機械翻訳と日本語質問応答システムを利用した日英言語横断質問応答システム

川岸 将実 † 森 辰則 ‡
† 横浜国立大学 大学院 環境情報学府 ‡ 横浜国立大学 大学院 環境情報研究院
E-mail: {kawagisi,mori}@forest.eis.ynu.ac.jp

1 はじめに

近年，文書情報に対するアクセス技術として，質問応答が注目されている．質問応答は，利用者が与えた自然言語の質問文に対し，その答を知識源となる大量の文書集合から見つける技術である．

知識源となる文書は種々の言語で書かれているため，単言語だけではなく言語を横断するような検索技術が求められてきている．

言語横断質問応答は，利用者の与える質問文に対し，その答を言語の異なる文書集合から見つけるタスクである．本稿で扱う JE-Task は，日本語の質問文に対し英語の知識源から答えを抽出するものである．言語横断質問応答に関する評価型ワークショップ NTCIR5 CLQA-1[6] においては，答は原文中の表現 (英語) のまま提出することが求められた．

本稿では，上記の背景の下，日英言語横断質問応答の一手法について提案する．本手法では，機械翻訳と既存の日本語依存の質問応答手法を組み合わせる．その中で，文書検索の手がかりとなる語（以下，キーワードと呼ぶ）の対訳方法と，機械翻訳と知識源中の語の表現の差異について考慮し，最終的な解を決定する手法を提案する．

2 関連研究

CLEF[2] では，ヨーロッパで使用される言語間での言語横断質問応答タスクが行なわれており，フランス語など 6 言語での単言語質問応答と，フランス語 - 英語など 50 種の言語横断質問応答（実際に参加があったのは 13 種）の結果について報告されている [3].

言語横断質問応答では，質問文と知識源の言語が異なるためいずれかの段階で翻訳を行なう必要がある．翻訳には対訳辞書や機械翻訳を使うことが考えられるが，翻訳精度が全体の精度に大きく影響してくる．また，言語依存の質問応答を用いるか，言語非依存の質問応答を用いるかという問題もある．

言語依存の質問応答を流用したシステムとして，関根は英語とヒンディ語における言語横断質問応答について報告している [8]．このシステムでは既存の質問解析部分を流用しており，解決定には知識源側言語のシステムを用いているが，簡単な処理である程度の精度が得られているとしている．

佐々木は言語横断質問応答に対する新しいアプローチとして，機械学習を用いて，質問文の特徴，文書の特徴から質問文タイプを用いずに解答を抽出する手法を提案している [12]．この手法では質問文タイプを用いることをせず，かつ機械学習を用いて解を抽出するため，新たに知識源側の質問応答システムを用意する必要がない．しかし，この方法では問題と解を対応付けた大量のデータを用意しなくてはならない．

一方で，文書集合を翻訳するのは好ましくないという報告がある [1]．その理由は対象となる言語が複数となった場合，その組合せが膨大になってしまうためである．また，Web のような文書を対象とした場合に適当ではないということがあげられる．

それに対して，我々は日本語に依存した質問応答システムを開発してきており [4]，日本語質問応答システムを基本としたシステムを提案する．検索した英語文書を日本語に翻訳し，その文書を対象に日本語質問応答システムを用いる．この時翻訳は機械翻訳を用いて行なう．

キーワードの対訳語の多義性の解消については，言語横断情報検索分野で盛んに研究されていて，対になる知識源から対訳語を決定する手法 [5] や，機械翻訳により決定する手法 [7] などがある．しかしこれらは対訳語を拡張して検索を行なうことに主眼が置かれているので言語横断質問応答に適用できるか疑問が残る．それに対し，我々は EDR 対訳辞書の訳語種別を用いて対訳語を重みづけしているところが異なる．また，我々は予備実験の結果から，対訳語を拡張せず候補数を絞った．

3 提案手法

我々の日本語質問応答は，質問文解析，文書検索，パッセージ抽出，命題照合の 4 つのモジュールからなっている．質問文解析では，質問文の特徴となる語（以下，キーワードと呼ぶ）と質問が何について尋ねているのか（以下，質問文タイプと呼ぶ）を決定する．文書検索では，抽出されたキーワードを手がかりに質問に関連する文書を検索する．パッセージ抽出では，検索文書の中から，正解を含む可能性の高い連続する数文（パッセージ）を抽出する．そして命題照合において，質問文タイプ，質問文の解析結果をもとに解候補を決定する．この解候補は質問文解析の結果などからスコアリングを行ない，スコアの高い方から優先して出力される．このシステムでは，文書検索モジュール以外は日本語依存となっている．

