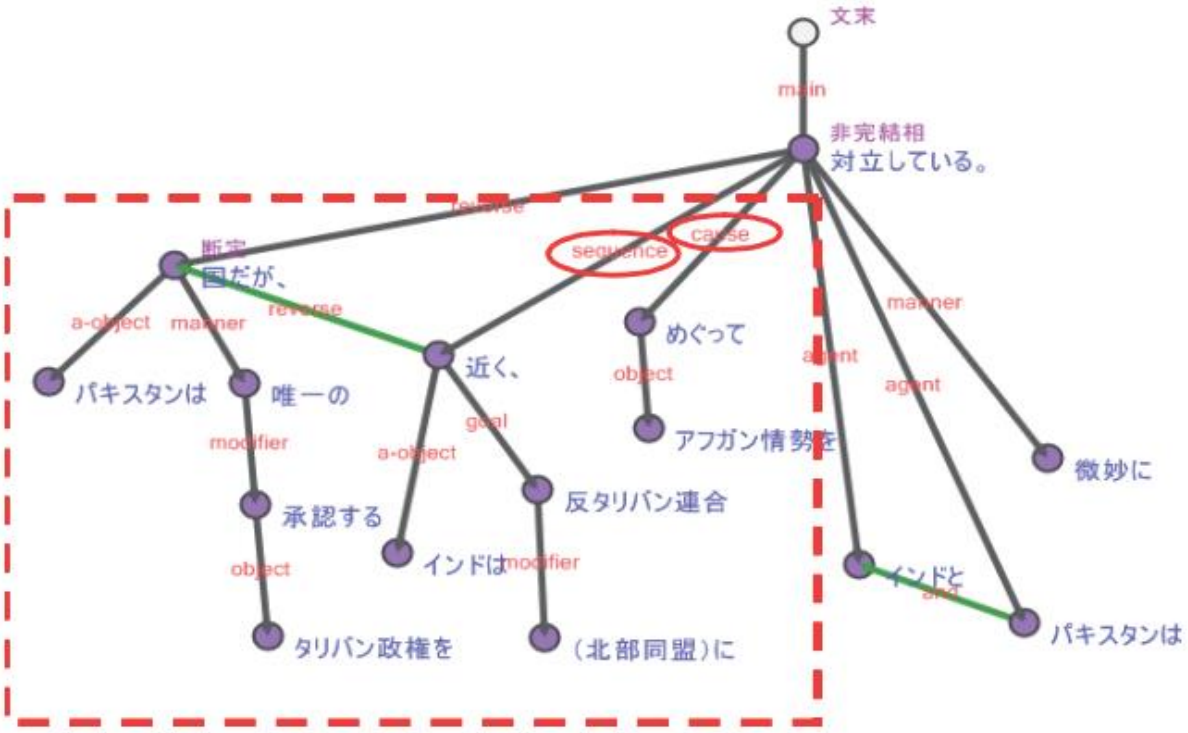

図 2-21: why 型質問の回答抽出 [22]

原田らは、理由と方法と定義を問うノンファクトイド型の質問が対象となっている NTCIR QAC のテストコレクション 100 問を用いて Metis の評価実験を行った。その結果、5 位までの累積検索成功率は 42%であった [22]。

高山らは、ノンファクトイド型質問に対する Metis の性能改善を行った[24]。Metis の課題と高山らによる対応策の概要を表 2-11 にまとめる。

表 2-11: Metis の課題と対応策 [24]

No.	課題	対応策
1	ノンファクトイド型の質問タイプ判定において、登録されていない疑問詞があった場合や２つの質問タイプ条件を満たす疑問詞や表現があった場合、正しく判定できない。	概念 ID の利用によって、より多くの表現に対応させること、及び判定条件の細密化により、質問タイプ判定の改良を行った。
2	解を含む文を検索できるが、抽出規則が少ないために回答を抽出できない場合がある。	主題ノードの選定方法の改良、回答抽出規則の追加を行った。
3	質問文と類似度が高い文からしか回答を抽出しないため、その周りに正解が存在した場合には抽出できない。	

課題１の質問タイプ判定における改良について、概念 ID による対応策を**表 2-12** に示す。表中の疑問詞の列が質問タイプ判定のパターンを表わす。すなわち質問文がこのパターンにマッチするかどうかで質問タイプを判定する。パターンには EDR 概念辞書における概念 ID(意味クラス)が使われ、単語をパターンとする場合よりも多様な表現に対応できる。また、パターンは従来の Metis で使われていたものよりも判定条件が追加されており、質問タイプの判定誤りを減らすように工夫されている。

表 2-12: 質問タイプ判定ルール （[24]より作成）

質問タイプ	No.	疑問詞	概念 ID 詳細（ID	語意）	
why 型	(1)	1f3b0e、100f27	1f3b0e	どうして
	(2)		100f27	どのような理由でという疑問を表わす気持ちであるさま
	(3)	(1014db、3cf234、3d1b8f、1ecb6f) ＋ (為、ため、訳、わけ、0e4135、10e038、3ce8af、3cf780)	1014db	どのような
	(4)		3cf234	不特定の物事の何か
	(5)		3d1b8f	いずれの
	(6)		1ecb6f	何のための
	(7)		0e4135	ある物事がそのようになった事情
	(8)	definition 型(1)(2)の表現があり、文中に(訳、わけ、為、ため、0e4135、10e038、3ce752、3ce8af、3cf780)を含む	10e038	結果や結論に至る事情
	(9)		3ce752	成し遂げようとして目指す事柄
	(10)		3ce8af	物事の原因である事柄
			3cf780	ある物事に背後から影響を与えている勢力や事情や環境
how 型	(1)	(1014db、3d04f9)＋(動詞節、する、やる、～する、～やる)	1014db	どのような
	(2)	(どう～、1014db、3d04f9)＋(方法、手段、経緯、メカニズム、3bbe28、3d0201)	3d04f9	内容としてどのように
	(3)	definition 型の(1),(2)の表現があり、文中に(方法、手段、3bbe28、3d0201)を含む	3bbe28	方法
	(4)	どうする	3d0201	ある事を実現する手段となる行動
definition 型	(1)	(～とは、～って)＋疑問属性を持つノード	1014db	どのような
	(2)	(1014db、3d04f9)を含み文末が動詞節	3d04f9	内容としてどのように
	(3)	(1014db、3d04f9)＋定義		
Yes/No 型	(1)	上記のどの疑問詞タイプにも該当せず、文末が「～か？」である		

次に課題２，３の回答抽出における改良について述べる。まず、主題ノードの選定方法の改良を行った。ここで述べる主題とは、回答抽出の手掛かりとなる質問文中の重要な語句であり、主語とは異なるものを指す。一方、質問文との類似度の高い知識データベース中の文を照応文と呼ぶ。また、照応文もしくはその周りの周辺文の中で主題ノードと対応する語句を主知識ノード、主知識ノードと特定の深層格の関係で結ばれるノードを根拠ノードと呼ぶ。主題ノードの選定方法を表 2-13 に示す。

表 2-13: 主題ノードの選定方法 [24]

質問タイプ		表現	主題
why	1	動詞節、動名詞節 ＋（理由、原因、目的）	動詞節、動名詞節
	2	（X のは）＋（どうして、なぜ、何故、なんで）	X
	3	1,2 に該当しない場合	文末
how	1	動詞節＋（方法）	動詞節
	2	1 に該当しない場合	文末
definition	1	（どんな、どういう、どのような、どういった） ＋ X（※もの、こと以外）	X
	2	1 に該当しない場合	なし

why 型の場合について、主題ノード選定の例を挙げる。

① 述語を主題とする。(表 2-13 の why の 3)

例１：なぜインドとパキスタンは**対立しているのか**。

② ただし、述語が検索に有効でない場合は疑問詞や疑問を表わす表現の前の語を主題とする。(表 1-13 の why の 1)

例２：インドとパキスタンが**対立している**理由は何ですか。

次に、以下の 3 つの場合に分けて、why 型質問に対する回答抽出規則を追加した。

1. 理由を表わす深層格を持つ場合（cause-deep-case）
2. 理由を補足する文がある場合（cause-supplementation）
3. 理由が宣言され主題が一致している場合（cause-declaration）

１．については、照応文に回答が含まれる可能性が高いため、周りの文に対しては照応文より回答を抽出する条件を厳しくして誤りを少なくする必要がある。具体的には以下のような処理となる。

（１）照応文での処理

主題ノードと対応する主知識ノードから「cause」,「condition」,「purpose」,「scene」,「location」,「sequence」,「reason」という理由を表現する深層格で結ばれている根拠ノードがあった場合、照応文を回答文とする。

（２）照応文の周りの文での処理

周辺文が主題ノードと類似度が閾値以上の主知識ノードを持ち、かつこの主知識ノードから「cause」,「condition」,「purpose」,「scene」という理由を表現する深層格で結ばれている根拠ノードがあった場合、この周辺文を回答文とする。

例を以下に示す。

質問文：９８年１月に日本が漁業協定を終了すると韓国に通告したのはなぜですか。

照応文：交渉がこう着状態になったことから、日本は今年１月２３日に現行協定の終了を通告、１年後の協定失効をにらみながら交渉を続けた。

この質問文の主題ノードは主述語である「通告した」であり、照応文の主知識ノードである「通告」より「cause」格で結ばれている根拠ノードが存在する(図 2-22 を参照)ので、この照応文を回答として抽出する。

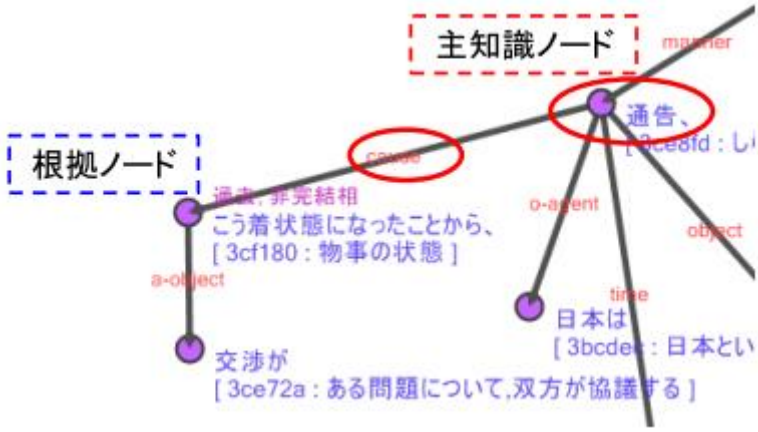

図 2-22: 理由を表わす深層格を持つ場合 [24]

２．については、理由を表現する「ため」を主辞にもち、文節のタイプが断定節であるノードから「modifier」という深層格で結ばれているノードがあった場合、「ため」を持つノードを根とする部分木を回答として抽出する。もしくは、理由を表現する「ため」を副主辞に持ち、文節のタイプが動名詞節であるノードから「modifier」格で結ばれているノードがあった場合、「ため」を持つノードを根とする部分木を回答として抽出する。例を以下に示す。

質問文：ＮＰＯが出来ても法人化されていないＮＰＯが多く存在しているのはどうしてですか。

照応文：団体の関心は高いが、当面、申請を見送ったり様子見の団体が多いのは「税金面で優遇措置がない」「申請するため必要な団体の定款づくりに時間がかかる」などネックが多いためらしい。

この場合、照応文に「ためらしい」という、主辞が「ため」で文節のタイプが断定節であるノードがあり、そのノードから「modifier」格で結ばれている「多い」というノードが存在する(図 2-23 を参照)。よって、「ため」をもつノードを根として、「「税金面で優遇措置がない」「申請するため必要な団体の定款づくりに時間がかかる」などネックが多いため」が回答として抽出される。

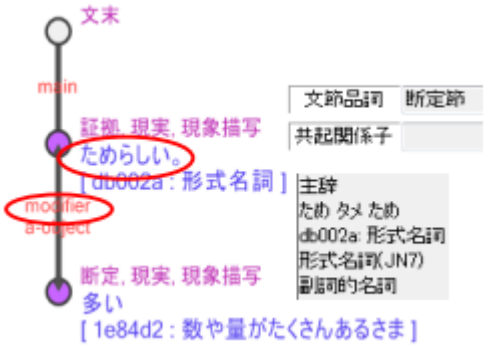

図 2-23: 理由を補足する文がある場合 [24]

３．については、理由を示す語「理由」,「原因」,「目的」,「利点」,「欠点」,「メリット」,「デメリット」,「背景」,「きっかけ」などを持つノードから modifier 格と結ばれるノードがあり、なおかつ主辞ノードに属性「断定」があった場合(文のモダリティが断定のとき)、照応文全てを抽出する。例を以下に示す。

質問文：イスラム原理主義台頭の背景には何がありますか。

照応文：原理主義の台頭の背景には、欧米型の近代化と生活様式の浸透でイスラムの宗教的、文化的なアイデンティティー、つまり主体性が失われることへの危機感があります。

この場合、質問文の主題ノードは「背景」の修飾節の「台頭の」である。知識文には、「背景には」というノードがあり、このノードと modifier 格で結ばれかつ質問文の主題ノードと一致する主知識ノード「台頭の」をもち、主辞ノードの「あります。」は属性「断定」を持つ(図 2-24 を参照)ので、「原理主義の台頭の背景には、欧米型の近代化と生活様式の浸透でイスラムの宗教的、文化的なアイデンティティー、つまり主体性が失われることへの危機感があります。」が回答として抽出される。

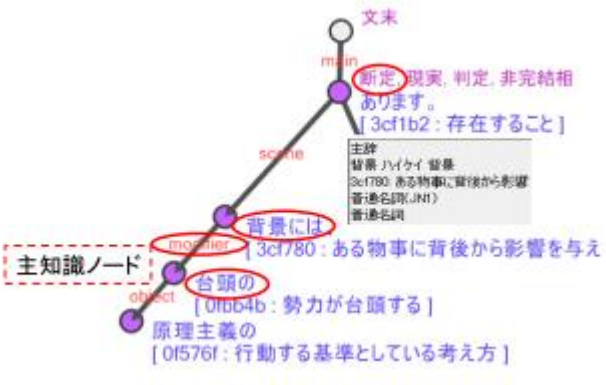

図 2-24: 理由が宣言され主題が一致している場合 [24]

上記の回答抽出規則及び規則の評価値の一覧を表 2-14 に示す。

表 2-14: 回答有望度評価方法（why 型の場合） [24]

質問タイプ	回答パターン	回答有望度	
		パターン別スコア	主語を含む 1)
why	理由を表わす深層格を持つ場合	10	5
	理由を補足する文がある場合	5	5
	理由が宣言されている場合	5	5
	理由を表わす文間深層格で接続する文がある場合	15	5

1)質問文の主語が回答文に含まれる場合

この評価値を回答有望度と呼び、従来のグラフ類似度が質問文と照応文のノードグラフ類似度とアークグラフ類似度の和であるのに対し、これらと回答有望度の 3 つのスコアの和を新しい総合類似度として回答候補の順位付けに用いた。

以上の改良により、高山らは回答抽出の精度が 19.0%から 61.0%に向上したと報告している[24]。

2.5. Web 検索エンジンを利用する手法

田村らのシステムは、質問文の言い換えを行い、検索キーワード（以下検索ＫＷ）と回答候補抽出パターン（以下回答候補抽出ＰＴ）を作成する [4]。検索ＫＷをクエリとして Web 検索エンジンに与えて、検索エンジンが返す結果から概要文（スニペット）を取り出す。得られた概要文と回答候補抽出ＰＴとの照合により回答候補を抽出する。図 2-25 は質問文解析の例であり、質問文が「空はどうして青いのですか。」のときに得られる検索 KW と回答候補抽出 PT を示している。一方、図 2-26 は回答候補の抽出の

例であり、スニペット内の適合文と回答候補抽出 PT から得られる回答候補を示している。

＜入力質問文（文中）＞空はどうして青いのですか。
＜言い換え後質問文（文末）＞空が青いのはなぜ
＜検索ＫＷ＞空が青いのは
＜回答候補抽出ＰＴ＞空が青いのは*

図 2-25: 質問文解析の例 [4]

＜適合文＞それゆえに空が青いのは、主に空気の分子による光の散乱です。
＜回答候補抽出ＰＴ＞空が青いのは*
＜回答候補＞空が青いのは、主に空気の分子による光の散乱です。

