2004－NL－162　(21)
2004／7／16

質問応答技術は情報アクセス対話を実現できるか

加藤 恒昭[1]　福本 淳一[2]　桝井 文人[3]　神門 典子[4]
東京大学[1]　立命館大学[2]　三重大学[3]　国立情報学研究所[4]

概要

対話的な情報アクセスの場面での利用を考えた場合に，質問応答システムにはどのような能力が必要であるかについて，情報アクセス対話を模擬した場面での質問の収集と分析を通じて明らかにした．名称を解答の範囲とするような質問応答システムが有効である一方，システムは様々な語用論的現象を処理できる必要がある．この結果に基づき，そのような場面で必要となる質問応答システムの能力を定量的客観的に評価するタスクを提案する．その設計とテストセットについて説明し，本タスクが，分析で得られた対話の特徴を適切に反映していることを示す．

Are Open-domain Question Answering Technologies Useful for Information Access Dialogues?

Tsunaeki Kato[1]　Jun'ichi Fukumoto[2]　Fumito Masui[3]　Noriko Kando[4]
The University of Tokyo[1]　Ritsumeikan University[2]　Mie University[3]　National Institute of Informatics[4]

Abstract

In this paper, we empirically examine what kind of abilities question answering systems need in order to participate in information access dialogues. Our study shows that it is useful to use question answering systems that cover factoid questions having values and names as those answers, while such systems need to handle a wide range of anaphoric phenomena observed in those situations. We also propose a challenge for evaluating those abilities objectively and quantitatively, and explain its design and a test set constructed, emphasizing that the challenge is properly reflected the characteristics of information dialogues revealed through our empirical study.

1　はじめに

質問応答技術は自然言語によって表現された質問に文書でなくその情報そのもので回答する事を可能とするもので，情報アクセスの新しい形として期待されている [12]．幾つかの例外はあるが [1][11]，事実に関するお互いに独立した質問に一問一答形式で回答するシステムが現在の研究の中心である．質問応答技術の持つ可能性は大きく，現状からの様々な展開が模索されているが，その中にアナリストやレポータが利用しうる枠組みへの発展がある [2]．そのひとつとして，新人レポータがある事件の記事を執筆するために，彼の記事で答えられるべき大きな質問をより簡単な質問の集まりに翻訳してシステムに訊ねるという利用形態が考えられている．

一方，複数文書要約と質問応答との関連も指摘されている．彼の講演の中で Hovy は複数文書の要約を一連の質問応答に還元する可能性について論じているし [6]，SUMMAC では文書要約の評価方法のひとつとして，要約対象文書のトピックにおいて必須となる質問を複数用意しそれにどの程度回答できるかを要約のよさの指標とすることが試みられている [10]．このことは，与えられた一連の質問に回答できるような質問応答システムが文書要約を支援できることを示唆している．

本稿では，まず，現在の質問応答システムがこのような期待に応えうるかを検証する．つまり，質問応答システムが対話的に行われる一連の情報アクセスを支援できるか，質問応答システムを対話相手とした情報アクセス対話が成立しうるかについての経験論的研究について報告する．ここで情報アクセス対話とは，利用者が特定のトピックに関するレポートを作成するための情報を対話的に収集する場合や，その興味の赴くままに情報をブラウジングする場合になされるような対話をさす．このような対話場面においてどのような質問が現れるのか，どのような語用論的現象が観察されるのかを質問の収集と分析を通じて明らかにした．次に，そこで得られた知見に基づいて設計された評価用タスクについて述べる．このタスクは情報アクセス対話において必要となる質問応答システムの文脈処理能力を定量的に評価することを目的としたものである．本稿では特にこのタスクの設計と経験論的研究で得られた対話の特徴との関係について述べる．

2　情報アクセス対話における質問

質問応答システムが情報アクセス対話に参加するためには様々な能力が必要となる [2]．まず，実時間の応答が必須である．そして，与えられた質問を対話文脈を考慮して適切に解釈し，必要な情報を付加した協調的応答をする必要もある．更にはシステムからの質問によって利用者の意図やゴールの曖昧性を解消したり，システムが提案を行うよ