日英言語横断質問応答とは，質問文が日本語で知識源となる文書が英語の場合の質問応答を指す．そのため，質問文解析においては既存の日本語質問応答システムのものをそのまま利用することが出来るが，文書検索以降では機械翻訳を行なう必要がある．

既存のシステムを使用するにあたり，以下のような手法が考えられる．

1. 知識源（英文）をあらかじめ機械翻訳しておく．
2. キーワードを翻訳し文書検索を行なう．
   (a) 検索された文書（英文）を翻訳し，パッセージ抽出以降は既存のシステムを使用する．

(b) パッセージ抽出までは英語で行ない，パッセージを翻訳する．命題照合は既存のシステムを使用する．

1 は，実際には使われない文書まであらかじめ翻訳する必要があるためコストが膨大なものになってしまう．また，外部の検索エンジンを利用せざるをえない Web 文書などを知識源とする場合に対応できない． 2 は，キーワードの翻訳が正しく行なわれないと知識源から必要な文書を取り出せないことがある．そのためキーワードの翻訳について検討する必要がある．

2(a) は検索された文書を全翻訳するため膨大な時間がかかる．しかし，パッセージ抽出以降で既存のシステムを使用できるというメリットがある． 2(b) はパッセージ抽出において日本語に依存した処理ができないというデメリットがあるが，パッセージのみを翻訳するため計算コストは小さくなるというメリットがある．

この他に，解提示の問題がある． 2005 年に行なわれた NTCIR5 の日英言語横断質問応答タスクにおいては，解答を知識源側の言語で提示する必要があった．しかし，1， 2 いずれの方法においても，命題照合は既存のシステムを使用するため日本語の解が抽出される．したがって，日本語で命題照合を行ない解候補を抽出した後に，解候補を翻訳する必要がある．予備実験の結果，翻訳された解候補と知識源となる文書において表現の差異があることがわかった．そのため，解候補と知識源の表現の差異に対応する必要がある．

本稿では 2 の方式を用いた日英言語横断質問応答システムを提案し，(a)(b) のどちらの方式がより適切であるかを比較検討する．また，解対応について検討を行ない実験により評価する．システムの概略を図 1に示す．

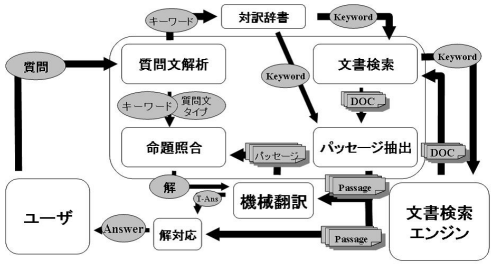

図 1: 提案する英日言語横断質問応答システム

質問文解析までは，既存の日本語質問応答システムと同様である．抽出された日本語のキーワードは対訳辞書を用いて英語のキーワードへと変換される．

(a) では英語のキーワードを文書検索に，日本語のキーワードをパッセージ抽出と命題照合にそれぞれ渡す．英語のキーワードをもとに検索されてきた英語の文書を日本語に翻訳し，日本語のキーワードと質問文解析結果を手がかりにパッセージ抽出，命題照合を行なう．

(b) では英語のキーワードを文書検索とパッセージ抽出に，日本語のキーワードを命題照合に渡す．英語のキーワードにより検索されてきた文書からパッセージを抽出する．命題照合では日本語のキーワードと質問文解析結果を手がかりに解候補を抽出する．

命題照合により抽出された日本語解候補は機械翻訳を用いて英語の解候補へと変換される．この解候補と原文書を比較して最終的な解候補を決定する．

なお，キーワードの翻訳に関しては節 3.1で，解対応に関しては節 3.2で詳しく述べる．

3.1 キーワードの翻訳

質問文解析で抽出されたキーワードは日本語のため，そのままでは英語の知識源の文書を検索してくることができない．そのため，英語のキーワードへと翻訳する必要がある．言語横断情報検索の分野で議論されているように，一般に対訳語は一つの語に対して数種類あり，適切に対訳語を決定する必要がある．本論文では，対訳語決定のために利用する語の重みづけの一手法を提案する．

対訳語候補の選定には EDR 日英対訳辞書を用いる．また， EDR 日英対訳辞書には語の概念と対訳語の種別があり，それを利用して訳語候補に対する重み (重要度) を以下のように決定する．