図 2-26: 回答候補を抽出する例 [4]

さらに、システムの改善策として「検索用ＫＷのスリム化」と「名詞頻度テーブルの最適化」を行っている。

検索用ＫＷのスリム化に関しては、検索エンジンに与えるクエリが長い場合、検索結果が無いあるいは数件になり、回答候補が抽出されないことがあるため、検索結果を増やすために検索ＫＷの文字数を削減する。検索ＫＷのスリム化とは、ここでは検索ＫＷを「従属節の名詞＆主節」とする方法を指す。主節の判定には助詞の「は」,「が」を利用する。日本語において主節は文の後半に置かれることが多いことから、検索ＫＷから「のは」を除いた文に対して、最も後ろにある「は」,「が」を含む節を主節と判定する。

名詞頻度テーブルとは、検索エンジンから得られた複数の概要文（回答候補）について名詞（ただし非自立語を除く）を抽出し、その出現頻度をスコアとして登録した表である。回答候補選択の際には、名詞頻度テーブルを参照し、高頻度の名詞を多く含む回答候補を選択する。名詞頻度テーブルは回答候補ランキングに大きく影響する。名詞頻度テーブルの最適化として次の４つの手法を挙げている。

1.低頻度語の削除

Web 検索結果に存在する不必要な情報（ノイズ）の影響を減らすために、出現頻度 1 以下の名詞を削除する。

2. スコア加算

why 型質問の正解には「から」をはじめとする原因、理由を表わす表現が含まれることが多いため、それらを含む回答候補に対してスコアを加算する。

3. スコア減算

why 型質問の回答に含まれにくい「当然」などの名詞は回答候補ランキングに悪影響を与えるため、それらを含む回答候補に対してスコアを減算する。

4. 名詞限定

代名詞「あれ」などの意味の無い単語による回答候補ランキングの影響を防ぐため、「一般名詞」と「固有名詞」のみを登録する。

実験結果によると、各改善手法のうち最も有効であったのは「スコア加算」であったと報告されている。

田村らは、システムの精度を確認するために、次の人手評価に基づく F 値、MRR、累積検索成功率による評価を行っている。

・人手評価

評価 A　出力された回答は正解とほぼ等しい内容である
評価 B　出力された回答は正解に加え他の情報を含む
評価 C　出力された回答は正解の一部を含む
評価 D　出力された回答は正解の内容を含まない
評価なし なし

学研サイエンスキッズ [25]の「科学なぜなぜ 110 番」の自然ジャンルを参考にして作成した 50 文の why 型質問文に対して実験を行い、F 値が 0.352、MRR が 0.409、success@10 が 54%という結果を得た[4]。ここでの F 値は、人手による評価が A または B の回答を正解とみなしている。

大友らは、田村らのシステム[4]に対して回答抽出方法の改良を行った。図 2-27 に示すように、質問文に含まれる名詞、形容詞及び動詞を任意に含み、「のは」という文字パターンを追加した回答抽出フィルタを適用した [26]。

＜質問文＞携帯電話の料金はなぜあんなに高いのか
＜回答抽出フィルタＴ＞(携帯電話 or　料金 or 高い) + “のは”

図 2-27: 回答抽出フィルタ改良例 [26]

さらに、実験結果の分析から、より多くの正答を抽出する工夫として、理由を示す文字パターンとして、「のは」の他に「から」と「ため」を追加することを試みている。このように、理由を示す文字パターンを複数用いることにより、ランキングの精度は下がるが正答抽出数は増加したと報告している。

また、正答の得られなかった質問に対し、言い換えを行うことにより正答を新たに獲得する手法を提案している。言い換え例を図 2-28 に示す。

＜質問文＞オーロラができるのはなぜ
＜言い換えた質問文＞オーロラが発生するのはなぜ
＜正答一例＞オーロラがこの領域でよく発生するのは、オーロラ発光の原因であるプラズマ粒子がほぼ磁力線に沿って動くという性質を持っているから

図 2-28: 質問文の言い換えの例 [26]

質問文の言い換えを人手で行い、質問文の再構築を行った。言い換え方法の例を図 2-29、図 2-30 に示す。

・似たような意味を持つ動詞を用いて言い換え
・似たような意味を持つ名詞を用いて言い換え
・動詞を、似たような意味の名詞に変更
・漢字を正しいものに訂正
・代名詞など不明確な表記を明確な表記に変更

図 2-29: 言い換え方法一覧 [26]

＜質問文＞どうして海には引き潮と満ち潮があるの
＜再構築後質問文＞どうして海では引き潮と満ち潮が起こるの

図 2-30: 言い換えによる質問文の再構築の例 [26]

言い換えにより正答抽出数は増加したが、言い換えを自動で行う手法の確立を課題として挙げている。

回答抽出フィルタの改良によっても回答を抽出できない例を図 2-31 に示す。

＜質問文＞
3ポイントシュートができたのはどうして
＜回答抽出フィルタＴ＞
(3ポイントシュート or　でき）+ “のは or から or ため”
＜正答＞
理由についてはいろいろですが、インサイドばかりでは・・・というのとプロのリーグでは現在のNBAと前記のABLやCBA（NBAの対抗リーグ）はあまり関係が良くなくて差別化を図るのと同時にショー的要素を強めるという理由で導入されたというのもあるみたいですね。

図 2-31: 回答抽出フィルタで回答が抽出出来ない例 [26]

この例では質問文中の単語が一切含まれず、理由を示す文字パターンも含まないため、

正答が抽出されなかった。このように、回答抽出フィルタに変換して回答を抽出する手法には限界があり、文章構造から回答を抽出する手法が必要であると報告している。

大友らは、学研サイエンスキッズ [25]の「科学なぜなぜ 110 番」の自然ジャンルを参考に 25 件、雑学のすゝめ [27]の「素朴な疑問」を参考に 75 件、計 100 件の質問を作成し、実験を行った。その結果、改良前と改良後では、F 値が 0.080 から 0.177 に向上し、検索成功率が 8%から 39%に向上した。

石下らは､質問タイプに依らない non-factoid 型の質問応答システムを検討した [28]。non-factoid 型質問は、表 2-15 に示すように､「定義型」,「理由型」,「方法型」といった典型的な型もあるが､その他の型も多く存在し､包括的に型を定義することは難しい。質問応答システムでは最初に質問文の型(質問タイプ)の判定を行うことが多いが、non-factoid 型質問応答システムではあらかじめ質問タイプを定義した上で質問文の型判定を行うことは困難である。そこで、石下らは質問の型分類を行わない手法を提案している。

表 2-15: non-factoid 型質問の分類 [28]

質問の型	質問の 記述スタイル例	回答の 記述スタイル例
定義型 (definition)	～とは何 って何	～とは・・・である ～は・・・のこと
理由型 (why)	なぜ～ ～の理由は何	～ため ～から
方法型 (how)	～にはどうしたらいい ～の方法は何	～するにはまず・・・ ～のやり方は・・・
その他 (other)	X と Y の違いは何	X は～だが，Y は・・・
	～したらどうなる	～した場合，・・・

石下らは、non-factoid 型質問に対する回答候補の適切性は、

【尺度１】質問の内容との関連性

【尺度２】質問の型に応じた記述スタイルを満たす度合

の組合せで見積もることが多いという研究報告 [29]を基礎に置いている。【尺度１】については、質問文内の内容語をキーワード群として、キーワード群に関連する語群を、Web 検索エンジンが出力する要約表示であるスニペットから収集し、質問文中のキーワード群及びそれらの関連語群の検索文書中での密度を調べることで見積もっている。【尺度２】については、あらかじめ「質問の型」に応じた回答の特徴表現を用意しておくの

ではなく、Q&A コーパスから記述スタイルが類似する質問を検索し、入力の質問と似ている質問事例を収集し、収集した質問事例に対応する回答事例から特徴表現を動的に取得することにより見積る。Q&A コーパスとして Yahoo!知恵袋のデータを利用する。約 311 万件の質問と約 1347 万件の回答の内、質問とベストアンサーのペアを「Q&A ペア」とし、この集合を Q&A コーパスとする。

石下らの提案手法の概要を図 2-32 に示す。

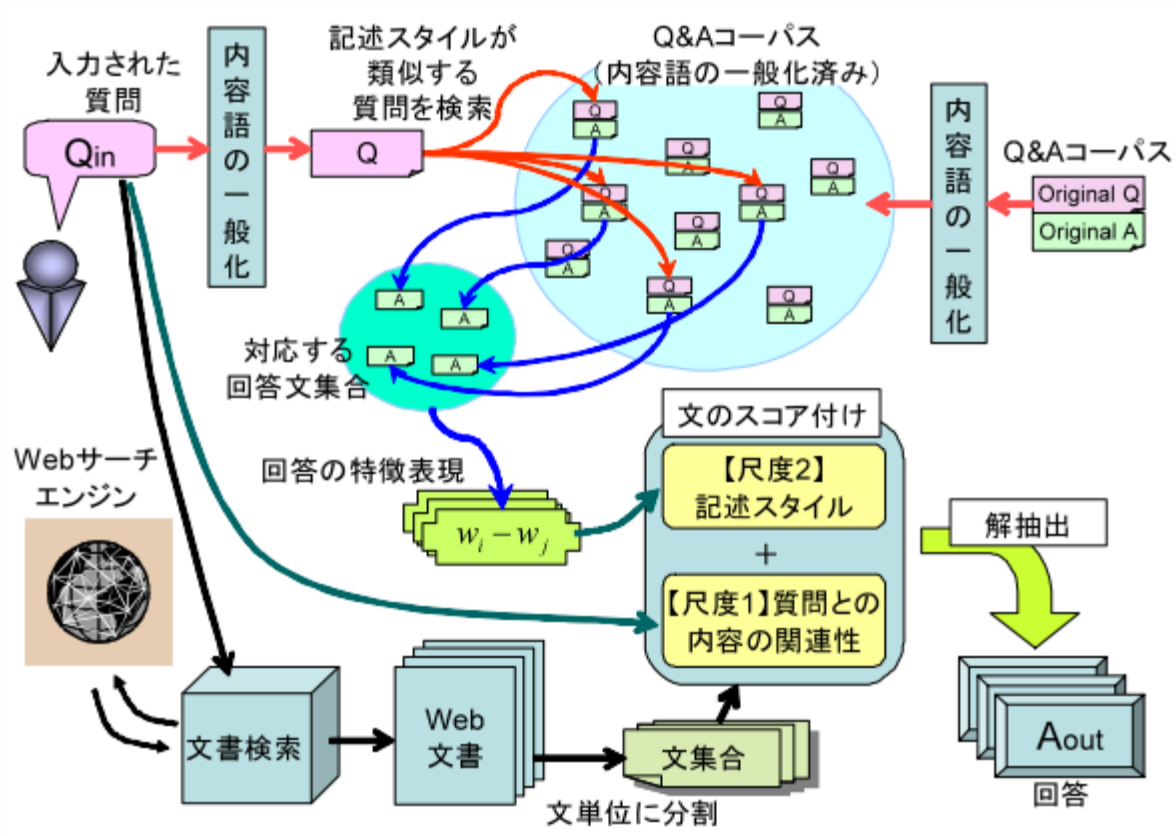

図 2-32: 石下らの提案手法の概要 [28]

石下らは、この手法の特徴として、以下の利点を挙げている。

- 型分類の失敗を考慮しなくて良い。
- 質問が入力された時点で「その質問」に応じた回答の特徴表現を動的に生成するため、入力された質問に応じて適合する特徴表現を見つけることができると期待される。
- Q&A コーパスは口語表現による質問・回答事例集合であるため、ここから取得した回答の特徴表現を用いることで、同じく口語表現で書かれた個人ホームページやブログ等の Web 文書から回答が抽出できると期待される。

石下らは、質問セットとして QAC-4 タスクの質問文 100 問（この内、理由型質問は

33 問）を用い、Web 検索エンジンとして Yahoo!JAPAN を用いて Web 文書を情報源とし、提案手法の評価実験を行った。評価結果は、33 問の理由型質問文に対して、MRR は 0.393、正解質問数は 19 問であった。

2.6. 要素技術の改良

2.6.1. 統語的特徴の利用

Verberne ら [30]は、質問応答システムのベースラインとして、質問文とパッセージ間の用語の一致度（term overlap)、パッセージの長さ、各用語のコーパス中での出現頻度を変数としてパッセージをスコアリングする QAP アルゴリズム [31]を用いた。これらの変数の長所・短所を示すために、図 2-33 の例を挙げている。

1. Why do people sneeze?
2. Why do women live longer than men on average?
3. Why are mountain tops cold?

図 2-33: why 型質問の検討例 [30]

1.は、質問文中の用語 people と sneeze に関して、日常的に使われる名詞である people よりも用語 sneeze の方がユニークであり、コーパス中の出現頻度も sneeze の方が大きくなる。回答候補のランキングの際にも sneeze を含む回答により高い重み付けがされることになり、問題はない。しかし、2.の場合、women, live, longer と average に関して、これらのコーパスにおける出現頻度に大差はないため、通常ではそれぞれは等しい重要度を有するが、この質問文の場合は後者の方が質問の本質に関して末梢的である。3.の場合、mountain と tops は通常では独立した用語としてみなされるが、この質問文では、mountain tops という複合語とし解釈する方がより適切である。このように、上記 2.と 3.の場合、回答候補の適切なランキングのためにはコーパス中での単語出現頻度の情報以外の要素が必要と考えられる。

Verberne らは、上記に関連して、最も重要度の高い回答が回答候補の上位に挙がらない問題に対して、回答候補の再ランキングの手法を導入した。再ランキングの方法として、統語的特徴（syntactic structure）を用い、質問と回答候補に対して 31 個の統語的特徴を付加し、各特徴の重み付けを遺伝的アルゴリズムにより最適化した。その結果、ベースライン・スコア、手掛かり語の存在、質問の主動詞、質問対象の焦点と文書タイトルの関係が精度向上のための主な特徴であることを報告している。

Verberne らは、評価実験のために、オンライン QA システム answers.com における

質問文を含む Webclopedia 質問セット [32]中の 805 個の why 型質問を用い、また回答コーパスとしては Wikipedia XML コーパス [33]を用いた。実験の結果、段落単位の MRR は 0.380、success@10 は 54.8%であった。

2.6.2. why テキストセグメントの識別

田中らは、why 型質問応答システムの要素技術として、why テキストセグメントを識別する識別器の構築方法を提案している [34]。why テキストセグメントとは、“原因・理由”のような因果関係の内容を含むテキストセグメントである。図 2-34 の例では、2 と 3 が why テキストセグメントであるのに対し、1 はそうではない。与えられたテキストが why テキストセグメントであるかを判別する識別器を構築することがここでの目標となる。

1. “空（そら）とは、地上から見上げたときに頭上にひろがる空間のこと。天。”
2. “青などの短波長の光が大気中の水蒸気、塵やゴミなどの微粒子にぶつかり散乱するため、空の色は青く見える。”
3. “初めのうちは波長の短い青が多く散乱され、波長の長い黄や赤の光はまだあまり散乱されないが、進んだ距離が長くなるにしたがい、波長の長い黄色や赤の光も散乱されていく。こうなると青は初めのうちにほとんど散乱されつくしてしまっていて、黄色や赤の光ばかりが地上に到達する。このため太陽が赤く見える。”

図 2-34: テキストセグメントの例 [34]

従来のルールベース型の手法 [1] [5]では、ルール作成の手間とルールの網羅性不足のため、why テキストセグメント識別精度に問題がある。一方、機械学習による why 型質問応答の従来研究[12]では、正解ラベルが付与された学習データが必要であり、学習データの大量入手が困難であること、ラベル付けされる情報は EDR 辞書の知識に制限されるため、EDR 辞書の知識に依存しない新しいルールの学習が困難であり、ランキング学習のための学習データの収集も容易ではないことを課題として挙げている。

why テキストセグメント識別器の構築手法として、Bag of Grammar（BOG）の利用を提案している。単語自体に意味のない機能語（助動詞、助詞など）を除いた単語集合、あるいは全ての品詞の単語集合は Bag of Words（BOW）と呼ばれるが、この BOW を拡張し、文法情報を表わす機能語集合を BOG と定義している。