うな主導権混在型対話により利用者の問題解決をリードすることも重要となろう．これら様々な必要性の中で，まずは，そもそも情報アクセス対話を扱うためにはどのような質問に答えられる必要があるのか，そして，対話の実現の基本となる対話文脈を考慮した質問の解釈，つまり照応解消や省略処理等のいわゆる文脈処理はどの程度必要なのかに着目し，情報アクセス対話でなされる質問を収集し分析した．情報アクセス対話には様々なバラエティが考えられるが，ここでは与えられたトピックについてのレポートを書くための情報を得るような対話に焦点を当てた．

2.1 質問の収集

新聞記事から選択した人物，組織，出来事等のトピックを被験者に提示し，それに関するレポートを作成するという状況を想定し，そこに含めたいと考える情報を質問文の形式で表現するように指示した．レポートは与えられたトピックの事実関係をまとめたもので予測や意見はそこに含めないものとし，質問文型は疑問代名詞を含む Wh 型に限定した．作成した質問に次々と回答が得られるという想定で，ひとつのトピックについて複数の質問を作成させ，質問中に代名詞等の参照表現を含めることを許した．これにより自然な質問の系列が作成されることを期待した．

与えられたトピックに関する知識の量と質問の内容や形式との関係への興味から，トピックの提示については以下の 3 種のパタンを設けた．

(少) TREC のトピック定義 [14] における title に相当する短い記述のみを与える．

(中) そのトピックに関する代表的な記事，一定の長さより長い場合はそのリード部分を添える．この情報は TREC トピック定義の narrative にほぼ相当する．

(多) そのトピックに関する記事を 5 件ずつ添える．

質問作成の際には，自分の訊ねる質問の解答が与えられた記事中に含まれているかは意識しないように指示している．つまり，記事によって与えられた情報はそのトピックを理解するためにのみ用い，その後，答を既に知っているかどうかとは関係なく，自分のレポートにとって必要な情報を訊ねる質問を作成するようにさせた．

今回は，2 年分の記事（1998, 1999 年の毎日新聞）から 60 のトピックを選んだ．30 人の被験者に協力してもらい，ひとりの被験者について，それぞれのトピック提示パタン毎に 10 トピックについて質問を作成させた．作成する質問数は 1 トピックあたり約 10 問を目安とした．

収集した質問のうち，40 トピックについて分析を行った[*1]．また，作成した質問全体をチェックし，意見を求める等趣旨に反した質問を一定数以上している被験者の作った系列をすべて除き，与えられたトピックについて両極端の知識の状態である被験者，つまり，事後に行ったアンケートでそのトピックについて非常に詳しい，あるいはそれについて添付された記事を読んでも理解できなかったと解答した人が作ったそのトピックに関する質問の系列を除いた．その後，各トピック各提示パタンについて 3 系列ずつを無作為に選択した．つまり，40 トピックについて各 9 系列の質問を解析の対象とした．質問の総数は 3,014 であった．選ばれたトピックは人物 6 件，組織 3 件，人工物 9 件，事件出来事 19 件，動物等 3 件であった．このうち 5 件はビール大手 3 社，同時多発テロ，祭礼行事のような集合的なものである．ハイブリッド車やニホンカワウソ等，個体ではなくクラスもトピックとなっている．新聞記事から選択したため，人物や組織など出来事以外のトピックでも，何らかの形で新聞記事となるような出来事に関連している場合が多い．例えば，ジョージ・マロリーは彼の遺体がその年に発見されている．トピックの例を図 1 に示す．

指揮者・小沢征爾のプロフィールと９８、９９年の動向
登山家・ジョージ・マロリー
ビール大手３社の９８、９９年の動向
米大手製薬会社ファイザー
ヤンキースタジアム
家庭用ゲーム機「ドリームキャスト」
ハイブリッド車
ミュージカル「ライオンキング」
犬型ロボット「ＡＩＢＯ」
ＮＡＴＯによるユーゴ空爆での中国大使館の誤爆
テレビ朝日によるダイオキシン報道
室生寺五重塔の台風による被害
タンザニア、ケニアでの米国大使館同時爆破事件
お水取り
特別天然記念物・ニホンカワウソ
トキ保護センターで９９年５月に生まれたトキ