- 同義語：日本語と英語の意味が同一であるもの．重みは 1.0．対訳語が文脈によらずに同一の意味を持つ場合 0.2 を加算する．（例: 犬 -dog）
- パラフレーズ：意味は同一ではないが，同じような意味に取れるもの．いいまわし．重みは 0.9．（例: 犬 -spy）
- 逐語訳：日本語を英語に直訳した語．重みは 0.7
- ローマ字表記：日本語をローマ字に直した語．重みは 0.65
- 説明文：対応する英語がない場合に，日本語の意味を英語で説明したもの．重みは 0.3

重み付けされた語の上位 N 語を使用して文書検索を行なう．本稿では $N = 1$ とした．これは予備実験において，対訳語を増やした方が精度が悪かったためである．この訳語決定法を訳語決定法 1 とする．

訳語決定法 1 では語の種別のみを扱っているため，生起頻度などは一切考慮していない．そこで，知識源中の単語の生起確率を導入した重みを提案する．この決定法を訳語決定法 2 とする．重みの式は次式のようになる．

$$W(w_n) = TW_n \frac{\alpha_n}{\sum_i \alpha_i + \beta} (\alpha_i \neq 0 \text{の場合})$$

$$W(w_n) = TW_n \frac{\beta}{\sum_i \alpha_i + \beta} \frac{1}{\beta_{num}} (\alpha_i = 0 \text{の場合})$$

$$W(w_n) = TW_n \gamma (\alpha_i \in stopword \text{の場合})$$

ここで W は重み， w_n は n 番めの対訳語候補， TW_n は訳語決定法 1 による w_n の重み， α_n は知識源中における w_n の出現頻度， β_{num} は対訳語候補中で $\alpha_n = 0$ の w_n の個数である． $stopword$ は and などの一般的によく使われる語の集合である． $stopword$ に含まれる語が選択された場合は，文書検索におけるキーワードには使用されないが，パッセージ抽出と命題照合においては使用される． β は知識源中に w_n が出現しない場合に，生起確率を 0 にしないための定数であり， γ は候補 w_n が $stopword$ だった場合の重みとして用いる定数である．

3.2 解対応

我々のシステムでは，文章全体あるいはパッセージを日本語に翻訳することで質問応答を行なうため，抽出さ

れる解は日本語となる．一方で，NTCIR-5 CLQA1 JE-Taskでは解を知識源の言語で提示する必要があったため，その解を英語へと翻訳することを考えた．この時に，翻訳した解と知識源で，同一の物を指しているのに，表層が異なるために正解とは認められない可能性がある．そこで，機械翻訳による表現を知識源の表現に対応させることを提案する．

解を対応させる主な規則として，次の三点を挙げる．

- 単位表現を知識源中の表現に統一する．
- 数値の前に `about` や `nearly` などの表現がある場合は，それも解の一部とみなす．
- 表層の異なる表現があった場合は，知識源中の表現に統一する．

具体的な方法としては，翻訳された解を基に，知識源から取り出したパッセージ（あるいは文書全体）にタグを付ける．そのタグを基に最終的な解候補を決定する．タグ付けと解候補決定の手順を以下に示す．

1. 翻訳された解答を単語ごとに分割する．
2. 得られた各単語をパッセージ中の各単語と照合し，以下の条件を満たした時にパッセージ中の単語にタグを付与する．
   (a) 翻訳された解答から得られた単語が3文字以下で，パッセージ中の単語と完全に一致した場合
   (b) 翻訳された解答から得られた単語が4文字以上で，パッセージ中の単語と前方一致した場合
3. 以下の条件を満たすものを解候補として決定する．ただし，複数個候補がある場合は，最長のものを解とする．
   (a) 翻訳された解答の単語数と，連続してタグづけされた単語数が一致した場合，そのまとまりを解とする（以下，解対応法1と呼ぶ）．
   (b) 翻訳された解答の単語数を W_{num}，タグ付けされた単語から単語までの単語数（タグ付けされた両端の単語を含む．また，間にタグ付けされていない単語が含まれていてもよい）を D_{tag} とする時，以下の式を満たす場合，そのまとまりを解とする（以下，解対応法2と呼ぶ）．
   $$W_{num} \leq D_{tag} \leq W_{num} + 1$$
   また，数値にタグ付けされている場合，直前の単語が `about` などの表現の場合それも解の一部とみなす．

これにより，単数，複数形の問題や，解の表現の違いに対処できると考える．

4 評価実験

4.1 使用データ

NTCIR5 CLQA1 JE-Taskのサンプルテストセット300問を開発用として使用した．評価用にはNTCIR5 CLQA1 JE-Taskで使用された200問を使用した．