機能語の形態素の基本形と品詞の組を Grammar 特徴と定義し、文から抽出された Grammar 特徴を BOG 特徴と呼び、BOG 特徴の集合を BOG と呼んでいる。BOG の構築に必要な情報は形態素解析の結果から得る。

① 私は学生です。
② テストなので、学校に行った。

図 2-35: Grammar 特徴の具体的概念 [34]

例えば、図 2-35 において、二つの文は機能語と内容語（下線部分）に分けられており、文①は事実を表わし、文②は理由を表わしている。この文で、名詞部分の“学生”や“学校”を例えば“教授”や“塾”に変えても、①は事実を表わし、②は理由を表わしており、提供される情報は変わるが述べられている形態（事実・理由）は変わらない。これは、“～は～です”や“～なので～”といった機能語の部分が、事実や理由を表わす要素であるためである。このような文法情報を表わす単語を Grammar 特徴と定義している。

提案手法では、Yahoo!知恵袋の約 300 万件の質問からなる質問コーパスと、その回答からなる回答コーパス（ベストアンサーコーパス）を用いて、BOG を素性として SVM（Support Vector Machine）を学習し、why テキストセグメント識別器を構築した。その結果、ルール作成に手間がかかる、識別器がドメインに依存する、ラベル付けされた学習データの入手が困難である、等の問題が改善された [34]。

2.6.3. 意味的極性の利用

呉らは、「否定的な事象の理由は否定的な事象であることが多い」、「肯定的な事象の理由は肯定的な事象であることが多い」という意味的極性に関わるパターンが why 型質問とその回答によく現れるという観察から、意味的極性のパターンを機械学習で学習し、why 型質問応答システムの性能改善を試みた [35]。

Q1: なぜガンになりますか？

A1-1: ニトロソアミンなどの発がん因が細胞のもつ遺伝子を変化させ、ガンのリスクを高める。

A1-2: 健康的な体重を維持することはガンのリスクを下げる。

図 2-36: 意味的極性に関わる質問回答の例 [35]

図 2-36 におけるように、「なぜガンになりますか？」という「否定的な事象」を説明する A1-1 と、ガンを予防するための「望ましいこと」を説明する A1-2 が回答候補として出てきた場合、「否定的な事象の理由は否定的な事象であることが多い」という意味的極

性に関する観察から、A1-1 が正しい回答候補となる可能性が高い。

更に、why 型質問が含む単語とその回答が含む単語間の意味的な相関関係を用いて性能向上を図っている。これは、上記例では、質問文における「病気」の原因を求める場合、回答には「有害物質」、「ウィルス」、「身体の部位」などを表わす単語を含む場合が多い。このように質問文中の単語と回答文中の単語の相関関係を把握し、回答抽出へ応用する。このために、単語クラスタリング手法を用いて、大規模なウェブ文書から単語の意味的クラス（意味的に類似する単語の集合）を自動獲得し、機械学習の素性として活用している。

なお、上記の質問回答例において、図 2-37 に示すように、A1-2 が肯定的な表現と否定的な表現を持つ場合は、意味的極性のパターンの効果が期待できない。

> がんに良いとされる食品を食べ過ぎるのはガンの予防に効果的ではないが、健康的な体重を維持することはガンのリスクを下げる。

図 2-37: 肯定的表現と否定的表現の混在例 [35]

このような問題を解決するため、why 型質問と回答候補における評価表現の極性と共に評価表現の内容（評価表現を構成している単語、係り受け関係など）も考慮した。この際のデータの過疎性を回避するために、評価表現の内容を単語と単語の意味的クラスの両方で表現している。

回答検索では、Murata らの手法 [36]を実装し、回答候補の抽出を行った。質問 q に対して抽出された回答候補 ac は式(2-9)(2-10)によってランク付けが為され、上位 20 個の回答候補を次の段階である回答のリランキングへの入力とする。

$$\mathrm{S(q,ac)} = max_{t_1 \in T} \sum_{t_2 \in T} \varphi \times \log(\mathrm{ts}(t_1, t_2)) \tag{2-9}$$

$$\mathrm{ts}(t_1, t_2) = \frac{N}{2 \times dist(t_1, t_2) \times df(t_2)} \tag{2-10}$$

ここで、T は q と ac の両方に現れる単語（名詞、動詞、形容詞など）の集合を表わす。N は文書の数、$\mathrm{dist}(t_1, t_2)$は ac においてのt_1 とt_2 間の距離、df(t)は t が現れる文書の頻度、$\varphi \in \{0,1\}$は$\mathrm{ts}(t_1, t_2) > 1$であるか否かを示す indicator である（$\mathrm{ts}(t_1, t_2) > 1$であれば 1, それ以外は 0）。次に回答のリランキングを行う。S(q,ac)の上位の回答候補について、それが回答に該当するかを教師あり学習によって作られた分類器（SVM）で判定し、そのスコア(判定の信頼度)によってリランキングする。

人手で作成した 362 個の why 型質問と Web から抽出された回答候補（1 つの質問に

つき 20 個）で評価実験を行った結果、意味的極性の利用によって約 11%の性能改善が確認された [35]。

2.6.4. 回答文と周辺文との関係利用

車らは、質問文と回答文との関係だけでなく、回答文とその周辺文との関係を利用することにより、回答抽出の性能改善を図った [37]。why 型質問の回答として適切な説明文は、周辺の文で補足して説明されていたり、回答候補文書中で重要な文となっていたりすることに着目し、文書の重要文抽出の手法としてよく用いられる Personalized PageRank [38]を用いて回答候補文をノードとした回答候補のランキングを行った。

Personalized PageRank は、PageRank の Random Suffer Model に優先すべきノードにジャンプするテレポーテーション確率を導入したものである。ノードXの PageRank スコアP(X) は、式(2-11)を反復的に処理することにより求められる。

$$\mathrm{P(X)} = (1-\mathrm{d}) * \mathrm{V(X)} + \mathrm{d} * \sum_{Y \in G} w(Y,X) * P(Y) \qquad (2\text{-}11)$$

ここで、G はノードの集合、V(X)はノードの重み、w(Y,X)はYからX方向のエッジの重みを表わす。dはテレポーテーション確率を表わすパラメータであり、実験では最も性能が向上した値$\mathrm{d} = 0.05$が用いられた。

ノードの重みとして質問文との類似度が用いられた。2 文間X,Yの類似度の測定には式(2-12)を用いた。

$$\mathrm{sim(X,Y)} = \frac{1}{|X|} \sum_{x \in X} max\{dice(x,y), y \in Y\} \qquad (2\text{-}12)$$

ここで、x,yは文X,Yの内容語（名詞、動詞、形容詞）、dice(x,y)は 2 単語間x,yのdice係数を表わす。dice係数は、高度言語情報融合フォーラム（ALAGIN）で公開されている単語共起頻度データベース（Version 1.1）を利用し、式(2-13)で算出した。

$$\mathrm{dice(x,y)} = \begin{cases} 1 & (\mathrm{x} = \mathrm{y}) \\ \dfrac{2 * cnt(x,y)}{cnt(x) + cnt(y)} & (\text{otherwise}) \end{cases} \qquad (2\text{-}13)$$

ここで、cnt(x),$cnt(y)$は ALAGIN コーパスの文書集合中での単語x,yの頻度、cnt(x,y)は単語x,yが 4 単語以内で共起した回数を表わす。次に、同じ回答候補文書内の文同士でエッジを張る。エッジの重みは式(2-12)の文同士の類似度とした。

評価実験では、Personalized PageRank の効果を調べるため、既存の手法として渋沢ら [5]の手法を用い、Personalized PageRank で回答候補のランキングを行うシステムと行わないシステムを比較した。回答候補検索のコーパスとして、Yahoo!知恵袋に投稿された回答文書約 50 万件を用いた。正解データの作成は以下のように行った。まず、コーパス中に正解の回答文が存在し、かつ回答候補が 10 文以上得られる 5 個の why 型

の質問を人手で用意した。次に、各質問文のクエリによりコーパスから抽出した回答候補文書の各文を回答候補とし、各文が質問に対する理由の説明になっているかどうかを判定し、理由になっているものを正解データとした。1 つの質問あたりの平均回答候補数は約 133 件、平均正解数は約 9 件であった。評価実験の結果、既存手法と比較し、MRR は 0.232 から 0.254 に向上した。また、上位 N 個の全体的な精度を表わすための P@N(Precision at the top-N answer)に関しては、P@3 が 0.467 から 0.600 に、P@5 が 0.400 から 0.440 に、P@7 が 0.300 から 0.360 に向上した [37]。

2.7. まとめ

why 型質問応答システムは、情報源の文書から、理由・原因部分と帰結部分を抽出する方向で研究が進められてきた。文章抽出の方法として、手掛かり語などによるルールベース型の手法が用いられた [1] [5]。更に、自動で網羅的にルールを生成するために抽象化された語彙のルールを機械学習で獲得する手法が研究された [12]。

意味解析ネットワークによる手法 [21]や Web 検索エンジンを利用する手法 [4]、質問タイプに依らずに包括的に文章抽出を行う手法 [28]なども研究されている。

要素技術としては、why テキストセグメントの識別 [34]、意味的極性の利用 [35]、回答文と周辺文との関係利用 [37]などの研究がある。

これまでの研究は主として、手掛かり語や理由・原因の表現パターンなどの表層的な文章情報を対象にしたものであるため、言語表現の多様性や因果関係の連鎖性への対応は難しい [5]。意味解析ネットワークは、語彙間の意味という深層的な情報に踏み込んだ手法であるが、現状の意味解析は、precision は高いが recall が低くなることが指摘されている [2]。今後、why 型質問応答システムの性能向上のためには、手掛かり語などによる文章抽出と意味解析の組み合わせ、更に文章全体の表わす意味を理解する技術の研究が必要であると考える。

第3章　含意関係認識

3.1. 含意関係認識による why 型質問応答システムの改善

一般に、why 型質問応答システムでは、図 3-1 に示すように、検索対象文書から理由文と事実文を抽出する。

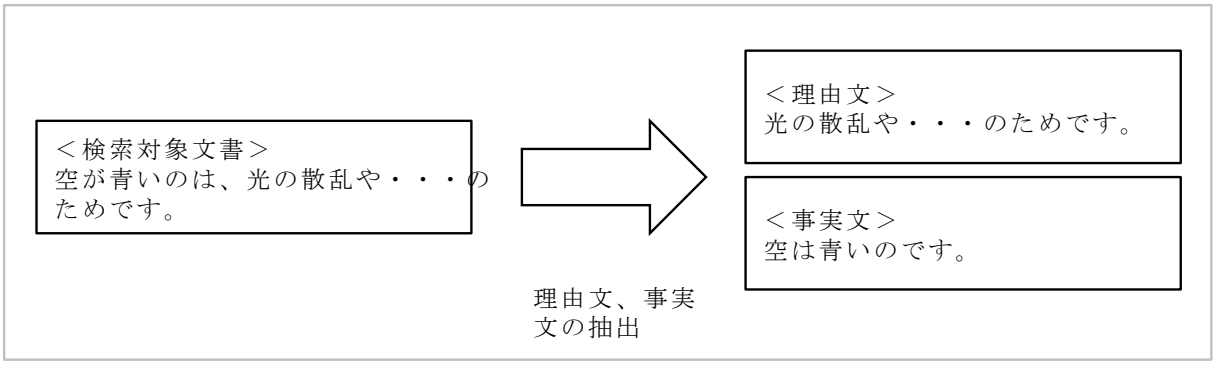

図 3-1: 検索対象文書から理由文・事実文の抽出

しかし、ほぼ同じ内容を述べている説明文であっても多様な表現があり得る。例えば、図 3-2 は、同じ内容が異なる表現で表わされている事実文が抽出された例を示している。

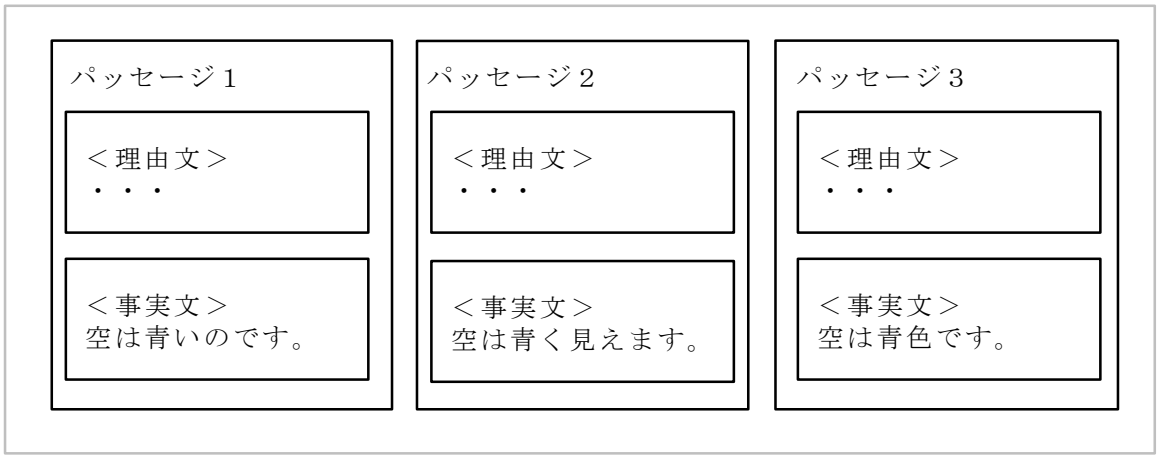

図 3-2: 説明文の多様な表現例（事実文を例として）

図 3-2 に示した事実文が「空は青い」という内容を表わすことを認識できれば、これらの事実文に対応する理由文を元の質問の回答として抽出できるようになる。したがって、上記のような文章表現の多様性に対応することによって、質問応答システムの精度向上が図れると考えられる。

文章表現の多様性や言い換え、内容の含意に関する研究は、含意関係認識（RTE; Recognizing Textual Entailment）と呼ばれる。含意関係認識は、質問応答システムや文書要約、機械翻訳等、広い応用が期待されており、RTE タスクとして精力的に研究さ

れている。本課題研究では、質問応答システムへの応用の観点から、含意関係認識を実現するためにどのような推論手法が研究されているか調査を行った。

含意関係認識を質問応答システムへ応用する際に、質問文と回答候補の含意関係を評価する場合がある。その際に、質問文には疑問詞（英語であれば、what, where, when, who 等の wh 型の語）が含まれているため、疑問詞の部分を変数として質問文を肯定文に変換する手法がある。英語の場合について表 3-1 に例を示す。

表 3-1: 質問文の肯定文への変換

No	質問文	変換後の肯定文	出典
1	What is the capital of France?	[X] is the capital of France	文献 [39]
2	Who painted 'The Scream' ?	[ANSWER] painted 'The Scream'.	文献 [40]

このような変換を行うことにより、質問文も処理対象とした含意関係認識を行うことができる。

次に、why 型質問応答システムに含意関係認識モジュールを組み込む方法について考察する。まず、上述のように why 型質問文を肯定文に変換する。そして、質問文に対して得られた回答候補群の事実文に対して、肯定文との含意関係認識を行い、両者に含意関係が成立する度合(含意関係認識の信頼度)に基づいて再ランキングを行い、最終的な回答を得る。この処理の流れを図 3-3 に示す。