図 1: 与えたトピックの例

2.2 質問と解答のタイプ

質問の種類，質問が何をたずねているかを分類した結果を表 1 に示す．ここで，4W 質問は「小沢征爾氏は誰に師事しましたか」のように具体的な人名等を訊ねる質問で，「〜って誰ですか」「〜とは何ですか」という質問は定義・記述・説明を求める質問に分類している．レポートを構成する情報を訊ねる質問を収集した今回の状況では，「なぜ」を訊ねる質問（Why 質問）は少ない．加えて，説明や定義を求める質問も予想されたほど多くはない．これは，「小沢征爾って誰ですか」という質問が，例えば彼の誕生日や出身地を訊ねるような具体的な質問に展開されているためであると考えられる．

表 2 は予想される解答のタイプによる分類である．ここで，「一般名称」とは種の名称，機械や身体的部品の名称等

[*1] この収集と分析は，後述するタスク設計と並行的に行っており，60 トピックのうち 12 トピックはタスクの試験実施（Dry Run）の材料として用いた．その他の 8 トピックは，得られた質問のバラエティが少ない等の理由で解析を行わなかった．

表 1: 質問で訊ねている内容の分類

何を訊ねているか	
4W 質問（誰，何，どこ，いつ等） 数量に関する質問を含む	70.4%
Why 質問 （なぜ等，理由を訊ねるもの）	4.4%
How 質問 （どうやって等，手順や手法をを訊ねるもの）	9.8%
定義・説明・記述を訊ねるもの	15.5%

表 2: 予想される解答による分類

何で答えられるか	
数値や日付表現	27.8%
固有名称	20.6%
一般名称	9.5%
多分名称	16.7%
節・文・文章	25.3%

であり，統語的には複合名詞とほぼ重なる．「固有名称」には小説や映画のタイトルが含まれる．この分類は表 1 に示した分類と強く関連する．例えば Why 質問に解答するためには一般に節や文が必要となる．しかし一方で，4W 質問に分類された質問がすべて名称（固有名称あるいは一般名称）によって回答できるわけではない．例えば，同じ場所を訊ねる質問でも「シェイクスピアの出身地はどこですか」は固有名称で回答可能であるが「ロブスタが好んで住むのはどこですか」には名称での回答は困難で，一定の量の記述や説明を必要とする．この分析は質問だけを見ることで行ったため，幾つかの質問については断定的な分類が行えなかった．「多分名称」と分類されたものは「AIBO の由来は何ですか」のような質問で，AIBO が何かのアクロニムであれば名称の範囲に収まるが，その由来が長い物語となるかもしれないものである．このように質問だけからは予想される解答が複数の分類にまたがるものは他でも存在するが，簡単のためにそれらは複雑な方に分類した．表 2 からわかるように，レポート作成のための質問のうち，58%–75% が数値や名称を解答とする質問となる．4W 質問とほぼ同じ割合の質問が名称を解答とすると分析されたのは，上のロブスタの例のように一部の 4W 質問が名称で解答できない一方で，AIBO の例のように定義を尋ねるもの等に名称で解答できる可能性のあるものが含まれたためである．

トピック提示のパタン，つまり質問作成者のそのトピックに関する知識の差による質問の種類の分布の違いを表 3 に示す．多くの知識を持っている被験者ほど，説明を求める等複雑な質問が多くなり，それに従って節や文で回答すべきものが増える弱い傾向が見られる[*2]．与えられたトピックの場合は具体的な質問に展開される「〜とは何ですか」のような質問が，周辺的な事物については漠然とした形のままで質問される場合があり，知識の増加と共にそのような周辺的な事物がより多く視野に入って来ることがこの傾向の原因ではないかという印象を受けている．飛行機事故がトピックの場合の「計器着陸装置（ＩＬＳ）とは何ですか」やトピックである製品を発売した会社についての「セガ・エンタープライゼスとはどのような会社ですか」がその例である．ただし，質問を眺めて受ける印象では，個人差の方が知識の差よりも大きい．知識の量と質問の内容や形式との関係について明らかにするためには，より精密な実験が必要であろう．

[*2] 質問の種類の分布は知識量によって有意に異なる．
($\chi^2_{(6)} = 14.12, p < 0.05$)

表 3: 知識の量と質問内容との関係

質問内容	知識（少）	知識（中）	知識（多）
4W 質問	722(73.7%)	699(70.2%)	700(67.4%)
Why 質問	36(3.7%)	42(4.2%)	55(5.3%)
How 質問	95(9.7%)	100(10.0%)	99(9.5%)
定義説明等	126(12.9%)	155(15.6%)	185(17.8%)
総質問数	979	996	1039