また，知識源はDairy Yomiuriの2000年，2001年の記事2年分，機械翻訳は市販の翻訳ソフトウェア[10]を，対訳辞書としてはEDR対訳辞書[11]を使用した．

4.2 システム性能評価

NTCIR5 CLQA1 JE-Taskの200問を使用しシステムの性能を評価した．結果を表1に示す．

表 1: システム性能評価

	Acc	MRR	Acc+U	MRR+U
JEQA-1	0.030	0.046	0.030	0.054
JEQA-2	0.085	0.115	0.090	0.128
JEQA-3	0.080	0.110	0.085	0.123
JEQA-4	0.045	0.069	0.060	0.092

Acc:Accuracyによる評価
MRR:MRR値による評価
+U:Unsupported(解一致，文書ID相違)
JEQA-1: 解翻訳のみ，訳語決定法1
JEQA-2: 解対応法1，訳語決定法1
JEQA-3: 解対応法2，訳語決定法1
JEQA-4: 解対応法2，訳語決定法2

ここで，JEQA-1は解を翻訳したのみのもの，JEQA-2は解対応法1を使用したもの，JEQA-3は解対応法2を使用したものであり，この3つは訳語種別のみの訳語決定法1を用いている．JEQA-4は解対応法2を使用し，生起確率と訳語種別を組み合わせた訳語決定法2を用いている．また，Accは Accuracy[1]による評価，MRRは MRR値による評価[2]，+Uは Unsupported (解は一致だが，解の根拠となる文書IDが違うもの) を加えたものである．

開発用テストセットにおける評価についてはJEQA-4と同じ設定の場合の精度が最も高かったが，上記の結果によれば，最も良かったのはJEQA-2であった．このことより，問題によっては解の拡張や，訳語決定法2が有効に働かないこともあることがわかった．しかし，訳語決定法2を使用することで正解が得られるようになった問もあり，単純に精度が落ちてしまうと結論づけることはできない．

4.3 NTCIR5 CLQA-1による評価

我々は本提案システムによりNTCIR5 CLQA-1に参加し評価を行なった．その結果を図2にしめす．

我々の参加システムはJEQA-1～3であり，節4.2のシステムに対応している．また，システムA-Eは他の参加システムである．

最も精度の良かったシステムAは，質問応答部に英語に依存したものを使用しており，システムB-Dはベースラインで，言語非依存の質問応答システムを用いたものである．システムEは日本語依存の質問応答を用いたものである．

結果を見ると質問応答部が日本語依存のシステムは，英語依存のものより精度は低く，日本語依存の質問応答を利用することは必ずしも適切とはいえない．しかし，

[1] システムは質問に対して解答を一つだけ提示し，全質問に対する正解数の割合を求めるもの
[2] Mean Reciprocal Rank．各問について最上位正解の順位の逆数を評価値とし，それを平均したもの．

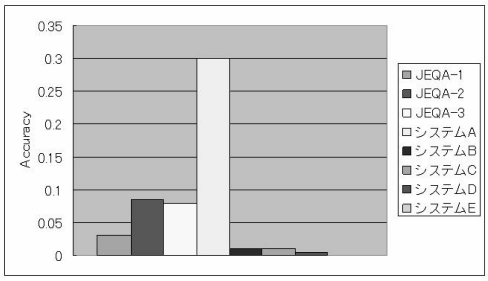

図 2: NTCIR-5 CLQA1 JE-Task の結果

日本語依存の質問応答を用いてもある程度の精度が得られることが分かった点については興味深い．

5 考察

ここでは，システムの問題点などを考察する．
失敗を解析すると，以下の種類に分けられる．

1. 日本語依存の命題照合において正解が取れていない．
2. キーワードの対訳がうまくいっていないため，関連文書を検索できていない．
3. パッセージの翻訳がうまくいっていないため，日本語依存の命題照合において正解が取れない．
4. 解対応において，知識源から正しい解を抜き出せていない．

1 の解決には日本語 QA システムの精度向上が望まれる．

2 は，「シーガイアはどこにあるのでしょうか？」のキーワード「シーガイア」のように，対訳辞書に対訳語が存在しない場合に多く見られた．このような例は固有名詞に多く，人名やビル名などの対訳に見られた．この対処法としては， Web などの情報源を用い，動的に対訳語を探す方法などが考えられる．