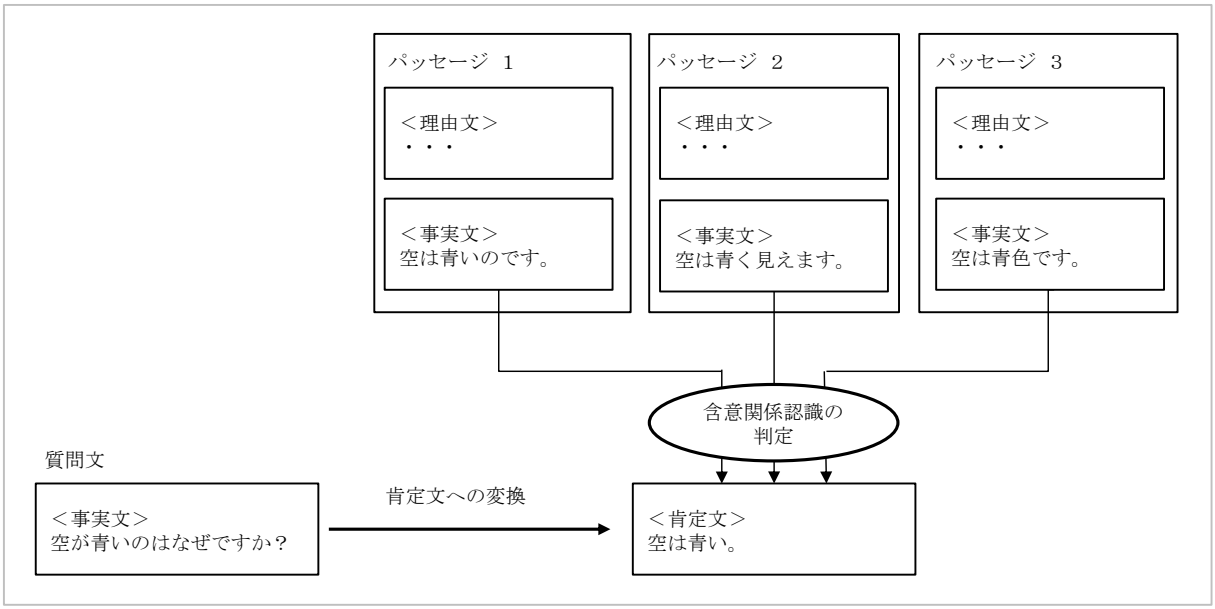

図 3-3: why 型質問応答システムへの含意関係認識の組込み

3.2. 含意関係認識の質問応答システムへの応用

含意関係認識を質問応答システムに応用する例として、Harabagiu らは含意関係認識モジュールを質問応答システムに組み込み、質問・回答の再ランキングを行う方法を提案している[41]。図 3-4 は Harabagiu らの質問応答システムの処理の流れを示している。

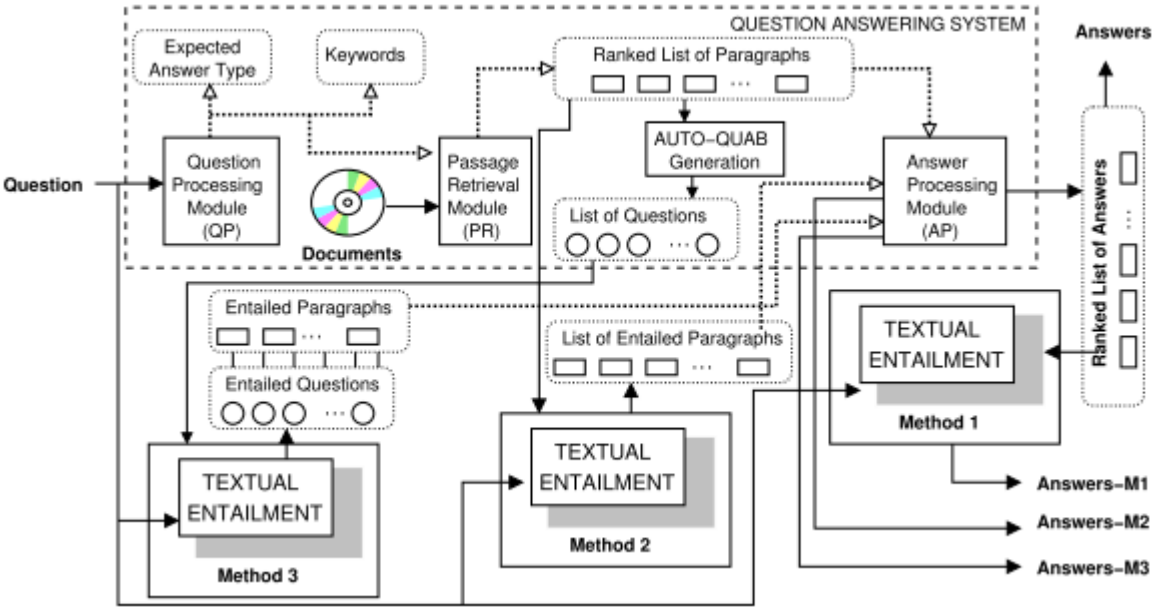

図 3-4: 含意関係認識の質問応答システムへの組み込み [41]

図 3-4 に示すように、一般的な質問応答システムとして以下のモジュールを含む。

(1) 質問生成モジュール（QP）

(2) パッセージ取得モジュール（PR）

(3) 回答生成モジュール（AP）

これらのモジュールから成る一般的な質問応答システムに、含意関係認識モジュールを付加して組み込む構成となっている。含意関係認識を適用する手法として、方法 1（Method 1)、方法 2（Method 2)、方法 3（Method 3）が提案されている。

（1）方法 1（Method 1）

AP モジュールが出力した回答候補群に対し、質問文との含意関係認識を行い、回答候補の再ランキングを行う。質問文の肯定文への変換は前述（表 3-1）の方法で行う。

表 3-2: 方法 1 による回答の再ランキング [41]

Q_1: "What did Peter Minuit buy for the equivalent of $24.00?"				
	$Rank_1$	TE	$Rank_2$	Answer Text
A1	6th	YES (0.89)	1st	Everyone knows that , back in 1626, Peter Minuit bought **Manhattan** from the Indians for $24 worth of trinkets.
A2	1st	NO (0.81)	-	in 1626, an enterprising Peter Minuit flagged down some passing locals, plied them with **breads, cloth and trinkets** worth an estimated $24, and walked away with the whole island.

表 3-2 において、$Rank_1$ は方法 1 を組み込む前の回答候補のランキング、TE は方法 1 による質問文と回答候補文との含意関係認識の結果、$Rank_2$ はその含意関係認識の結果による再ランキングを表わす。ここでは、再ランキングにより、正解の回答候補 A1 のランキングが 6 位から 1 位に上がったことを示している。

（2）方法 2（Method 2）

PR モジュールが出力するパッセージ群に対し、質問文との含意関係認識を行い、含意関係が成立しないパッセージを除外する。このように候補となるパッセージ群の絞り込みを行うことにより、処理対象とするパッセージの量を制限することができる。

（3）方法 3（Method 3）

PR モジュールが出力するパッセージ群から、パッセージに含まれる質問文相当の文章を自動生成し、それら自動生成質問文（AGQ: automatically-generated questions）と元の質問文との含意関係認識を行う。パッセージから自動生成された質問文の例を表 3-3 に示す。

表 3-3: 自動生成質問文（AGQs）と質問文との含意関係認識 [41]

Q_2: "How hot does the inside of an active volcano get?"	
A_2	Tamagawa University volcano expert Takeyo Kosaka said lava fragments belched out of the mountain on January 31 were as hot as **300 degrees Fahrenheit**. The intense heat from a second eruption on Tuesday forced rescue operations to stop after 90 minutes. Because of the high temperatures, the bodies of only five of the volcano's initial victims were retrieved.
Positive Entailment	
AGQ_1	What temperature were the lava fragments belched out of the mountain on January 31?
AGQ_2	How many degrees Fahrenheit were the lava fragments belched out of the mountain on January 31?
Negative Entailment	
AGQ_3	When did rescue operations have to stop?
AGQ_4	How many bodies of the volcano's initial victims were retrieved?

以上の方法１～３によって含意関係認識の機能を質問応答システムに付加することにより、why 型質問応答システムの性能向上が実現できる。

Harabagiu らは、TREC Q/A 評価で用いられた質問セットからランダムに選んだ 500 個のファクトイド型質問文を用いて、含意関係認識機能の効果を評価した。質問文 500 の内、335 文（67.0%）は回答タイプが付加されており、残りの 165 文（33.0%）は回答タイプが不明なものである。前述の方法１～３，及びそれら方法の組み合わせによるハイブリッド方法による実験結果を表 3-4 に示す。含意関係認識機能の追加によって、正解率及び MRR が向上することが確認された。Harabagiu らの実験ではファクトイド型質問文が用いられたが、ノンファクトイド型質問文に対しても同様な効果が期待できる。

表 3-4: 質問応答システムへの含意認識機能の追加効果 [41]

	回答タイプあり		回答タイプ不明	
	正解率	MRR	正解率	MRR
ベースライン	32.0%	0.3001	30.6%	0.2978
方法 1	44.1%	0.4114	39.5%	0.3833
方法 2	52.4%	0.5558	42.7%	0.4135
方法 3	41.5%	0.4257	37.5%	0.3575
ハイブリッド方法	53.9%	0.5640	41.9%	0.4010

3.3. 含意関係認識の手法

含意関係認識の研究は近年精力的に行われている。含意関係認識の評価型ワークショップも実施されている。PASCAL RTE-1～RTE7 [42] [43] [44] [45] [46] [47] [48]は英語を対象とした含意関係認識のタスクであり [40]、日本語に関しても RITE1, RITE2 [49] [50] のタスクが行われてきた。宮尾らは、大学入試センター試験から正誤を問う設問を抽出し、含意関係認識の評価データを開発し、RITE タスクに提供した[51]。また、国立情報学研究所が 2011 年度から推進している「人工頭脳プロジェクト ―ロボットは東大に入れるか」 [52]でも含意関係認識の研究が行われている。本節では、含意関係認識の過去の主な研究を概観する。

3.3.1. 類似度をベースとした手法

Adams は語句類似度によるアプローチをとり、Text と Hypothesis 間の語句の重なりを測定し、重なりが大きいときに含意関係が成立すると判定する手法を提案している [53]。ここで Text と Hypothesis は含意関係を認識する 2 つのテキストを表わし、Text が真のときに Hypothesis が真となるか(Text が Hypothesis の意味を包摂するか)を判定することが含意関係認識の一般的な問題設定となっている。

Mehdad ら [54]は、 編集距離モデルに基づいた EDITS(Edit Distance Textual Entailment Suite)というオープンソースの RTE ソフトウェアを開発した。編集距離（Edit Distance）モデルでは、Text と Hypothesis の編集距離を計算し、その距離が十分小さいときに含意関係が成立すると判定する。EDITS では、1）文字単位の編集距離、2）トークン単位の編集距離、3）構文木単位の編集距離（構文木のノード単位の編集距離）の 3 レベルの編集距離を用いている。そして、求められた編集距離ED(T, H)から、式(3-1)により entailment スコアを計算する。

$$\text{entailment(T,H)} = \frac{ED(T,H)}{\left(ED(T,\_) + ED(\_,H)\right)} \tag{3-1}$$

ここで、$\text{ED}(T,H)$は T と H 間の編集距離を表わし、$\text{ED}(T,_)$は T の全テキストを削除するコスト、$\text{ED}(_,H)$は H の全テキストを挿入するコストに相当する。entailment スコアは 0 から 1 の範囲の値となり、T と H が同一であるときは 0 に、T と H が全く異なる場合は 1 になる。この entatilment スコアを用いて学習を行い、含意関係の成立を判定するための最適なスコア閾値を求めた。

EDITS の処理フローを図 3-5 に示す。なお、図 3-5 中の ETAF (EDITS Text Annotation Format)は EDITS 内で使われる言語情報の内部表現を表している。

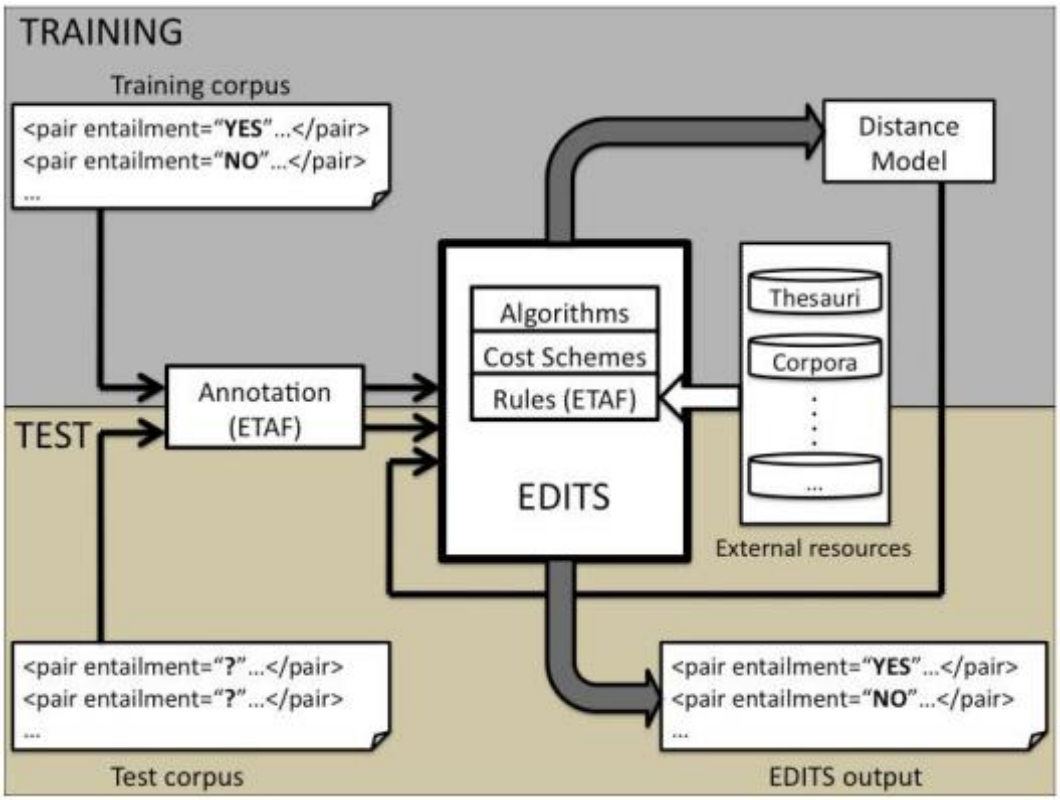

図 3-5: EDITS の処理フロー [54]

Heilman らは、編集距離モデルにおいて、編集操作は局所的に限定するという通常の仮定を置かないことにより、統語構造のより柔軟な変更を可能にしている[55]。編集操作として、INSERT-CHILD, INSERT-PARENT, DELETE-LEAF, DELETE-&-MERGE, RELABEL-NODE, RELABEL-EDGE, MOVE-SUBTREE, NEW-ROOT, MOVE-SIBLING を使用している。図 3-6 は Hypothesis の依存木から Text の依存木を得る際の編集過程の例を表わしている。

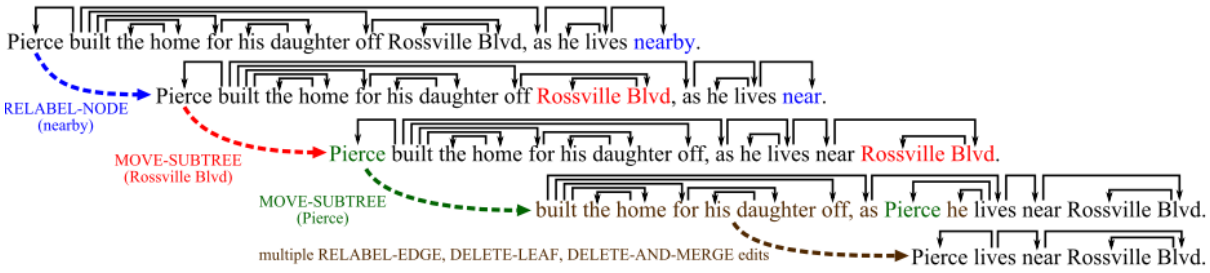

図 3-6: 編集操作による含意関係認識の例 [55]

3.3.2. アラインメントをベースとした手法

単語や文節等のアラインメントによって Text と Hypothesis の文間関係認識を行い、含意関係認識に適用するアプローチがある。文間関係認識とは、同意、矛盾、限定、根拠などを含む、文間のより多様な関係を同定する技術である。アラインメントによる文間関係認識の大まかな流れは以下の通りである [56]。

1. 解析 形態素解析、構文解析といった基本的な解析
2. アラインメント 文間で対応する単語間の対応付け
3. 関係分類 これらの結果から文間の関係を判別

De Marneffe らは、Text と Hypothesis のテキストを意味グラフに変換し、意味グラフ間のアラインメントを行った[72]。例えば、次の Text, Hypothesis の例に対しては、Hypothesis は図 3-7 のような意味グラフに変換され、図 3-8 のようにアラインメントを行ってアラインメント・スコアを計算している。