2.3 語用論的特徴

前方照応のための手段を代名詞，ゼロ代名詞，定名詞句，省略の 4 つに分けて，その出現頻度を調べた．更に省略を除く参照表現の参照物について，それが出来事であるか，与えられたトピックつまり大域的トピックであるかを分析した．出現頻度をまとめたものを表 4 に示す．括弧内はそのうち出来事を参照物とするものの数である．合計は質問総数である 3,014 を越えている．これは例えば，「それまで誰がその国の指導者だったのですか」のように複数の参照表現がひとつの質問文中に含まれる場合があるためである．このような複数の参照表現を含む質問は 50 質問 (17%) であった．

省略を除く 1,915 の参照表現のうち，546 表現 (29%) は与えられたトピック，つまり大域的トピック以外を参照するものであった．しかも，そのうち 502 表現 (92%) は同じ質問文中に大域的トピックを参照する表現を持たない．このような質問の存在は質問系列中で焦点が推移しており，その時点の焦点が大域的トピックと一致していなかったことを示している．質問数にして 496 質問（総質問数の 16%）がこのような質問であった．

参照表現を含まない 1,135 質問のうち，329 質問 (29%) は系列の先頭である．それ以外のものには，焦点を代名詞化しないでそのまま質問中に表現する場合 (39%) とサブダイアログの先頭となる質問である場合 (32%) とがある．前者は，焦点である人物を姓のみで参照する場合や，個体ではなくクラス（ニホンカワウソやハイブリッド車）を焦点とする場合に多い．後者は，例えば，あるニュース番組

表 4: 収集した質問中の語用論的現象

分類	
参照表現を含まない質問	1135
代名詞	421 (136)
ゼロ代名詞	1230 (417)
定名詞句	264 (189)
省略	14

のダイオキシン汚染に関する誤報道をトピックとした場合に，その番組に対する一連の質問に続いて「ダイオキシンの毒性はどのくらいですか」と訊ねるような質問である．ただし，作成された質問系列の後半でこのような独立した質問が比較的多く観察されるので，それらが目安とした作成質問数にひきずられて，付け足し的に訊ねられた質問であるという可能性も否定できない．

3 情報アクセス対話のための評価タスク

本章では，前章で述べた特徴を有する情報アクセス対話での利用を想定した質問応答システムの能力を評価するタスクについて述べる．このタスクは対話文脈を考慮した適切な質問の解釈，つまり照応解消や省略処理等のいわゆる文脈処理に焦点を当てており，その定量的客観的評価を目的として設計された．本稿ではこのタスクの設計と経験論的研究で得られた対話の特徴との比較について述べる．本タスクは NTCIR4 Workshop[3] における質問応答技術の評価型ワークショップ QAC2 において Subtask3 として設計されたものである．評価手法に関する議論や本タスクの実施によって明らかとなった技術の現状については，参考文献 [7][8][9] に詳しい．

3.1 タスクの設計

QAC は日本語による質問応答技術の評価チャレンジであり，本稿で述べる Subtask3 を含めて 3 つのタスクからなる [4]．QAC では共通して，名称（固有名称と一般名称）や日付，数値を解答とするような事実に基づく（factoid）質問を対象としている．システムはこれらの名称をそれを含んだ部分でなく，過不足なく正確に抜き出すことを求められる．回答に利用される文書セットは新聞 2 紙各 2 年分の記事（QAC1 では 1 紙 2 年分）で，それを使って分野に依存しない質問に回答する．

従って，我々の関心のひとつは，このようなタイプの質問を扱う質問応答システムが情報アクセス対話の場面で利用できるかというものであった．前章で述べたように，レポート作成のための質問では，Why 型の質問，説明や定義を求める質問は共にそれ程多くはなく，全体のうち，58%–75%が名称や数値を解答とする質問であったことから，このような質問を扱う質問応答システムで情報アクセス対話を扱いうるということが明らかになった．また，この名称を解答とするであろう 75 %の質問のうち，737 問について調査したところ[*3]，人手で探して新聞記事から回答が得られたものは 84%であった．この点でも新聞記事を知識源とした質問応答システムと対話しながらレポートを作成するという設定が現実的であることがわかる．

