

Cross-language 情報検索のための BMIR-J2 を用いた一考察

酒井 哲也 梶浦 正浩 住田 一男

{tetsuya.sakai,masahiro.kajiura,kazuo.sumita}@toshiba.co.jp

(株) 東芝 研究開発センター

本論文では、日本語テストコレクション BMIR-J2 およびこれを英訳したデータを用い、情報フィルタリングシステム NEAT と機械翻訳システム ASTRANSAC による日・英間の cross-language 情報検索の検索精度評価実験を行う。英語検索要求による日本語文書の検索実験では、文書の翻訳と検索要求の翻訳のアプローチを、さらに異なる翻訳者による検索要求の翻訳を比較する。日本語検索要求による擬似英語文書の検索実験では、検索要求の翻訳の前後にローカルフィードバックを行う。以上により、日本語単言語検索の 90% 以上の精度を実現する。

A Study on Cross-language Information Retrieval using BMIR-J2

Tetsuya SAKAI Masahiro KAJIURA Kazuo SUMITA

Toshiba R&D Center

We study a cross-language IR approach using the NEAT information filtering system and the ASTRANSAC machine translation system. The BMIR-J2 standard Japanese test collection and our own translated data are used for evaluation. In the *English-to-Japanese* experiments, we consider both document translation and query translation, and also compare the retrieval performance when the queries are translated by different translators. In the *Japanese-to-pseudo-English* experiments, we perform local feedback both before and after query translation. We achieve over 90% of Japanese monolingual performance.

1 はじめに

近年のインターネットの普及を背景に、世界中のユーザーが自分の使いやすい言語で検索要求を表現し、これに適合する情報を世界中の多様な言語で表現された情報データベースから探し出すという多言語情報検索 (multilingual information retrieval: MLIR) の環境が現実的なものとなりつつある。本論文では、MLIR のサブタスクとして、言語 L1 で記述された検索要求を満たす言語 L2 (\neq L1) で記述された情報を検索するタスクと定義される cross-language IR (CLIR) ¹ を扱う。CLIR は、例えば以下の場合に有用である [23]。

- 母国語が L1 であるユーザー A が、言語 L2 で記述された情報データベースから情報収集しようとしている。

¹ 同義語として crosslingual IR, cross-linguistic IR, translanguagual IR があるが、ここでは [16] の経緯に従う。なお、MLIR は CLIR の他にも、言語 L1 で記述された情報と言語 L2 で記述された情報が混在する場合や、各情報中に言語 L1, L2 が混在する場合なども含むより広い概念であると考えられる。また、英語以外の単言語検索を MLIR と称した事例もある [9][23]。

うとしている。A 氏は、検索要求を言語 L2 により自分でうまく記述する自信はないが、言語 L2 で記述された情報を眺めて必要な情報を得る程度の語学力は有している。

- 母国語が L1 であるユーザー B が、言語 L2 で記述された情報データベースから情報収集しようとしている。B 氏は、言語 L2 に関する知識を殆どもたないため、検索された情報をシステムが選択的に言語 L1 に翻訳してから提示する。

このようなニーズから、欧米ではここ数年 CLIR の研究が盛んである。例えば、米国規格協会 (NIST) が 1992 年から開催している情報検索システムの評価会 TREC (Text REtrieval Conference) では、1997 年の TREC-6 から欧米語の検索要求および文書を扱う CLIR がタスクに加わった [23]。これに習って、日本でも学術情報センターが日本語検索要求と英語文書を扱う CLIR のタスクを含む評価会 NTCIR を 1999 年 9 月に開催する予定である [15]。

CLIR では、以下のようなリソースが利用される。

- (a) 機械可読辞書 (machine-readable dictionaries: MRD) やシソーラス

(b) parallel/comparable/unlinked/monolingual corpora²

(c) 機械翻訳システム (MT)

また、CLIR における検索要求と検索対象の関連性を求める手段、換言すれば CLIR を単言語検索 (monolingual IR: MIR) に帰着させる手段としては以下が考えられる。