T: Mitsubishi Motors Corp.'s new vehicle sales in the US fell 46 percent in June.
H: Mitsubishi sales rose 46 percent. (FALSE)

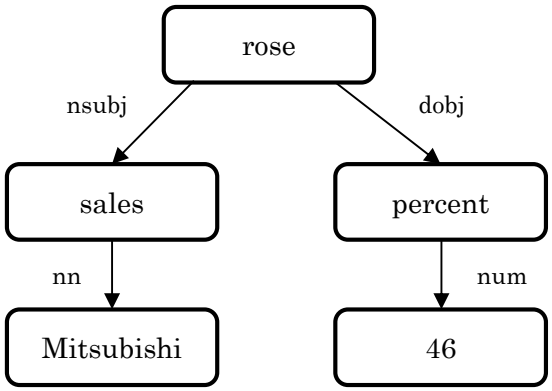

図 3-7: 意味グラフの例 [72]

```
rose                →   fell
sales               →   sales
Mitsubishi          →   Mitsubishi_Motors_Corp.
percent             →   percent
46  →    46

Alignment score: -0.8962
```

図 3-8: アラインメント結果 [72]

De Marneffe らは、式(3-2)(3-3)のようにアラインメント・スコアを定義し、このスコアが最大となるアラインメントを探索した。

$$\mathrm{s(D)} = \sum_{(t,h,a)\in D} s(t,h,a) \tag{3-2}$$

$$s(t,h,a) = \sum_{i\in h} s_w\big(h_i, a(i)\big) + \sum_{(i,j)\in e(h)} s_e\Big(\big(h_i,h_j\big),\big(a(h_i),a\big(h_j\big)\big)\Big) \tag{3-3}$$

ここで、$\mathrm{a}(x)$はアラインメントaにおいて Hypothesis 中の単語xに対応付けられる Text 中の単語を表わす。$\mathrm{e}(x)$は Hypothesis x 中のエッジの集合を返す関数を表わす。式(3-3)において、最初の項s_wは各単語のアラインメント・スコアの合計に相当し、2 番めの項s_eは Hypothesis グラフ中のエッジとそれに対応する Text 中のエッジとのアラインメント・スコアの合計に相当する。なお、De Marneffe らは、アラインメントの処理を経て、3.3.7 に述べる言語的特徴の分析を行い、最終的な含意関係の判定を行っている。

Sammons ら [57]は、 図 3-9 に示すように、テキストに対して、固有表現や共参照の抽出、意味役割付与を行って、Text と Hypothesis のアラインメントに利用している。

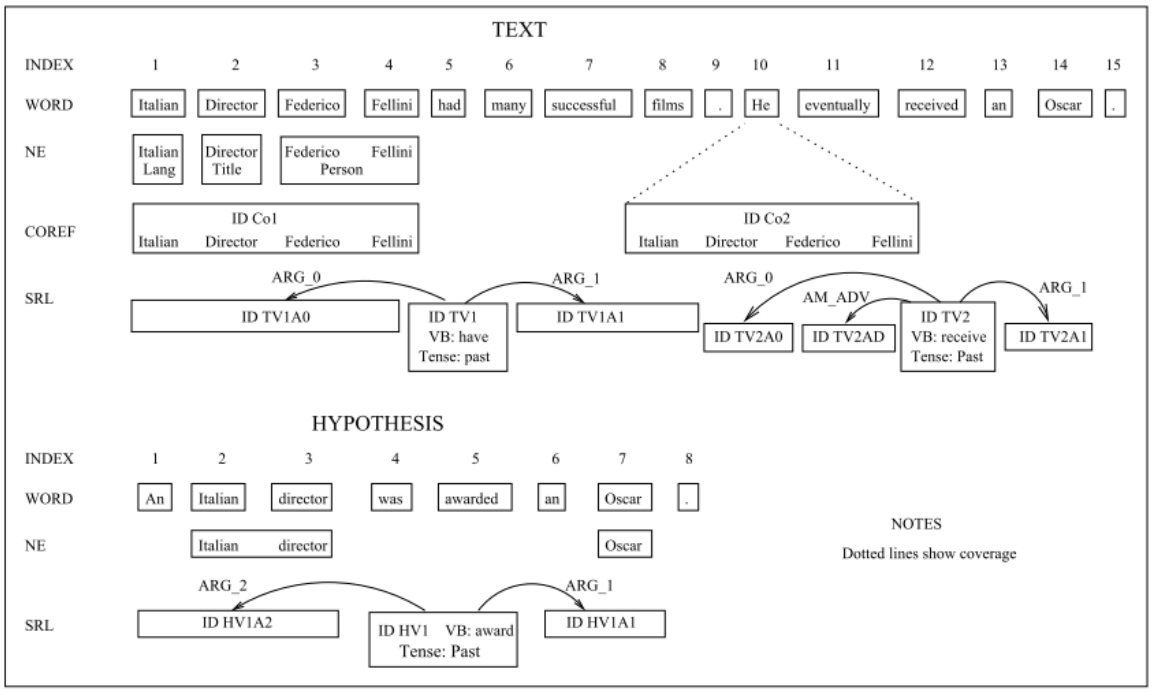

図 3-9: テキストに対する固有表現・共参照の抽出及び意味役割付与 [57]

図 3-9 において、NE は固有表現（named entity）、COREF は共参照（co-reference）、SRL は意味役割付与（semantic role labeling）を指す。Sammons らは、これらの言語情報を用いてトークン間の対応付けをスコアリングし、アライメントを行った。テキストには多様な言語情報が含まれるため、各言語情報に対する重み付けを機械学習により最適化した。

後藤ら [56]は、文間関係認識におけるアラインメントについて、単語同士のアラインメントに加えて、単語間の依存関係の対応付けをとる処理をアラインメントの中で行う方法を提案している。単語間の依存関係の対応付けを「局所構造アラインメント」と呼ぶ。局所構造アラインメントは、図 3-10 において、HYP(Hypothesis)における w_i,w_j はそれぞれ TEXT における $A(w_i)$,$A(w_j)$に単語アラインメントされているとした場合、w_i,w_j 間の意味的関係が $A(w_i)$,$A(w_j)$間でも成り立っているかどうかを判別する問題となる。

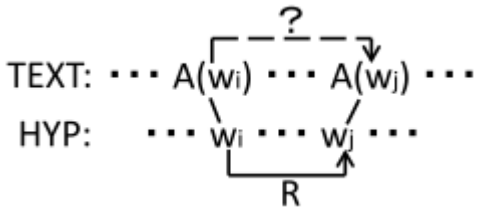

図 3-10: 局所構造アラインメント [56]

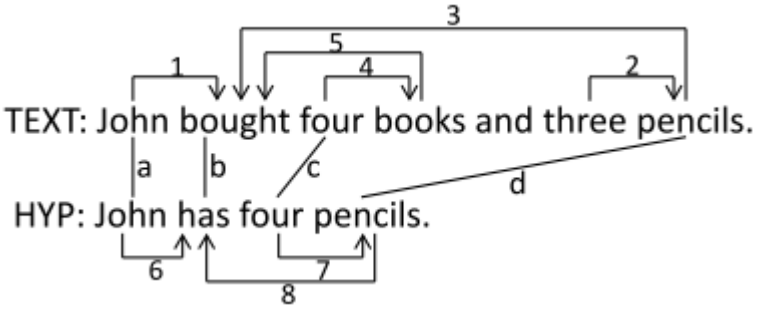

図 3-11: 局所構造アラインメントの例[56]

図 3-11 を例に局所構造アラインメントの処理を説明する。

- HYP 中の依存構造 6 については、a, b の単語アラインメントがあり、TEXT においてアラインメント先の 2 つの単語(John と bought)には依存構造 1 が存在しているので、1 と 6 の間に局所構造アラインメントをとる。
- HYP 中の依存構造 7 については、c, d の単語アラインメントがあるが、TEXT においてはアラインメントされた 2 つの単語の間に依存構造はない。従って、7 に対して局所構造アラインメントはとられない。
- HYP 中の依存構造 8 については、b, d の単語アラインメントがあり、TEXT においてアラインメント先の 2 つの単語には依存構造 3 が成立しているので、8 と 3 の間に局所構造アラインメントをとる。

そして、全体を単語アラインメントと局所構造アラインメントの 2 つの段階に分け、それぞれをルールベースでプロトタイピングし、得られたシステムで事例を収集、課題の性質の分析などを行い、どの部分を統計的学習で最適化するのが効率的であるかを検討することを提案している。

後藤らは提案する構造的アラインメント手法について、評価実験を行った。関係認識を行う文対は、クエリに対する検索結果のテキストとし、表 3-5 に示す 7 種類のクエリから検索で得られた文集合に対して提案手法を適用し、人手で正解判定を行った。また、正解データを人手で各クエリに対して 20 文ずつ作成した。

表 3-5: 検索質問と検索数 [56]

検索質問（下線部が検索エンジンに与えたクエリ）	検索数
キシリトールは虫歯予防に効果がある	95
イソフラボンは健康に効果がある	523
ステロイドは副作用がある	81
還元水は健康を守る	118
バイオエタノールは環境に良い	172
クローン技術は規制が必要だ	175
CO_2 は温暖化の原因である	905

上述のコーパスに対して提案手法を適用した評価結果について、その精度と再現率を表 3-6 に示す。

表 3-6: 局所構造アラインメントの実験結果 [56]

関係	同意	矛盾	不明
精度	54.38%	52.08%	47.08%
再現率	51.77%	51.04%	44.04%

De Marneffe らは、アラインメントは、Text の比較的少量の部分が Hypothesis にとって重要であることを示した[61]。その結果に基づくと、アラインメントは、その重要部分を特定し、推論ステップを単純化することが目標となる。

3.3.3. 統語構造変換による手法

Braz らは、統語構造の階層的グラフ表現を元に、グラフの包摂（Subsumption）が成立するかという問題を解くことで含意関係認識を実現する手法を提案した[58][59]。テキスト変換ルールを用意し､含意関係認識対象のテキストを統語構造グラフに変換し、Text の統語構造グラフが Hypothesis の統語構造グラフを包摂するかを判断する。
Braz らの例を以下に示す [59]。例では、S が Text に、T が Hypothesis に相当する。

S: “The New York Times reported that Hanssen sold FBI secrets to the Russians and could face the death penalty.”
T: “Hanssen sold FBI secrets to the Russians.”

S, T を構文解析して、次の S-A, S-B 及び T-A を生成する。S において階層構造で入れ子になっている“Hanssen sold FBI secrets to the Russians”の部分が抽出されて S-B が生成され、T-A とマッチングし、T が S に包摂されていると判定することができる。

S-A:“[The New York Times]/A0 reported [that Hanssen sold FBI secrets to the Russians...]/A1”
S-B: “[Hanssen]/A0 sold [FBI secrets]/A1 to [the Russians]/A3”
T-A: “[Hanssen]/A0 sold [FBI secrets]/A1 to [the Russians]/A3”

Braz らは、更に言語の多様な表現に対応するために、WordNet [75]や名詞句の包摂ルール（NP-subsumption rule)を利用して、テキスト変換を行った。WordNet 利用の例として、“destroy”と“raze”、“build”と“construct”のような同義語の関係や、肺癌（lung cancer）に対する癌（carcinoma）のような上位語の関係の認識などがある。名詞句の包摂ルールとして、例えば“Jazz singer Marion Montgomery”と“singer”が IS-A 関係にあることの認識などがある。多様な表現に対する包摂及び含意関係認識の例を以下に示す。

S: Lung cancer put an end to the life of Jazz singer Marion Montgomery on Monday.
T: The singer died of carcinoma.

Braz らの方法では、S を変換して、次の S1, S2 のような複数のテキストを生成する。

S1: Lung cancer killed Jazz singer Marion Montgomery on Monday.
S2: Jazz singer Marion Montgomery died of lung cancer on Monday.
…
そして、これらの変換されたテキストに T が包摂されるかを判定していく。この例の場合、“X put end to Y’s life → Y die of X”というルールから S2 を生成し、“Jazz singer Marion Montgomery”と“singer”が IS-A 関係にあるというルールから、S2 が T を包摂することを認識することにより、結果的に、S が T を含意する、と判定することができた。Braz らは、多くの表現を生成して包摂を判定するアルゴリズムを整数計画問題として定式化し、コスト最小になる包摂を探索するようにした。

Bar-Haim らは、構文構造の変換ルールを用いることによる含意関係認識のアプローチをとっている[60]。

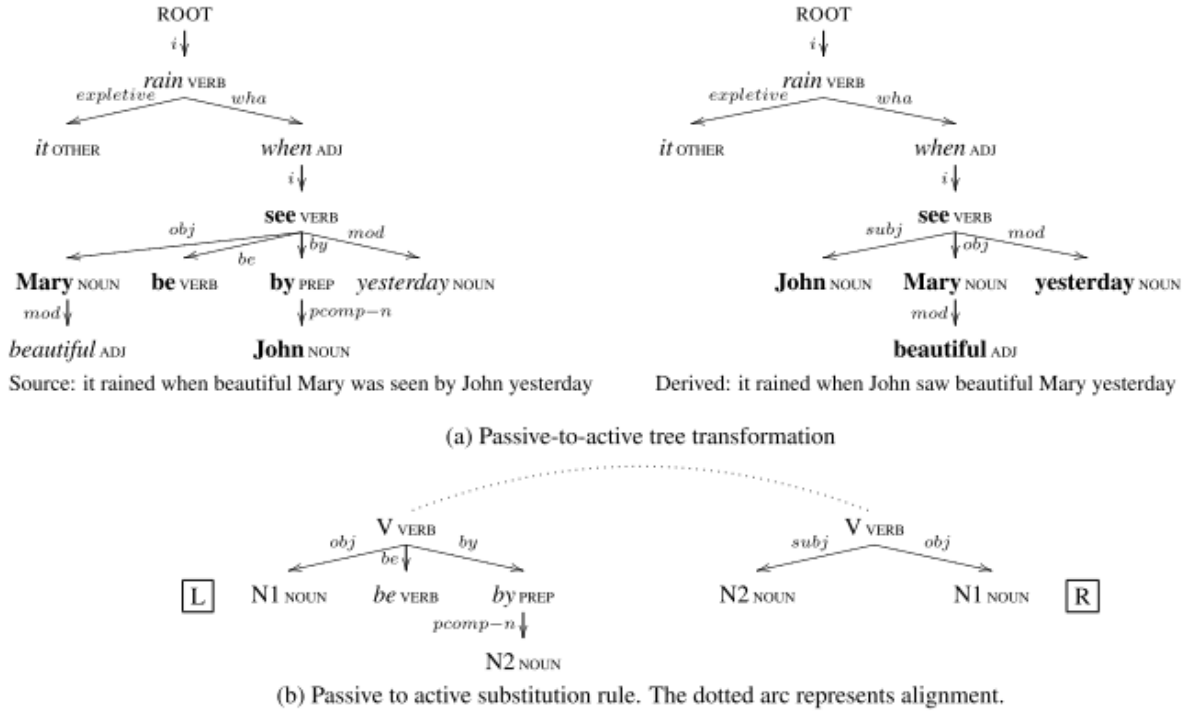

図 3-12: (a)変換例及び(b)変換ルール [60]

図 3-12 は受動形と能動形の変換例を示している。図 3-12 (b)に示す変換ルールのテンプレートをあらかじめ用意しておき、図 3-12 (a)の構文木へのマッチングを行う。マッチングが成功したら、各ノードを具体的な単語へ置き換えるが、例に示す“yesterday ”のようにテンプレートに含まれていない単語は、変換元の構文木構造に従って、変換先の構文木に追加する。この例では受動態の文を能動態の文に正規化することにより含意関係認識の精度を向上させることを狙っている。Bar-Haim らがカテゴリ分類した変換ルールを表 3-7 に示す。

表 3-7: カテゴリ分類した変換ルール [60]

#	Category	Example: source	Example: derived
1	Conjunctions	Helena's very experienced and has played a long time on the tour.	⇒ Helena has played a long time on the tour.
2	Clausal modifiers	But celebrations were muted as many Iranians observed a Shi'ite mourning month.	⇒ Many Iranians observed a Shi'ite mourning month.
3	Relative clauses	The assailants fired six bullets at the car, which carried Vladimir Skobtsov.	⇒ The car carried Vladimir Skobtsov.
4	Appositives	Frank Robinson, a one-time manager of the Indians, has the distinction for the NL.	⇒ Frank Robinson is a one-time manager of the Indians.
5	Determiners	The plaintiffs filed **their** lawsuit last year in U.S. District Court in Miami.	⇒ The plaintiffs filed **a** lawsuit last year in U.S. District Court in Miami.
6	Passive	We have been approached by the investment banker.	⇒ The investment banker approached us.
7	Genitive modifier	Malaysia's crude palm oil output is estimated to have risen by up to six percent.	⇒ The crude palm oil output of Malasia is estimated to have risen by up to six percent.
8	Polarity	Yadav was forced to resign.	⇒ Yadav resigned.
9	Negation, modality	What we've **never** seen is actual costs come down.	What we've never $\overline{\text{seen}}$ is actual costs come down. (⇏ What we've seen is actual costs come down.)