Series 14
小沢征爾さんはいつ生まれましたか。
どこの生まれですか。
大学はどこを卒業しましたか。
師事した先生は誰でしたか。
誰に認められましたか。
９８年にはどこで指揮を行っていましたか。
２００２年からどこの指揮者になりますか。

Series 20
ジョージ・マロリーはどこの国で生まれましたか。
彼の有名な言葉は何ですか。
それを言ったのはいつのことですか。
彼が初めて山に登ったのは何歳の時ですか。
彼がエベレストの頂上付近で行方を絶ったのは何次遠征のときですか。
それは何年のことですか。
彼が最後に目撃されたのはエベレストの何メートル付近ですか。
彼の遺体を発見したのは誰ですか。

Series 22
ニューヨーク・ヤンキースの本拠地となっている球場はどこですか。
何年に造られたものですか。
そこには何人の記念碑が飾られていますか。
１９９９年に飾られたのは誰ですか。
彼が新婚旅行で来日したのは何年ですか。
その時の結婚相手は誰ですか。
彼女をポップ・アートに描いているのは誰ですか。
彼が描く缶詰はどこの会社のものですか。

図 2: 質問シリーズの例

本タスクでは，システムに一連の質問（シリーズと呼ぶ）を与え，それに次々と回答させてゆく．質問シリーズの例を図 2 に示す．この一連の質問とそれへの回答が，情報アクセス対話を模擬したものとなる．本タスクでは，複数のシリーズをバッチ的に与え，それに回答することをシステムに求める．対話の展開があらかじめ定められていることで対話本来のダイナミクスが失われているが，その一方で，相互比較可能な評価結果が得られるという利点に加え，テストセットも再利用可能となる．

質問シリーズについては，情報アクセス対話が，利用者があるトピックについてのレポートや要約を作成ために情報を収集する等の目的でそれに関する一連の質問を行なう

[*3] 解答が存在するかの確認はコストが高く，先の分析対象すべてについて行うことはできなかった．

ような対話（収集型）と利用者の興味の赴くところに従って対話の進行と共にトピックが変わっていくような対話（ブラウジング型）との２種類に分類されるという直観から，それに応じた２つの型を設定している．

あるシリーズの型は，そこに含まれている語用論的現象から定義される．収集型は，やや広い意味ではあるが，共通のトピックに関する質問からなり，そのトピックはシリーズ先頭の質問で導入される．収集型で特に狭義のものは，すべての参照表現（０代名詞を含む）がそのトピックを参照するもので，それ以外のものは含まれない．広義のものは，複数の参照表現を持ちその一方だけがトピックを参照するような質問や，トピックが関連した出来事を参照する表現を持つような質問も含まれている．図 2 に示した Series14 は狭義の収集型で，先頭質問で述べられている「小沢征爾」を補うことで，すべての質問の照応解消が行える．Series20 は広義の収集型での第 3 問は複数の参照表現を含み，第 6 問は先頭質問文で述べられているトピックであるジョージ・マロリーが関連したで出来事を参照している．ブラウジング型はそのような共通のトピックを持たず，質問中の参照表現は，直前の質問の解答や以前の質問中で言及された事物を参照している．series22 はブラウジング型である．

タスクにおいて，あるシリーズがどちらの型であるかは与えられず，システムはそれを自分で判定しなければならない．また，システムはある質問がシリーズの先頭であるという情報は利用してよいが，ある質問に回答する際にそれに続く質問を参照することは許されない．これは本タスクが対話的な状況でのシステムの利用を模擬していることからの制約である．

レポート作成を目的とした情報アクセス対話を前提とするのであれば，ブラウジング型は不要ではないかという考えもあるが，前章の分析で明らかになったように，特定のトピックに関する情報収集を目的とした質問系列でもその中では焦点の移動があり，そのような焦点の移動はブラウジング型でのみ起こりうる．その意味でブラウジング型は必要であり，本タスクのシリーズは実際の情報アクセス対話で現れる対話の一部を語用論的な特徴に着目して切り出したものとなっていると位置づけられる．シリーズの型の判定をシステムに行わせるのも，このような位置づけによる．