- (i) 検索要求を翻訳する
- (ii) 検索対象を翻訳する
- (iii) 言語に依存しない中間言語表現を用いる

CLIR における様々なアプローチにはそれぞれ一長一短がある。例えば、多言語かつ多方向の CLIR を扱えるという長所をもつ (iii) は、一般に (b) の入手可能性に依存する。また、入手可能性が比較的高い (a) を利用する場合も、多義語の曖昧性解消のために (b) を併用せざるを得ない場合が多い³。一方、(c) はもともと曖昧性解消の機構を有しているが、多言語および多方向への拡張のために新たな開発労力を要する。現状では、(a)(b) と (i)(iii) の組合せが主流といえる [1] [2] [3] [4][6][9] [11][12] [14] [18] [24] [26]。しかし、例えば TREC-6 の比較評価では (c) の有効性が再認識されており、さらに (c) のもとでは (i) よりも (ii) のほうが有効であることが示唆されている [17][23]。(ii) は翻訳コストの観点から敬遠されることが多いが、例えばユーザーの母国語が既知である場合に、前処理の段階で検索対象を MT にかけておくことにより高精度な CLIR を実現するといった応用可能性がある。

本論文では、情報フィルタリングシステム NEAT [20][22] および機械翻訳システム ASTRANSAC [7] [13] を用い、日本語と英語を扱う CLIR のうち (c) と (i)(ii) の可能性を探るための検索精度評価実験を行う。評価の題材としては、日本語情報検索システム評価用テストコレクション BMIR-J2 [21] およびこれを独自に英訳したものを用いる。2章で本研究の位置づけについて述べ、3章で CLIR の評価のベースラインとなる日本語単言語検索実験について、4章で英語検索要求による日本語文書の検索実験について、5章で日本語検索要求による英語文書の検索実験について述べ、最後に6章でまとめを述べる。

²制約が強い順、したがって実際の入手可能性が低い順に列挙してある。parallel corpora は文書とそれを直接翻訳したものが対応付けられているものであり、comparable corpora は同一トピックについての記述を含む複数言語の文書が対応付けられたものである。対応付けの単位には文書・文・語レベルがあり、対応関係の定義も様々である [14][24]。これに対し unlinked corpora は対応付けのされていない複数言語の monolingual corpora の集合である。

³逆に (b) からスタートして多言語ソーラスなどを構築し、これを CLIR に利用する事例もある [14]。

2 本研究の位置づけ

前節で述べたように、検索要求あるいは検索対象を翻訳するために MT を利用した事例は少ない⁴。さらに、MRD やコーパス利用の研究の中には「現状の MT の性能では高精度な CLIR は期待できない」という主張さえしばしば見られる [9][18][26]。一方、MT を積極的に利用している研究者は「現状の MT は人が読んで満足するレベルには達していないが、CLIR には十分有用である」と主張する [5]。我々もこの立場に立ち、実際に MT の利用により高精度な CLIR が実現できることを示す。

本論文は以下の4つの特徴をもつ。

- (1) MT を用いた日英間の CLIR を扱い、かつその検索精度をテストコレクションを用いて定量的に評価している。

MT を用いた CLIR で日本語を扱い標準的な検索精度評価を行った従来研究は、我々の知るかぎり英語検索要求と日本語文書を扱った我々自身の研究 [10] のみである。

- (2) 日英間の双方向の CLIR を扱っている。

TREC-6 では欧米語間の双方向 CLIR を扱っている例が見られるが [3][6]、日本語を扱った CLIR では従来このような試みは見られなかった。ただし日本には CLIR 評価用テストコレクションが存在しないため⁵、本論文では、日本語コレクション BMIR-J2 の全文書を MT で翻訳することにより、日本語検索要求と擬似英語文書を扱う CLIR の実験を可能とした (本論文ではこれを Je-CLIR と表記する。小文字の “e” は、純粋な自然言語ではなく日英 MT の出力であることを表している)。さらに、BMIR-J2 の検索要求を手で英訳することにより、英語検索要求と日本語文書を扱った CLIR の実験を可能にした (これを EJ-CLIR と表記する)。

- (3) EJ-CLIR において、検索要求を翻訳した場合と文書を翻訳した場合を比較検討している。

Oard らは、TREC-6 の英語検索要求とドイツ語文書を扱った CLIR において検索要求を MT にかけた場合と文書を MT にかけた場合の性能比較を行っており、一般には後者のアプローチのほうが多くの文脈を活用できることから高精度を実現する可能性が高いことを示唆している [17]。日本語を扱う CLIR においてはこのような比較はこれまでに報告されていない。

⁴CLIR の検索結果の提示のために MT を利用する試み [26] や、parallel/comparable corpora の代わりに monolingual corpora と MT を併用する試み [3][14] などは除く。

⁵1998年11月には、学術情報センターにより NTCIR 参加者のみに対して英日の大規模テストコレクション(テスト版)がリリースされ、これにより日本語検索要求と英語文書を扱った CLIR の実験が可能となった。この評価実験については別途報告予定である。なお、日本語を扱う CLIR においてテストコレクションを用いた標準的な評価を行っている従来研究は、コーパス利用の学術情報センター自身による研究 [11] のみである。