3.3.4. 形式論理をベースとした手法

Akhmatova ら [64]、Bos ら [65]、Raina ら [66]による含意関係認識手法は定理証明系を基礎としている。自然言語形態の含意関係認識の対象とする文のペアから、宣言や依存構造を抽出し、それぞれの文を論理式に変換する。この際、言語的な分解や構文構造の等価性、WordNet [75]から抽出した語句表現のルールなどを利用している。そして、Text の論理式から Hypotheis の論理式を導き出すことができれば、含意関係が成立と判定する。

定理証明に基づく手法の例として Raina らの手法 [66]を紹介する。Raina らの手法では、例えば“Bob purchased an old convertible. ”というテキストを図 3-13 に示す統語的依存木（syntactic dependencies）に変換する。

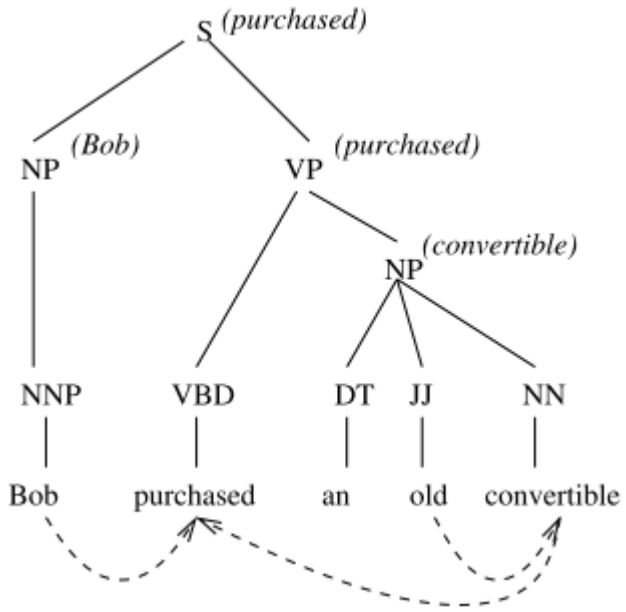

図 3-13: 統語的依存木 [66]

次に、統語的依存木の係り受け構造を式(3-4)に示す Horn 節に変換する。

(∃A, B, C) Bob(A) ∧ convertible(B) ∧ old(B) ∧ purchased(C, A, B) (3-4)

そして、論理式の導出原理を用いて、Text から Hypothesis を導けるかを判定するが、含意関係認識では、通常、Text 以外に暗黙の仮定文を付与することが必要であり、この仮定文の選択のためにコスト関数を導入している。仮定文を A とすると、コスト関数 $C_w(A)$は式(3-5)で表わされる。

$$C_w(A) = \sum_{d=1}^{D} w_d f_d(A) \quad (3\text{-}5)$$

ここで、$f_1, \ldots f_D$ は、WordNet [75]などの知識リソースから得られる A の特徴を示す特徴関数であり、$w_1, \ldots w_D$ は各特徴関数の重みを表わす。特徴関数を重み付けして総和を

とり、A のコスト関数 $C_w(A)$が得られる。Raina らは、各特徴関数の重み付け$w_1, \ldots w_D$ を Text と Hypothesis の組から成るデータセットから機械学習することにより決定している。仮定文を付与することは、結論（ここでは Hypothesis）を導くために仮説を推し量ることに相当するため、論理学におけるアブダクションに相当すると考えられる。従って、Raina らの方法は、アブダクションによる論理推論と機械学習を組み合わせた手法である。

Moldovan, Tatu らが開発したシステム COGEX[62][63]では、Text と Hypothesis を論理式表現に変換し、Text の論理式、及び Hypothesis の論理式の否定形、及び背景知識として WordNet などから構築した公理から成る集合に対して、否定を証明する推論を探索する。否定が証明されれば、Hypothesis が否定形で入っているため、Hypothesis が成立する、すなわち、含意関係が認識されたことになる。否定を証明する推論が見つからない場合は、Hypothesis 中の述語項の引数を削除して、再び探索を行う。引数を削除しても探索が失敗する場合は、その述語項を削除していく、という処理を繰返し行う。そして、否定が証明された時点で、Hypothesis から削除した部分のスコアリングを行い、含意関係の判定の確からしさを[0,1]の範囲の値で算出する。

田らは、Dependency-base Compositional Semantics(DCS) [68]を基盤とし、外延の抽象表現を定義することで、データベースを明示的に与えずに、自然言語テキスト間の推論関係を計算できることを示した[67]。DCS は、図 3-14 のように、「学生が本を読む」という文は係り受け木の形で表わされ、木の各辺には SBJ や OBJ のような意味役割が付与される。DCS 木が定義されると、あらかじめ用意された図 3-15 のようなデータベースに対し、例えば「読む」の表の中から SBJ が「学生」、OBJ が「本」に入るレコード（「花子が平家物語を読む」など）を取り出し、その結果、「学生が本を読む」が表わすレコードの集合（外延）が得られる。

図 3-14: DCS 木の例 [67]

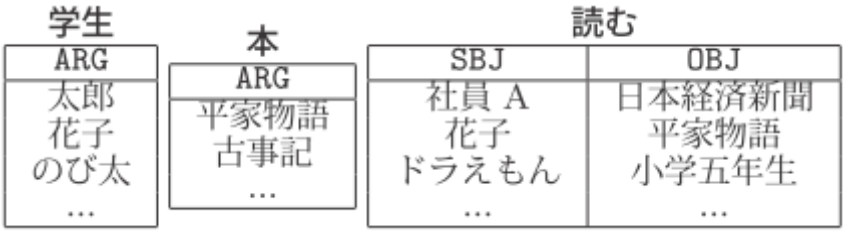

図 3-15: DCS 木のデータベース表現 [67]

しかし、DCS を一般の含意関係認識に適用する場合には、具体的な DCS 木のデータベースが必要となる。田らは、関係代数の演算子を用いて、データベースの検索命令を抽象化した「外延の抽象表現」を定義することにより、DCS 木の意味を具体的なデータベースに依存しない形で表現する手法を提案した。外延の抽象表現の例を図 3-16 に示す。

	自然言語表現	外延の抽象表現
複合名詞	*pet fish*	$\mathbf{pet} \cap \mathbf{fish}$
修飾	*nice day*	$\mathbf{day} \cap (W_{\mathrm{ARG}} \times \mathbf{nice}_{\mathrm{MOD}})$
時間	*boys study at night*	$\mathbf{study} \cap (\mathbf{boy}_{\mathrm{SBJ}} \times \mathbf{night}_{\mathrm{TIME}})$
関係詞節	*books that students read*	$\mathbf{book} \cap \pi_{\mathrm{OBJ}}(\mathbf{read} \cap (\mathbf{student}_{\mathrm{SBJ}} \times W_{\mathrm{OBJ}}))$
量化	*all men die*	$\mathbf{man} \subset \pi_{\mathrm{SBJ}}(\mathbf{die})$
否定	*no dogs are hurt*	$\mathbf{dog} \parallel \pi_{\mathrm{OBJ}}(\mathbf{hurt})$

図 3-16: 外延の抽象表現の例 [67]

外延の抽象表現間の関係として充足可能、包含、矛盾の 3 つを用意し(これを言明と呼ぶ)、また言明間の論理関係を導くための約 30 個の公理を実装した。公理の一部を図 3-17 に示す。この公理系は、全ての公理がホーン節であることが特徴である。これにより、自然言語における多くの推論を可能にし、計算機での高速処理を可能にした。

1. $W \neq \emptyset$	5. $(A \subset B \;\&\; B \subset C) \Rightarrow A \subset C$
2. $A \cap B \subset A$	6. $(A \subset B \;\&\; A \neq \emptyset) \Rightarrow B \neq \emptyset$
3. $B \times q_{\subset}(A, B) \subset A$	7. $(A \parallel B \;\&\; C \subset A) \Rightarrow C \parallel B$
4. $A \neq \emptyset \Leftrightarrow \pi(A) \neq \emptyset$	8. $(C \subset A \;\&\; C \subset B) \Rightarrow C \subset A \cap B$

図 3-17: 公理系（抜粋） [67]

図 3-18 は、左上が「太郎は全ての教科書を読む」、左下が「竹取物語は教科書にある」、右が「太郎は竹取物語のある本を読む」を表わす DCS 木の例であり、この木から図 3-19 のような抽象表現を生成する。

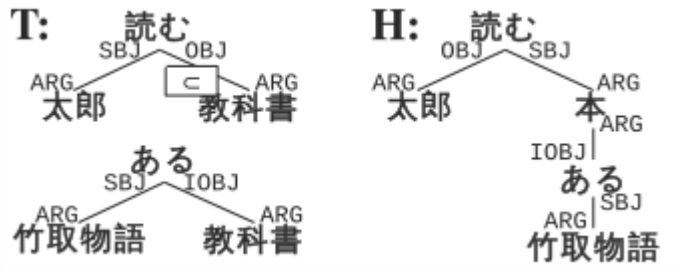

図 3-18: DCS 木の例 [67]

太郎が読むもの：

$$F_1 = \pi_{\texttt{OBJ}}(\textbf{読む} \cap (\textbf{太郎}_{\texttt{SBJ}} \times W_{\texttt{OBJ}}))$$

竹取物語は教科書にある：

$$F_2 = \textbf{ある} \cap (\textbf{竹取物語}_{\texttt{SBJ}} \times \textbf{教科書}_{\texttt{IOBJ}})$$

竹取物語のある教科書：

$$F_3 = \textbf{教科書} \cap \pi_{\texttt{IOBJ}}(\textbf{ある} \cap (\textbf{竹取物語}_{\texttt{SBJ}} \times W_{\texttt{IOBJ}}))$$

竹取物語のある本：

$$F_4 = \textbf{本} \cap \pi_{\texttt{IOBJ}}(\textbf{ある} \cap (\textbf{竹取物語}_{\texttt{SBJ}} \times W_{\texttt{IOBJ}}))$$

太郎は竹取物語のある本を読む：

$$F_5 = \textbf{読む} \cap (\textbf{太郎}_{\texttt{SBJ}} \times F_{4,\texttt{OBJ}})$$

図 3-19: DCS 木から生成した抽象表現 [67]

図 3-18 の T（左側）と H（右側）に対して、図 3-20 に示すように、T の言明（教科書 ⊂ 本）および公理から、H の表わす言明$F_5 \neq \emptyset$が証明される。

$$\dfrac{\pi_{\texttt{OBJ}}(F_5) = F_1 \cap F_4 \qquad \dfrac{\dfrac{\pi_{\texttt{IOBJ}}(F_2) = F_3 \quad \dfrac{\textbf{T}}{F_2 \neq \emptyset}\ \text{公理 4}}{F_3 \neq \emptyset} \qquad \dfrac{\dfrac{\textbf{教科書} \subset \textbf{本} \quad \dfrac{\textbf{T}}{\textbf{教科書} \subset F_1} \quad \dfrac{\text{公理 2}}{F_3 \subset \textbf{教科書}}\ \text{公理 5}}{F_3 \subset F_4} \qquad \dfrac{\dfrac{\textbf{T}}{\textbf{教科書} \subset F_1} \quad \dfrac{\text{公理 2}}{F_3 \subset \textbf{教科書}}\ \text{公理 5}}{F_3 \subset F_1}\ \text{公理 8}}{F_3 \subset F_1 \cap F_4}\ \text{公理 6}}{F_1 \cap F_4 \neq \emptyset}\ \text{公理 4}}{F_5 \neq \emptyset}$$

図 3-20: 外延の抽象表現を用いた証明の例 [67]

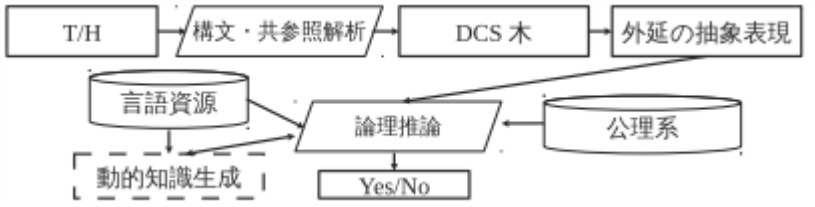

図 3-21: 含意関係認識システム [67]

田らの含意関係認識システム（図 3-21 を参照）は以下のように動作する[67]。

1. 各 T-H ペアに対し係り受け・共参照解析を行う。
2. 解析結果をパターンマッチングにより DCS 木に変換する。DCS 木は抽象表現に関する言明に変換される。
3. T の言明と言語知識を用いて、H の証明を試みる。
4. H を証明できなかった場合、T と H の DCS 木からアラインメントを生成する。類似度が閾値以上のアラインメントを言明に変換し、新たな知識として推論エンジンに追加する。
5. 再度 H の証明を試みて、最終的に H が証明されたペアに対して「含意関係あり」と出力する。

3.3.5. Natural Logic による手法

MacCartney らは、含意関係認識に関わる言語的特徴や単調性を利用して、自然言語を論理式に変換せず、自然言語のまま含意判定を行う Natural Logic を提案した[70]。Natrual Logic とは、自然言語のレベルで 1 つのテキストからもう 1 つのテキストを推論するための論理体系であり、自然言語を述語論理式に変換する必要がないため、より自然な形で推論することができる。Natural Logic に基づく含意関係認識の基本的な考え方については、次に述べる増田らの研究事例の紹介の中で説明する。

増田らは、日本語を対象とした Natural Logic による含意関係認識の手法を提案している[71]。２つのテキスト（t1, t2）に対して、単調減少と非単調の単語リストを用いて文節の単調性を求め、文節の位置関係に従って対応付けられた t1 と t2 の文節どうしを比較し、EDR 電子化辞書[16]における概念の上位下位関係を用いて文節単位で含意関係を導き、文節単位の含意関係を統合して文全体の含意関係を推論する。まず、含意関係の定義を表 3-8 のように定めている。

表 3-8: 含意関係の定義 [71]

含意関係	定義
=	t1→t2：○　かつ　t2→t1：○
<	t1→t2：○　かつ　t2→t1：×
>	t1→t2：×　かつ　t2→t1：○
?	t1→t2：×　かつ　t2→t1：×

次に、単語を単調増加、単調減少、非単調という 3 種類に分類する。ある単語をテキスト対の同じ位置に同じように付け加えたとき、含意関係が保たれる場合を単調増加の単語、含意の方向が反転する場合を単調減少の単語といい、単調増加でも単調減少でもない単語を非単調としている。含意関係が変わる例と変わらない例として表 3-9 の例を挙げている。t1=「東京にいる」、t2=「日本にいる」のとき、「今」という単語は、これを追加したときの含意関係は変わらないので単調増加となり、「ない」という単語は、これを追加したときの含意関係が反転するので単調減少となる。

表 3-9: 含意関係が変わる例と変わらない例 [71]

t1	t2	含意関係
東京にいる。	日本にいる。	<
今、東京にいる。	今、日本にいる。	<
東京にいない。	日本にいない。	>

増田らは、表 3-10 のように単調減少の単語リストと非単調の単語リストを作成し、リストに含まれない単語を単調増加としている。

表 3-10: 単調減少と非単調の単語のリスト [71]

単調性	単語
単調減少	ない，無い，たら，と（接続助詞）， は（文末は過去形以外），禁止，・・・
非単調	最も

含意関係認識の判定は以下のように行う。まず、t1 と t2 に含まれる文節について、文節の文内における位置や文節の係り受け関係を手がかりに文節間の対応付けを行う。次に、対応付けられた文節対 b1,b2 に対して含意関係の判定を行う。EDR 概念辞書を用いて、b1,b2 に含まれる単語の語義(概念 ID)c1,c2 を求め、EDR 概念体系において c1 が c2 の上位語(下位語)であるときには文節間の含意関係を b1>b2(b1<b2)とする。

b1>b2,b1<b2 の両方が成立するときは b1=b2、両方とも成立しないときは b1?b2 とする。さらに、テキストの係り受け構造をもとに、単調性を考慮しつつ文節単位の含意関係を組み合わせていく。係り側の文節の含意関係、受け側の文節の含意関係、これらを組み合わせた後の含意関係の対応を表 3-11 に示す。

表 3-11: 含意関係の組み合わせ [71]

		受け側の含意関係			
		=	<	>	?
係り側の含意関係	=	=	<	>	?
	<	< (>)	< (?)	? (>)	?
	>	> (<)	? (<)	> (?)	?
	?	?	?	?	?