3.2 テストセットの構築と特徴

タスクの実施のためのテストセットは前章で述べた手法で収集した質問を用いて構築した．収集型のシリーズは，収集し分析した 40 トピックから 26 トピック（人物５件，組織２件，人工物５件，事件出来事 11 件，動物等３件）を選び，そのトピックに関する質問を適当に選び並べることで作成した．様々な参照表現を含めて質問文の表現については，できるだけ収集したものをそのまま用いることに努めたが，意味的，語用論的な曖昧性除去のために編集を行っている．

表 5: テストセットに現れる語用論的現象

分類	
参照表現を含まない質問	36
代名詞	76 (21)
ゼロ代名詞	134 (33)
定名詞句	11 (4)
省略	7

ブラウジング型のシリーズは，残りのトピック中の質問と複数文書要約に関するチャレンジである TSC[15] のために収集された評価用の質問文を種とし，そこからもしくはそこまでの適当な流れを付加することで作成した．例えば，Series22 は，ヤンキースタジアムをトピックに収集した質問（前半の４問．ただし最初の質問は焦点を変えてある）に後半を付け加えたものである．

今回のテストセットは 36 シリーズ，251 質問で，収集型 26 シリーズ（狭義のものは５シリーズ），ブラウジング型 10 シリーズである．１シリーズを構成する質問の平均数は 6.92 となっている．

テストセット中に現れた語用論的現象を表 5 に示す．形式は表 4 と同様で，括弧内は出来事を参照物とする場合の数である．現状の質問応答システムは質問をキーワードの並びとして理解している場合が多く，その場合，固有名称で表現されない出来事を参照物として認識することは固有名称で同定される人物等の場合と較べて困難であると予測される．そのような出来事を指示対象とする参照表現を含め，情報アクセス対話に現れる様々な語用論的現象がテストセットにも含まれている．

なお，表からは明らかでないが，テストセットと前章で分析した質問系列には大きな違いがある．テストセット中では参照表現（０代名詞を含む）を含まない質問文はシリーズの先頭のみである．これは固有名詞を使って現在の焦点を参照するようなことはせず，参照表現を積極的に利用して，更にサブダイアログを含まないような場合に相当する．前章の分析が示しているように情報アクセス対話における質問の系列はその焦点が対話の進行によって推移しサブダイアログを含む複雑なものであるから，この点ではテストセットの特徴は観察された現象と異なるが，タスクとしてこの程度の制約は必要と判断した．

3.3 TREC 2001 Context Task との関連

TREC 2001 では，質問応答システムの文脈追跡（文脈処理）能力を測定するために一連の質問に解答させるという Context Task を設計し実施している [13]．その目的は本稿で提案したタスクと同じである．このタスクの実施では，システムがある質問に正解できるかがそれ以前の質問に正解したかに依存しないという予想に反する結果が得られた．これは最初の質問によってそのシリーズの質問すべ

ての解答を含んだ少数の記事が同定されてしまい，その後の質問に正解できるかは文脈処理の能力よりも特定のタイプの質問に解答できるかに依存してしまうためであるとされている．このため，このようなタスクは現状では文脈処理能力を測定するのに不適切と判断され，その後の TREC では実施されていない．

このような結果となった理由はひとつのシリーズを構成する質問の数が 3 から 4 と少ないこと，本稿でいうところのブラウジング型を含んでいないことにあると思われる．ブラウジング型の場合，文脈追跡の必要性は自明である．ニューヨーク・ヤンキーズからキャンベルスープまでを含んだ記事はありえないので，最初の質問に関する処理だけでその後の質問に正解できる記事集合が得られることはありえない．収集型についても，それに関する記事が比較的多いようなトピックを選んでいることや広義の収集型の存在により，そのトピックに関する記事すべてを検索してもそこから正しく回答を選択することは，何らかの文脈処理なしでは困難である．一例として，狭義の収集型であっても「小沢征爾」をキーワードとする記事は知識源中に 155 件あり，そのうちの 22 件が彼のウィーンフィルへの移籍を扱っているが，その中で彼の誕生日に言及しているものは 2 件のみである．しかも本タスクではこれら 2 種類のシリーズが混在しており，システムはその型を判断しなければならない．このような複雑さは，隣り合う質問の解答のうち 85%が同じパラグラフに存在したという TREC の状況 [5] と本質的に異なる