- (4) EJ-CLIRにおいて、複数の翻訳者が日本語検索要求を英訳した場合を比較検討している。

通常、言語 L_1 の検索要求と言語 L_2 の検索対象を扱う CLIR の実験は以下の手順で行われる。

1. 言語 L_2 の検索要求 Q_2 および検索対象 D_2 を用いて MIR の実験を行う。
2. Q_2 を人手により言語 L_1 の検索要求 Q_1 に翻訳する。
3. Q_1 と D_2 を用いた CLIR の実験により、MIR の場合の何%の精度が達成できるかを示す⁶。

しかし、CLIR の検索精度は人手による検索要求の翻訳結果に大きく依存すると考えられる。そこで本論文では、二人のバイリンガルにより検索要求を英訳し、それぞれをもとに日本語文書に対する CLIR の実験を行う⁷。このような試みは欧米でもこれまで報告されていない。

3 ベースライン: 日本語単言語検索 (J-MIR)

本章では CLIR のベースラインとなる日本語 MIR (以下、J-MIR) の実験について述べる。我々は文献 [20] において、確率検索モデルと local feedback (LF) を用いた高精度な J-MIR について報告した。本論文における J-MIR はこの実験を踏襲したものである。概要のみを説明する。なお、検索精度の評価尺度として 11 点平均適合率⁸ (11pt と表記する) を用いる。

1. BMIR-J2 の日本語検索要求 50 件 (Q_J と呼ぶ) を形態素解析し、名詞などを抽出して初期検索条件を生成する。
2. 式 (1) の検索語重み tw の総和に基づき NEAT により初期検索を行い、パラメータ K および b を最適化する。
3. LF を行う。すなわち、初期検索結果の上位 n 件を正解文書と見なし、この本文中から式 (2) の値の降順に新たな検索語 (展開語) を m 個選定して検索条件に付加する。
4. 再検索を行い n および m を最適化する⁹。

⁶このように、TREC では文書の言語を揃えて MIR と CLIR の比較を行う [23]。一方、NTCIR では、検索対象が parallel corpora に近いので、検索要求の言語を揃えて MIR と CLIR の比較を行う予定である [15]。

⁷検索要求の英訳を複数用意するアイデアは、Dr. Gareth Jones (University of Exeter) と Dr. Nigel Collier (東大) との東芝における共同研究の過程で得られた [10]。

⁸再現率 $r = 0.0, 0.1, \dots, 1.0$ の 11 点における適合率 p を平均した値 [25]。

⁹本論文では全ての実験において NEAT の text 条件のみを利用し、付加する検索条件の条件重み w は 0.2 で固定した [20][22]。

$$tw(t, d) =$$

$$\frac{\log(|C|/df(t)) * tf(t, d) * (K + 1)}{K * ((1 - b) + (b * L(d) * |C| / \sum_{d \in C} L(d))) + tf(t, d)} \quad (1)$$

$$rdf(t) *$$

$$\log \frac{(rdf(t) + 0.5) / (|R| - rdf(t) + 0.5)}{(df(t) - rdf(t) + 0.5) / (|C| - df(t) - |R| + rdf(t) + 0.5)} \quad (2)$$

ここで、 C : 検索対象となる文書集合、

R : 正解文書集合、

$L(d)$: 文書 d の文字数、

$tf(t, d)$: 文書 d 中における検索語 t の出現頻度、

$df(t)$: 文書集合 C の中で検索語 t を含む文書あるいは見出しの数、

K : $tf(t, d)$ の影響を調整するために経験的に決定される定数 ($0 \leq K$)、

b : $L(d)$ の影響を調整するために経験的に決定される定数 ($0 \leq b \leq 1$)、

$rdf(t)$: 検索語 t を含む正解文書の数。

今回の実験の文献 [20] との違いは、検索条件の展開 (query expansion) に加えて検索語の重み調整 (term reweighting) も行った点である。これは具体的には、式 (1) の \log の項を式 (2) の \log の項で置き換えることにより実現される [19]¹⁰。

表 1 に J-MIR の結果を示す。ここで、J- は J-MIR を意味し、INIT は K, b を最適化した初期検索結果を、LF は LF の n, m を検索要求全体について最適化した結果を表す。また GLF は、まず初期検索語数に応じて検索要求をグループ化し、表 2 のように各グループ毎に n, m の最適化を行ったものである¹¹。これは、LF によりどの程度文脈を補う必要があるかは、初期検索条件中にどの程度の情報が含まれているかに依存するという仮説に基づくものである。なお、符号検定 [8] で有意差が見られたのは J-INIT と J-GLF の間 ($\alpha = 0.01$) のみであった。