注)括弧内の関係は､係り先の文節の単調性が単調減少であるときの含意関係を表わす。文節の単調性は、文節内の単語の単調性の積で決める。

上記の係り受け構造による含意関係の導出を t1,t2 の 2 つのテキストに適用し、各テキストの含意関係を求める。そして、各テキストの文全体の含意関係を表 3-11 にしたがって組み合わせ、導出された含意関係を 2 つのテキストの最終的な含意関係とする。

増田らの Natural Logic の手法による含意関係認識の例を表 3-12、図 3-22 に示す。図 3-22 は、文節の含意関係を組み合わせて最終的に t1 全体の含意関係を「>」と決定している過程を示している。テキスト t2 についても同様に含意関係を求めると「>」となる。最後に表 3-11 にしたがって 2 つの含意関係を組み合わせると、最終的な含意関係は表 3-12 に示すように「>」となる。

表 3-12: Natural Logic による含意関係認識の例文 [71]

t1	ツバメが低く飛ぶと雨が降る。
t2	鳥が低く飛ぶと豪雨が降る。
含意関係	＞

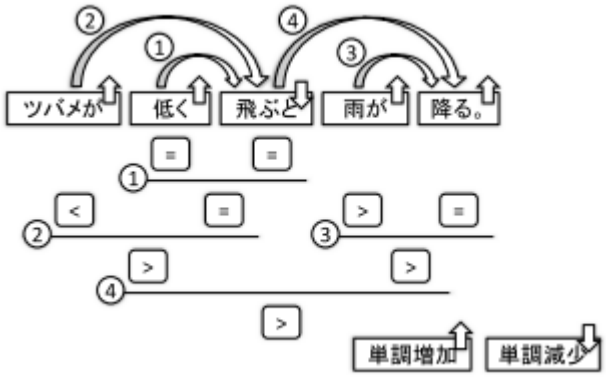

図 3-22: t1 の文全体の含意関係の導出 [71]

増田らは、377 ペアからなるテストセットに対する含意関係認識の正解率が 50.4%であったと報告している。課題として、単調減少と非単調の単語リストを充実させること、２つのテキストの文節を対応付ける際、多少構文構造が違うときにも位置合わせに失敗してしまわないようにすることなどを挙げている。

3.3.6. 含意関係認識のための知識ベース

含意関係認識には様々な知識が使われる。本項では、含意関係認識によく使われる知識ベースを紹介し、また含意関係認識に有用な知識ベースを構築する試みを紹介する。また、主に知識ベースを利用した含意関係認識手法についても述べる。

知識ベースによるアプローチでは、DIRT [73], VerbOcean [74], WordNet [75]等の言語リソースが利用されている。日本語のリソースとしては日本語 WordNet [76]が存在する。DIRT は依存木で表現されたパラフレーズのデータベースである。VerbOcean は動詞間の意味的関係を表わした言語資源である。WordNet は包括的な単語の概念辞書であり、単語は同義語のグループにまとめられ、さらにグループ間の意味関係が表現されている。これらの言語リソースは、含意関係認識に必要な背景知識として、本研究で調査した研究事例を含め、広く利用されている。

小谷らは、含意関係認識における問題点を議論しやすくするために、１つまたは２つの少数の要因によって含意関係が成立することを判断できる事例を集めた日本語 TE データを構築するとともに、類義表現に基づく推論関係を認識する手法を提案した[77]。日本語 TE データとして、含意関係の成立が推測できる要因を「包含」「語彙（体言)」「語彙（用言)」「構文」「推論」の５つに分類し、さらにそれぞれの分類に下位分類を設けて、各分類に当てはまる事例(含意関係にある文の組)を人手で作成した。それぞれの分類・下位分類に該当する事例の数を**表 3-13** に示す。

表 3-13: 作成した TE データの分類と属するデータ数 [77]

分類	下位分類	◎	○	△	×	計	合計
包含	節	43	5	1	43	92	263
	並列	34	1	1	34	70	
	補文	25	2	4	25	56	
	名詞句	21	3	0	21	45	
語彙（体言）	定義的	92	60	19	46	217	557
	同義語	57	18	5	45	125	
	下位→上位	35	3	3	20	61	
	上位→下位	9	8	4	25	46	
	対義語	18	5	1	14	38	
	名詞の格関係	18	5	0	13	36	
	性質	7	9	6	12	34	
語彙（用言）	言い換え	126	87	10	74	297	767
	前提的	66	61	16	22	165	
	含意	45	67	19	10	141	
	副詞	33	21	2	21	77	
	内包	26	9	5	7	47	
	対義語	14	4	3	19	49	
構文	主語の変換	60	7	7	35	109	219
	複文の変換	13	24	4	15	56	
	強調構文	19	3	1	11	34	
	名詞句	10	0	0	10	20	
推論	結果→原因	48	93	67	28	236	868
	原因→結果	43	95	51	41	230	
	省略の類推	65	51	6	19	141	
	副助詞＋対応	23	27	18	28	96	
	副助詞＋一般化	17	13	7	10	47	
	時間軸・数量	20	5	1	10	36	
	順接・逆接＋一般化	10	12	2	10	34	
	順接・逆接＋具体化	9	5	4	10	28	
	間接発話行為	3	8	8	1	20	
		1009	711	275	679	2674	

◎：t が真であったとき h が必ず真であるといえる場合

例）彼は人間である。→ 彼は哺乳類である。

○：t が真であったとき h が正しいと常識的には考えられる場合
　例）彼はオムレツを作った。→彼は卵を使った。
△：t が真であったとき h が真である可能性がある程度考えられる場合
　例）信号は赤ではなかった。→信号は青だった。
×：t が真であったとき h が全くの誤りだとわかる場合
　例）彼は読書が好きだ。→彼は読書が嫌いだ。

さらに、小谷らは、「語彙(体言)」に分類される事例の含意関係認識を実現する手法を提案している。「語彙(体言)」の事例とは、Text 内の名詞の意味や性質から Hypothesis の成立が推測できる事例である。語や句の同義・上位下位関係を国語辞典やウェブテキストから取得して、以下の推論パターンにより含意関係認識を行う。
（１） 語や句の同義関係による推論パターン
（２） 語や句の上位下位関係による推論パターン
（２）については、語や句の上位下位関係が導く推論は、語が現れる文のタイプや語の文中での役割によって変化するため、文を「行為を表わす文」と「性質・状態を表わす文」に分け、推論パターンを記述している。

大西らは、含意関係認識技術には、モノに関する知識の他にコト、即ち事態についても上位下位関係や部分全体関係、因果関係などの知識が必要であるという認識の下、岩波国語辞典タグ付きコーパスにおける動詞の語釈文を利用することにより、人手で事態間関係知識の整備を行った[78]。
川添らは、「人工頭脳プロジェクト ―ロボットは東大に入れるか」 [52]における世界史分野の問題について、頻繁に出題される史実であるか否かという問題に答えるために、イベントの分類とイベント成立に関わる必要条件を分析し、世界史オントロジーの構築を進めている[79] [80]。イベント分類として、
(1) 参加者の存在時間による分類
(2) 参加者の場所的関係による分類
(3) 再発可能性（単発かどうか）による分類
(4) 参加者の両立不可能性に関する分類
などを提案している。

3.3.7. 含意関係認識に有効な言語的特徴の分析

MacCartney ら [69]は、3.3.2 で述べた de Marneffe ら [72]のアラインメント処理に加えて、含意関係を判定するための言語的特徴を検討した。MacCartney らが用いた PASCAL RTE データセットの例を表 3-14 に示す。

表 3-14: PASCAL RTE データセットの例 [69]

ID	Text	Hypothesis	Entailed
59	Two Turkish engineers and an Afghan translator kidnapped in December were freed Friday.	Translator kidnapped in Iraq	no
98	Sharon warns Arafat could be targeted for assassination.	prime minister targeted for assassination	no
152	Twenty-five of the dead were members of the law enforcement agencies and rest of the 67 were civilians.	25 of the dead were civilians.	no
231	The memorandum noted the United Nations estimated that 2.5 million to 3.5 million people died of AIDS last year.	Over 2 million people died of AIDS last year.	yes
971	Mitsubishi Motors Corp.'s new vehicle sales in the US fell 46 percent in June.	Mitsubishi sales rose 46 percent.	no
1806	Vanunu, 49, was abducted by Israeli agents and convicted of treason in 1986 after discussing his work as a mid-level Dimona technician with Britain's Sunday Times newspaper.	Vanunu's disclosures in 1968 led experts to conclude that Israel has a stockpile of nuclear warheads.	no
2081	The main race track in Qatar is located in Shahaniya, on the Dukhan Road.	Qatar is located in Shahaniya.	no

Text と Hypothesis 間の単語のオーバーラップ率に基づくような方法だけでは、ID2081 のような例は間違えてしまうことを MacCartney らは指摘している。

MacCartney らは、含意関係判定に貢献する 28 個の言語的特徴を抽出した。最も重要度の高い特徴として、極性的特徴（Polarity features）、付加的特徴（Adjunct features）、反意的特徴（Antonymy features）、モダリティ的特徴（Modality features）、叙実性の特徴（Factivity features）、数量詞の特徴（Quantifier features）、数・日付・時間的特徴（Number,date,and time features）を挙げている。以下、これらについて述べる。

・極性的特徴（Polarity features）

単純な否定（not）、下向き単調数量詞（no, few）、制約（without, except）、最上位表現(tallest)などがある。

・付加的特徴（Adjunct features）

言葉の付加または削除によって、含意関係が変わることを指す。例えば、Dogs barked loudly ⊨Dogs barked は、言葉が削除されても含意関係が維持される例であり、Dogs barked ⊭Dogs barked today は、言葉の付加によって含意関係が変わる例である。表 3-14 中の ID 59 の例は、in Iraq という言葉の付加によって含意関係が no となる例である。

・反意的特徴（Antonymy features）

表 3-14 中の ID 971 の例のように、反意語を含むことによって含意関係が変わることがある。反意的な関係のペアについて、（ｉ）極性のマッチングの文脈でそのような単語が現れる場合、(ⅱ) Text 中の言葉のみが否定的極性の文脈で現れる場合、(ⅲ) Hypothesis 中の言葉のみが否定的極性の文脈で現れる場合、の３つの場合の特徴を抽出した。

・モダリティ的特徴（Modality features）

表 3-14 中の ID 98 のように、モダリティが含意関係に影響することがある。must や maybe のようなモダリティに関係する語の有無によって、Text と Hypothesis を possible, not possible, actual, not actual, necessary, not necessary の 6 種のモダリティのいずれかにマップするようにした。そして、Text と Hypothesis のモダリティのペアは、yes, weak yes, don't know, weak no, no のいずれかの含意判定にマップされる。例を以下に示す。

(not possible ⊨ not actual)? ⇒ yes

(possible ⊨ necessary)? ⇒ weak no

・叙実性の特徴（Factivity features）

動詞フレーズが文脈に埋め込まれることによって、現実に起きたことか否かの意味合いが変わることがある。例えば、以下の例を挙げる。

The gangster tried to escape ⊭ The gangster escaped.

そのため、叙実的(factive)な動詞と非叙実的な動詞のリストを作り、Hypothesis グラフのルートノードとアラインメントされる Text グラフのノードの親ノードが叙実的か非叙実的のどちらのクラスに該当するかを判定するようにした。

・数量詞の特徴（Quantifier features）

Text と Hypothesis 中の数量詞は、no, some, many, most, all の 5 種のいずれかにマッピングするように集約した。

・数・日付・時間的特徴（Number,date,and time features）

表現の正規化（日付など）を行い、曖昧なマッチングにある程度対応できるようにした。表 3-14 中の ID 1806 の例では、年のミスマッチを検出できたが、ID 231 の例では、over という単語の重要性を捉えられず、ミスマッチが報告される結果となった。

MacCartney らは、これらの特徴を学習素性とし、含意関係の有無を判定する分類器を教師あり機械学習により獲得することも試みている。機械学習アルゴリズムはロジスティック回帰を採用した。RTE 1 のテストデータを用いた評価実験では、含意関係認識の正解率は 59.1%であった報告している[69]。

3.4. まとめ

含意関係認識は、文章間の類似度やアラインメント、構文構造の変換、形式論理、Natural Logic 等のアルゴリズム的アプローチと、知識ベースによるアプローチが並行して研究されてきた。知識ベースの充実と共に、論理的推論の向上が有効であると考えられるが、自然言語を論理式に変換することの難しさも指摘されている [81]。また、文章表現の多様性以外に、文章の言外に隠された意味を正確に把握して処理するためには、より深い文章理解の技術や常識的な知識の獲得・整備が必要であると考えられる。

第4章　おわりに

本課題研究では、より知的な質問応答システムを実現するために必要となる技術のうち､原因･理由や含意関係認識を対象とした推論方式の研究についてサーベイを行った。原因・理由についての推論方式に関しては why 型質問応答システムを調査対象とした。

why 型質問応答システムについては、大量の文書情報に対して、手掛かり語などによるルールベース型手法や機械学習によるルールの自動獲得の手法、意味解析ネットワークによる手法などが研究されてきた。含意関係認識の研究と比較すると、機械学習のように共通して使われている手法はあるが、含意関係認識で行われている形式論理、Natural Logic 等の論理的推論によるアプローチは why 型質問応答システムの事例では見られなかった。今後、論理的推論によるアプローチも why 型質問応答システムの研究に有効ではないかと考えられる。また、文書情報から取得された原因・理由が本当に正しいかという妥当性や信憑性を判断する研究も今後の課題として考えられる。

含意関係認識については､語の類似度やアラインメント､構文構造の変換､形式論理、Natural Logic 等に基づくアルゴリズム的アプローチと、知識ベースを利用するアプローチが研究されてきた。双方のアプローチを組み合わせて質的向上を図っていくことにより、全体的な認識精度向上が期待できると考えられる。

含意関係認識は、英語を対象とした研究が先行していたが、日本語に関しても、「人工頭脳プロジェクト ―ロボットは東大に入れるか」[52]において含意関係認識が中核技術の一つとして位置付けられ、今後の更なる発展が期待されている。

謝辞

本課題研究を進めるにあたり、多大なご支援、ご指導を頂いた白井清昭准教授に深くお礼申し上げます。中間審査の場において貴重なご意見を頂いた東条敏教授、池田心准教授に厚く感謝いたします。また、北陸先端科学技術大学院大学先端領域社会人教育院関係者各位に深くお礼申し上げます。

参考文献

[1] 諸岡 心, 福本淳一: Why型質問応答のための回答選択手法, 信学技報, CL2005-107, 2006.