また，辞書中に適切な対訳語は存在するが，その対訳語が適切に選択されていない場合もあった．この解決法の一つとして，共起確率を用いた制約伝播法 [9] により対訳語を決定する方法を実験したが， Accuracy 値で 0.050 と訳語決定法 1 の結果に及ばなかった．訳語決定法については今後さらなる検討をしていく予定である．

3 は，誤訳や未知語が残ってしまうことにより，日本語依存の命題照合において，正しい構文情報が取得できなかったり，解として認識できないなどの問題が生じてしまう．この問題を解決するには，機械翻訳の精度に依存しないような方法が有効と思われる．例えば，英語のまま命題照合を行なうことや，未知語の補完を Web などを用いて行なうことが考えられる．

4 は，日本語で抽出した正解を英語に翻訳する時の誤訳に主な原因がある．例えば，解を翻訳したときに知識源中の正解より多い単語数で訳してしまい解が取れないもの，冗長な解を取ってしまうものがある．この問題の対処法としては， 3 と同様に英語のまま命題照合を行なうことや，解対応の条件をより細かく設定することなどが挙げられる．

6 おわりに

本稿では，機械翻訳と日本語質問応答を利用した日英言語横断質問応答の構築に関し，二つの手法を提案した．

第 1 に，解を決定する際に，日本語の解を翻訳したものを提示するのではなく，知識源中の表現と照らし合わせる手法を提案した．原文から解を決定する手法は全体的に有効であるが，その一方で機械翻訳の誤訳や表現の差異に対してはさらなる検討が必要であることもわかった．

第 2 にキーワードの翻訳において，重み付けを行ない，対訳語を決定する手法を提案した．二つの訳語決定法のうち，開発用テストセットを用いた実験では頻度情報を用いた決定法の方が精度が高い結果となったが，評価用テストセットにおいては対訳語種別のみを用いた決定法が高い結果となった．そのため，対訳語決定法に関してはさらなる検討が必要なことがわかった．

今後は，時間に関する質問に対して正しく答を導出する手法，及び，正解より少ない単語数となった場合の解対応について検討していく予定である．

参考文献

[1] Guillaume Bourdil, Faza Elkateb, Brigitte Grau, Garbriel Illouz, Laura Monceaux, Isabelle Robba, and Anne Vilnat. How to answer in English to questions asked in French: by exploting results from several sources of information. In Working Notes for the CLEF 2004 Workshop, 9 2004.

[2] Cross Language Evaluation Forum. http://clef.iei.pi.cnr.it/.

[3] Bernardo Magnini, Alessandro Vallin, Christelle Ayache, Gregor Erbach, Anselmo Pe nas, Maarten de Rijke, Paulo Rocha, Kiril Simov, and Richard Sutcliffe. Overview of the CLEF 2004 Multilingul Question Answering Track. In Working Notes for the CLEF 2004 Workshop, 9 2004.

[4] Tatsunori Mori. Japanese Q/A system using A* search and its improvement: Yokohama national university at QAC2. In Working Notes of the Fourth NTCIR Workshop Meeting, pp. 345–352, 6 2004.

[5] Tetsuji Nakagawa and Mihoko Kitamura. NTCIR-4 CLIR Experiments at Oki. In Working Notes of the Fourth NTCIR Workshop Meeting, 6 2004.

[6] NTCIR5 CLQA-1. NTCIR Workshop5 言語横断質問応答タスク. http://www.slt.atr.jp/CLQA/, 2005.

[7] Tetsuya Sakai, Makoto Koyama, and Akira Kumano. Toshiba BRIDJE at NTCIR-4 CLIR: Monolingual/Bilingual IR and Flexible Feedback. In Working Notes of the Fourth NTCIR Workshop Meeting, 6 2004.

[8] 関根聡. 言語横断質問応答システム. 言語処理学会 第 10 回年次大会 発表論文集, pp. 321–324, 3 月 2004.

[9] 麻野間直樹, 中岩清巳. 目的言語の単語共起情報を利用した訳語選択と未知語 の検出. 言語処理学会 第 5 回年次大会 発表論文集, pp. 442–445, 3 月 1999.

[10] 日本アイ・ビー・エム株式会社. 翻訳の王様バイリンガル Version5, 2002.

[11] 日本電子化辞書研究所. EDR 電子化辞書, 2001.

[12] 佐々木裕. 統合学習による質問応答システムの新しい構成法 ～ＣＬＱＡに向けて. 自然言語処理研究会報告 2004-NL-163, 情報処理学会, 2004.