本論文では、CLIR 評価のベースラインとして J-INIT および J-GLF を用いる。なお、BMIR-J2 を用いた MIR においてこのような高精度を達成した例はこれまでに報告されていない [20]。なお、2 章で述べたように、EJ-CLIR のベースラインとして J-MIR を用いることは文書の言語を揃えた TREC 方式、Je-CLIR のベースラインとして J-MIR を用いることは検索要求の言語を揃えた NTCIR 方式であると言える。

¹⁰予備実験において、LF における query expansion のみの効果は初期検索精度の 5% 程度、これと term reweighting を併用した効果は 6% 程度であった。

¹¹実用上は訓練用の検索要求を用いて表を作成しておき、これをもとに未知の検索要求に対する LF を行うことになる。

表 1: J-MIR results.

名前	Q_J の説明	llpt
J-INIT	初期検索 ($K = 0.4, b = 0.5$)	0.463
J-LF	LF ($n = 4, m = 15$)	0.491
J-GLF	LF で n, m をグループ毎に最適化	0.500

表 2: GLF table for Q_J .

初期検索語数	1	2	3	4	5	6
擬似正解文書数 n	8	6	8	4	4	4
展開語数 m	15	5	15	20	30	5

4 英語検索要求による日本語文書の検索 (EJ-CLIR)

4.1 EJ-CLIR のための英語検索要求の作成

まず, Q_J をバイリンガルの日本人 X および Y が独自に英訳した¹². 各英語検索要求セットを Q_{EX} および Q_{EY} と呼ぶ. これらの自然言語文を英日 AS-TRANSAC により品詞タグ付けし¹³, 名詞などを抽出して初期検索条件を作成した. 表 3 の (1)(3)(4) に Q_J およびこれに対応する Q_{EX} と Q_{EY} の例を示す. 検索要求の表現は翻訳者によりかなりばらつきがあり, これは CLIR の精度に直接影響を与えると考えられる.

4.2 文書の翻訳による EJ-CLIR

本節では, 日本語文書の機械翻訳による EJ-CLIR (EJD と呼ぶ) の実験について述べる. 実験手順は以下のとおりである.

1. BMIR-J2 の文書 5,080 件 (D_J と呼ぶ) を日英 AS-TRANSAC により擬似英語文書 (D_e と呼ぶ) に変換し, NEAT で検索可能にする¹⁴.
2. Q_{EX} および Q_{EY} から生成した初期検索条件を用いて, D_e に対して J-MIR と同様に初期検索および LF を行う.

図 1 に, 表 3 の (3) に対応する初期検索条件に対して LF を行った後の検索条件を示す. 「text :1」の後に続いているのが初期検索語, 「text: 0.2」の後に続いているのが展開語である [20].

¹² 同一の訳になった検索要求は 9 件あった. よってこれらから生成された初期検索条件も同一である.

¹³ 本論文の全ての実験において, CLIR の実験のための AS-TRANSAC の辞書の変更は英日・日英ともに一切行っていない. また専門分野辞書なども使用しなかった.

¹⁴ NEAT は英単語に対する stemming 機能を内蔵している.

```
text :1, agenc, depress,
government, measur,
text :0.2, busi, recoveri,
bottom, was, econom, plan,
hang, aspect, cycl, research,
april, consumpt, postwar,
still, low, monthli, februar,
reduct, adjust, escap, index,
consider, yen, world,
diffusion, recess, real, late,
longest, uptrend;
```

図 1: An example expanded query from Q_{EX} .

表 4 に EJD の実験結果を示す. INIT, LF, GLF の意味は表 1 と同様である. Q_{EX} については $K = 0.8, b = 0.5, n = 10, m = 25$, Q_{EY} については $K = 0.6, b = 0.7, n = 8, m = 25$ とした. 表中の「%base」の 2 つの値はそれぞれ 11 点平均適合率が J-INIT / J-GLF の何%であることを表している. この表の縦方向の対に関する符号検定で有意差が見られたのは, Q_{EX} については EJD-INIT と EJD-LF / EJD-GLF の間 ($\alpha = 0.05$), Q_{EY} については EJD-INIT と EJD-LF の間 ($\alpha = 0.05$), EJD-INIT と EJD-GLF の間 ($\alpha = 0.01$) であった. 一方, 横方向の対, すなわち Q_{EX} と Q_{EY} の対は全て有意差なしであった. Q_{EX} と Q_{EY} の平均精度に著しく差があるにも関わらず有意差が検出されないのは, 表 5 の (1) に示したように, ほとんどの検索要求について X が Y より勝っているわけではなく, X が勝っているものと Y が勝っているものが約同数あるためである. 実際, 第一筆者 (翻訳者 X) が Q_{EX} と Q_{EY} の英文を比較した結果, X よりも Y のほうがやや口語的である傾向は見られたものの, 前者の翻訳の質が後者より全体的に勝っているとは言えなかった¹⁵.