[2] 奥村学 監修, 磯崎秀樹, 東中竜一郎, 永田昌明, 加藤恒昭:質問応答システム, p.167, コロナ社, 2009

[3] NTCIR : http://research.nii.ac.jp/ntcir/index-ja.html

[4] 田村元秀, 村上仁一, 徳久雅人, 池原 悟:Web 検索エンジンを用いた Why 型質問応答システム, in NLP, pp.1021～1024 (2008)

[5] 渋沢 潮, 林 貴宏, 尾内理紀夫 : Why 型質問の回答文を WEB から自動抽出するシステムの開発と評価, 情処学論, vol.48, no.3, pp.1512-1523, 2007

[6] 大野 晋, 浜西正人 : 角川類語新辞典, 角川書店, 1992

[7] 生田目弥寿 : 日本語教師のための現代日本語表現文典, 凡人社, 1996

[8] 益岡隆志, 田窪行則 : 基礎日本語文法―改定版, くろしお出版, 2004

[9] 長尾 真（編）: 自然言語処理（岩波講座ソフトウェア科学, No.15）, 岩波書店, 1996

[10] 乾孝司, 奥村学 : 文書内に現れる因果関係の出現特徴調査, 情報処理学会自然言語研究会 NL-167, 2005

[11] 磯崎秀樹, 東中竜一郎 : パターンマイニングを用いて「なぜ」に答えるシステム, 言語処理学会第 14 回年次大会発表論文集, pp.1025-1028, 2008

[12] Higashinaka, R. and Isozaki, H.: Automatically Acquiring Causal Expression Patterns from Relation-annotated Corpora to Improve Question Answering for why-Questions, 2008

[13] 東中 竜一朗, 磯崎 秀樹 : 公開特許公報 特開 2009-157791,[発明の名称]「質問応答方法、装置、プログラム並びにそのプログラムを記録した記録媒体」, 2009

[14] 工藤拓, 松本裕治 : カーネル法を用いた言語解析における高速化手法, 情報処理学会論文誌, 45 No.9, pp.2177-2185, 2004, http://chasen.org/~taku/software/cabocha

[15] Kudo, T. and Matsumoto, Y.: A boosting algorithm for classification of semi-structured text, In Proc, EMNLP, pp.301-308, 2004

[16] 日本電子化辞書研究所 : EDR 電子化辞書 2.0 版 仕様説明書, 2001.

[17] Ikehara, S., Shirai, S., Yokoo, A., and Nakaiwa, H.: Toward an MT System without Pre-Editing -Effects of New Methods in ALT-J/E-, In Proc. Third Machine Translation Summit: MT Summit III, pp.101-106, 1991

[18] Fuchi, T. and Takagi, S.: Japanese morphological analyzer using word co-occurrence –JTAG–. In Proc. 17th COLING and 36th ACL. Vol. 1,

pp.409–413, 1998

[19] Fukumoto: Question answering system for non-factoid type questions and automatic evaluation based on BE method, In Proc. NTCIR, 441-447, 2007

[20] Isozaki, H: An analysis of a high-performance Japanese question answering system, ACM Transactions on Asian Language Information Processing (TALIP) 5, 3, pp.262-279, 2005

[21] Minoru Harada, Yuhei Kato, Kazuaki Takehara, Masatsuna Kawamata, Kazunori Sugimura, and Junichi Kawaguchi: QA System Metis Based on Semantic Graph Matching, Proc. of the 6th International Conference on NII Test Collection for IR Systems(NTCIR6), Tokyo, Japan, pp.448-459, 2007

[22] 原田実, 松田源立：インターネットへの高精度な質問応答システムの開発, 電気通信普及財団研究調査報告書第 23 号, pp.594-602, 2008

[23] 杉村和徳, 山本哲哉, 木村健太郎, 鳥居隼, 韓東力, 原田実：意味解析システム SAGE の精度向上と利便性の向上, 情報処理学会第 67 回全国大会論文集, 1J-02, 第 2 分冊, pp.67-68, 2005

[24] 高山真行, 今村泰香, 久保田裕章, 原田 実：NonFactoid 型質問文と回答文との意味的関係に基づく質問応答システム Metis, 情報処理学会研究報告. SLP, 音声言語情報処理 2010-SLP-81(17), 1-7, 2010

[25] 学研サイエンスキッズ： http://kids.gakken.co.jp/kagaku/index.html

[26] 大友謙一, 村上仁一, 徳久雅人, 池原 悟：WHY 型 QA システムにおける回答抽出方法の改良, 言語処理学会 第 15 回年次大会 発表論文集（2009 年 3 月）

[27] 雑学のすゝめ： http://homepage1.nifty.com/tadahiko/

[28] 石下円香, 佐藤 充, 森 辰則：Web 文書を対象とした質問の型に依らない質問応答手法, 人工知能学会論文誌, Vol. 24, No.4, pp.339-350, 2009

[29] Han, K.-S., Song, Y.-I., and Rim, H.-C.: Probabilistic model for definitional question answering, in SIGIR, pp.212-219, 2006

[30] S. Verberne, L. Boves, N. Oostdijk, and P. Coppen: Using syntactic information for improving why-question answering, Proc. COLING '08, pp.953-960, 2008

[31] Buttcher, S., C.L.A. Clarke, and G.V. Cormack: Domain-Specific Synonym Expansion and Validation for Biomedical Information Retrieval, 2004

[32] Hovy, E.H., U. Hermjakob and D. Ravichandran: A Question/Answer Typology with Surface Text Patterns, In Proc. of HLT 2002, 2002

[33] Denoyer, L. and P. Gallinari: The Wikipedia XML corpus, ACM SIGIR Forum, 40(1):64–69, 2006

[34] 田中克幸, 滝口哲也, 有木康雄：Bag of Grammer を用いたドメイン依存性の少ない Why テキストセグメント識別器の自動構築法, 電子情報通信学会論文誌. D, 情

報・システム J94-D(12), 2047-2057, 2011

[35] 呉 鍾勲, 鳥澤健太郎, 橋本 力, 川田拓也, De Saeger, Stijn, 風間淳一, 王 軼謳：意味的知識を用いたWhy型質問応答の改善, 言語処理学会 第18回年次大会 発表論文集, 2012年3月

[36] M.Murata, S. Tsukawaki, T.Kanamaru, Q.Ma, and H.Isahara: A System for Answering Non-Factoid Japanese Questions by Using Passage Retrieval Weighted Based on Type of Answer, In Proc. Of NTCIR-6, 2007

[37] 車 智修, 鍋島啓太, 水野淳太, 岡崎直観, 乾 健太郎: 文章構造を用いたWhy型質問応答システム, The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013

[38] Jeh, G. and Widom, J.: Scaling personalized web search, in Proceedings of the 12th international conference on World Wide Web, pp. 271-279 ACM, 2003

[39] Asli Celikyilmaz, Marcus Thint, and Zhiheng Huang: A graph-based semi-supervised learning for questions-answering, In Proceedings of the Annual Meeting of the Association for Computational Linguistics, pages 719-727, Suntec, Singapore, August 2009, Association for Computational Linguistics. 11

[40] Ido Dagan, Dan Roth, Mark Sammons, and Fabio Zanzotto: Recognizing Textual Entailment –Models and Applications, Morgan & Claypool publishers, 2013

[41] Sandra Harabagiu and Andrew Hickl: Methods for using textual entailment in open-domain question answering. In In Proc. of ACL-2006, pages 905-912, 2006

[42] Ido Dagan, Oren Glickman, and Bernardo Magnini: The PASCAL recognizing textual entailment challenge, In Quionero-Candela et al., editor, First PASCAL Machine Learning Challenges Workshop, pp.177-190, Springer-Verlag, 2006

[43] Roy Bar-Haim, Ido Dagan, Bill Dolan, Lisa Ferro, Danilo Giampiccolo, Bernardo Magnini and Idan Szpektor: The second PASCAL recognizing textual entailment challenge, In Proceedings of the Second PASCAL Challenges Workshop on Recognizing Textual Entailment, Venice, Italy, 2006

[44] Danilo Giampiccolo, BernardoMagnini, Ido Dagan and Bill Dolan: The third PASCAL recognizing textual entailment challenge, In Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp.1-9, Association for Computational Linguistics, Prague, June 2007

[45] Danilo Giampiccolo, Hoa T. Dang, Bernardo Magnini, Ido Dagan, and Bill Dolan: The fourth PASCAL recognizing textual entailment challenge, In Proceedings of the First Text Analysis Conference (TAC2008), 2008

[46] Luisa Bentivogli, Ido Dagan, Hoa T. Dang, Danilo Giampiccolo, and Bernardo

Magnini: The fifth PASCAL recognizing textual entailment challenge, In Proceedings of the Second Text Analysis Conference (TAC 2009), 2009

[47] Luisa Bentivogli, Peter Clark, Ido Dagan, Hoa T. Dang, and Danilo Giampiccolo: The sixth PASCAL recognizing textual entailment challenge, In Proceedings of the Third Text Analysis Conference (TAC 2010), 2010

[48] Luisa Bentivogli, Peter Clark, Ido Dagan, and Danilo Giampiccolo: The seventh PASCAL recognizing textual entailment challenge, In Proceedings of the Fourth Text Analysis Conference (TAC 2011), 2011

[49] H. Shima, H.Kanayama, C.-W. Lee, C.-J. Lin, T.Mitamura, Y. Miyao, S. Shi, and K. Takeda: Overview of NTCIR-9 RITE: Recognizing inference in text. In Proc. NTCIR-9, 2011

[50] Y.Watanabe, Y.Miyao, J.Mizuno, T.Shibata, H.Kanayama, C-W.Lee, C-J.Lin, S.Shi, T.Mitamura, N.Kando, H.Shima and K.Takeda: Overview of the Recognizing Inference in Text (RITE-2) at NTCIR-10, Proceedings of the 10th NTCIR Conference, June 18-21, 2013, Tokyo, Japan

[51] 宮尾祐介, 嶋英樹, 金山博, 三田村照子： 大学入試センター試験を題材とした含意関係認識技術の評価, 言語処理学会 第 18 回年次大会 発表論文集（2012 年 3 月）, pp.439-442, 2012

[52] 国立情報学研究所: ロボットは東大に入れるか Todai Robot Project, http://21robot.org/

[53] Rod Adams: Textual Entailment Through Extended Lexical Overlap, In Proceedings of the Second PASCAL Challenges Workshop on Recognizing Textual Entailment, 2006

[54] Yashar Mehdad, Matteo Negri, Elena Cabrio, Milen Kouylekov1 and Bernardo Magnini: EDITS: An Open Source Framework for Recognizing Textual Entailment, In Proceedings of the Second Text Analysis Conference (TAC 2009), pp.169-178, 2009

[55] Michael Heilman, Noah A. Smith: Tree edit models for recognizing textual entailments, paraphrases, and answers to questions, In Proceedings of the Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL), pp.1011-1019, Los Angeles, California, June 2010, Association for Computational Linguistics.

[56] 後藤隼人, 水野淳太, 村上浩司, 乾健太郎, 松本裕治： 文間関係認識のための構造的アラインメント, 言語処理学会第 16 回年次大会発表論文集 E3-7, 2010

[57] Mark Sammons, V.G.Vinod Vydiswaran, Tim Vieira, Nikhil Johri,Ming-Wei Chang, Dan Goldwasser, Vivek Srikumar, Gourab Kundu,Yuancheng Tu, Kevin

Small, Joshua Rule, Quang Do, Dan Roth: Relation Alignment for Textual Entailment Recognition, In Proceedings of the Second Text Analysis Conference (TAC 2009), 2009

[58] Rodrigo de Salvo Braz, Roxana Girju, Vasin Punyakanok, Dan Roth, Mark Sammons: An Inference Model for Semantic Entailment in Natural Language, In Proceedings of the National Conference on Artificial Intelligence(AAAI), pp.1678-1679, 2005

[59] Braz, R., Girju, R., Punyakanok, V., Roth, D., and Sammons, M.: Knowledge representation for semantic entailment and question-answering, In IJCAI'05 Workshop on Knowledge and Reasoning for Answering Questions, 2005

[60] Roy Bar-Haim, Ido Dagan, Iddo Greental, Idan Szpektor, Moshe Friedman: Semantic inference at the lexical-syntactic level for textual entailment recognition, In Proceeding of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp.131-136, Prague, June 2007, Association for Computational Linguistics

[61] Marie-Catherine de Marneffe, Anna N. Rafferty and Christopher D. Manning: Finding contradictions in text, In Proceedings of the Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT), pp.1039-1047, Columbus, Ohio, June 2008

[62] Dan Moldovan, Christine Clark, Sanda Harabagiu, Steve Maiorano: COGEX: A logic prover for question answering, In Proceedings of HLT-NAACL, pp.87-93, Edmonton, May-June 2003

[63] Marta Tatu, Brandon Iles, John Slavick, Adrian Novischi, Dan Moldovan: Cogex at the second recognizing textual entailment challenge, Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment, pp.104-109, 2006

[64] Elena Akhmatova: Textual entailment resolution via atomic propositions, In Proceedings of the First PASCAL Challenges Workshop on Recognizing Textual Entailment, 2005

[65] Johan Bos, Katja Markert: Recognising Textual Entailment with Logical Inference, In Proceedings of the Human Language Technology Conference and the Conference on Empirical Methods in Natural Language Processing, pp.628-635, Vancouver, 2005

[66] Rajat Raina, Andrew Y. Ng and Christopher D. Manning: Robust textual inference via learning and abductive reasoning, In Proceedings of the National Conference on Artificial Intelligence (AAAI), 2005

[67] 田然, 宮尾祐介： 外延の抽象表現を用いた論理推論, 言語処理学会第 20 回年次大会発表論文集, 2014 年 3 月

[68] P.Liang, M.Jordan, D.Klein: Learning dependency-based compositional semantics, in ACL, 2011

[69] Bill MacCartney, Trond Grenager, Marie-Catherine de Marneffe, Daniel Cer, and Christopher D. Manning: Learning to recognize features of valid textual entailments, In Proceedings of the North American Association for Computational Linguistics (NAACL), 2006

[70] Bill MacCartney, Christopher D.Manning: Natural Logic for Textual Inference, In ACL-07 Workshop on Textual Entailment and Paraphrasing, 2007

[71] 増田涼良, 杉本徹：Natural Logic を用いた日本語テキストの含意関係認識, 人工知能学会全国大会論文集, 26th Vol., 4K1-OS-2-3, 2012

[72] Marie-Catherine de Marneffe, Trond Grenager, Bill MacCartney, Daniel Cer, Daniel Ramage, Chloe´ Kiddon, Christopher D. Manning: Aligning semantic graphs for textual inference and machine reading, In AAAI Spring Symposium at Stanford 2007

[73] Dekang Lin, Patrick Pantel: DIRT: discovery of inference rules from text, In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining, pp.323-328, 2001

[74] Timothy Chklovski and Patrick Pantel: VerbOcean: Mining the Web for Fine-Grained Semantic Verb Realations, In Proceedings of the Conference on Empirical Methods for Natural Language Processing (EMNLP), pp.33-40, 2004

[75] Christiane Fellbaum (editor): WordNet: An Electronic Lexical Database, MIT Press, Cambridge, MA, 1998

[76] 独立行政法人情報通信研究機構: 日本語 WordNet, http://nlpwww.nict.go.jp/wn-ja/

[77] 小谷通隆, 柴田知秀, 中田貴之, 黒橋貞夫 : 日本語 Textual Entailment のデータ構築と自動獲得した類義表現に基づく推論関係の認識, 言語処理学会 第 14 回年次大会 発表論文集, pp.1140-1143, 2008.

[78] 大西良明, 乾健太郎, 松本裕治：事態間関係知識の整備と含意文生成への応用, 言語処理学会 第 14 回年次大会発表論文集, pp. 1152-1155, 2008.

[79] 川添愛, 宮尾祐介, 松崎拓也, 横野光, 新井紀子,計算機による大学入試問題への解答に向けた世界史オントロジーの設計,人工知能学会第 26 回全国大会, 2012 年 6 月

[80] 川添愛、宮尾祐介、松崎拓也、横野光、新井紀子,「史実としてありえない」という判断を可能にする世界史オントロジー,2013 年度人工知能学会全国大会（JSAI2013）論文集, https://kaigi.org/jsai/webprogram/2013/pdf/647.pdf

[81] 稲田和明, 松林優一郎, 井之上直也, 乾健太郎：効率的な推論処理のための日本語文の論理式変換に向けて, 言語処理学会 第 19 回年次大会 発表論文集, pp.608-611, 2013