加えて，このような特徴がレポート作成を目的とした情報アクセス対話という場面設定から自然に得られたことも重要である．対して，TREC Context Task にはそのような現実性のある裏付けが存在しない．

4 おわりに

本稿では，情報アクセスのための対話場面で質問応答システムが利用可能かについての経験論的研究を行い，名称を解答の範囲とするような質問応答システムが充分有効である一方で，様々な語用論的現象を処理できる必要があることを明らかにした．更に，情報アクセス対話という場面での質問応答システムの能力を評価するタスクを提案し，その設計とテストセットについて述べ，実際に起こりうると思われる対話とを比較した．

今後の課題として，今回の研究では明確な結果が出なかった知識の量と質問の内容や形式との関係について，より精密な実験を考えていきたい．タスク設計に関しては，ブラウジング型シリーズの構築の方法論を確立し，収集型と同じような現実性のある裏付けを与えたい．また，本稿では省略したが，評価尺度についても対話的場面での利用という設定に起因する難しさがあり，その点についても検討を続けたい．

謝辞

議論に参加していただき，貴重なコメントをいただいた QAC2 Subtask3 参加者の皆様に感謝いたします．本研究の一部は国立情報学研究所との共同研究として支援されています．ここに感謝します．

参考文献

[1] Joyce Y. Chai and Rong Jin. 2004. Discource Structure for Context Question Answering. *Proceedings of HLT-NAACL2004 Workshop on Pragmatics of Question Answering*, pp. 23-30.

[2] John Burger, Claire Cardie, and et al. 2001. Issues, Tasks and Program Structures to Roadmap Research in Question & Answering (Q&A) `http://www-nlpir.nist.gov/projrcts/duc/roadmapping.html`.

[3] NTCIR4 Workshop Home Page. 2003. `http://research.nii.ac.jp/ntcir/workshop/work-en.html`.

[4] Jun'ichi Fukumoto, Tsuneaki Kato and Fumito Masui. 2003. Question Answering Challenge(QAC-1) An Evaluation of question answering tasks at the NTCIR workshop 3. *AAAI 2003 Spring Symposium New Directions in Question Answering*, pp. 122-133.

[5] Sanda Harabagiu, Dan Moldovan, and et al. 2001. Answering complex, list and context questions with LCC's Question-Answering Server. *Proceedings of TREC 2001*.

[6] Eduard Hovy. 2001. `http://www-nlpir.nist.gov/projects/duc/pubs/2001papers/isi_hovy_duc.pdf`.

[7] 加藤恒昭, 福本淳一, 桝井文人, 神門典子. 2004. 質問応答から対話理解へ – NTCIR QAC Task3 の提案 –. 言語処理学会第 10 回年次大会, D2-7, pp. 317-320.

[8] Tsuneaki Kato, Jun'ichi Fukumoto, and et al. 2004. Handling Information Access Dialogue through QA Technologies – A novel challenge for open-domain question answering –. *Proceedings of HLT-NAACL2004 Workshop on Pragmatics of Question Answering*, pp. 70-77.

[9] Tsuneaki Kato Jun'ichi Fukumoto, and Fumito Masui. 2004. Question Answering Challenge for Information Access Dialogue – Overview of NTCIR4 QAC2 Subtask3 –. *Working notes on the Fourth NTCIR Workshop Meeting*, pp. 291-296.

[10] Inderjeet Mani, David House, and et al. 1998. The TIPSER SUMMAC text summarization evaluation final report. Technical Report MTR98W0000138, The MITRE Corporation.

[11] Sharon Small, Nobuyuki Shimizu, and et al. 2003. HITIQA: A Data Driven Approach to Interactive Question Answering: A Preliminary Report. *AAAI 2003 Spring Symposium New Directions in Question Answering*, pp. 94-104

[12] Ellen M. Voorhees and Dawn M. Tice. 2000. Building a Question Answering Test Collection. *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 200 - 207.

[13] Ellen M. Voorhees. 2001. Overview of the TREC 2001 Question Answering Track. *Proceedings of TREC 2001*.

[14] TREC Home Page. 2003. `http://trec.nist.gov/`.

[15] Text Summarization Challenge Home Page. 2003. `http://lr-www.pi.titech.ac.jp/tsc/index-en.html`.