TREC-6 における欧米言語対を扱う CLIR における精度が 50%-75% 程度であった [23] ことを考えると, 我々の EJ-CLIR は高精度であると言える¹⁶. とくに, Q_{EX} の EJD-LF / EJD-GLF は J-INIT の 90% 以上の精度を実現している. ただし, Q_{EX} と Q_{EY} の比較から明らかなように, MIR をベースラインとした評価値は翻訳者に大きく依存するため, この数値はひとつの目安としてとらえるべきであろう. ひとりの翻訳者を用いた CLIR の多くの従来研究においても, 翻訳者を差し替えることにより MIR をベースラインとした評価値は大きく変わることが予想される.

¹⁵ X は情報検索の研究者であるが, 訳語を選択する際に検索語として危険な多義語を意識的に避けるなどの「作為的な翻訳」は行っていない.

¹⁶ 欧米言語同士を扱う CLIR は日本語と英語を扱う CLIR よりも一般には容易であると考えられる. 例えば, 同一文字セットの欧米言語対においては, 固有名詞が同一の表記になることが多いので, 翻訳に失敗しても検索に成功する可能性がある.

4.3 検索要求の翻訳による EJ-CLIR

本節では、英語検索要求の機械翻訳による EJ-CLIR (EJQ と呼ぶ) の実験について述べる。実験手順は以下のとおりである。

1. Q_{EX} および Q_{EY} の自然言語文を英日 ASTRANSAC によりそれぞれ擬似日本語文 Q_{jX} および Q_{jY} に変換する。小文字の “j” は、英日 MT の出力であることを表している。
2. Q_{jX} および Q_{jY} をもとに、 D_J に対して J-MIR と同様に初期検索および LF を行う。

表 3 の (5)(6) に、同表の (3)(4) に対応する Q_{jX} および Q_{jY} の例を示す。この例では、“depression” や “recession” といった検索要求中の重要語の誤訳 (それぞれ「低下」, 「後退」) が見られ、また Q_{jX} の例では “related” を主動詞と解釈したことによる構文解析誤りが見られる¹⁷。図 2 に、表 3 の (5) に対応する初期検索条件に対して LF を行った後の検索条件を示す。

text :1, 関係, 機関, 政治, 低下, 反, 法案;
text :0.2, 成立, 本会議, 改革, 参院, 否決, 可決, 通過,
与野党, 関連, 審議, 同法, 案, 野党, 予算, 幹事, 護照;

図 2: An example expanded query from Q_{jX} .

表 6 に EJQ の実験結果を示す。 Q_{jX} については $K = 1.2, b = 0.3, n = 8, m = 15, Q_{jY}$ については $K = 1.6, b = 0.2, n = 6, m = 25$ とした。この表の縦方向の対に関する符号検定で有意差が見られたのは、 Q_{jX} については EJQ-INIT と EJQ-LF/ EJQ-GLF の間 ($\alpha = 0.01$), Q_{jY} については EJQ-INIT と EJQ-LF の間 ($\alpha = 0.05$) および EJQ-INIT と EJQ-GLF の間 ($\alpha = 0.01$) であった。また 4.2 節同様、 Q_{jX} のほうが Q_{jY} よりも全体的に精度は高いが、横方向の対は全て有意差なしであった。なお Q_{jX} と Q_{jY} との検索要求毎の優劣は表 5 の (2) の通りである。

表 4 と表 6 の比較から、Oard らの結果 [17] 同様、EJQ よりも EJD のほうが全体的に精度が高いことがわかる。一般に、翻訳時には EJD のほうが多くの文脈が活用できるため翻訳精度が高くなり、さらに検索時には EJD のほうが誤訳された語の影響が文書中の他の語で相殺され検索精度が高くなると考えられる。ただし、比較的小規模な BMIR-J2 を用いた今回の実験では、表 4 と表 6 で対応する結果の対は全て有意差なしであった。

なお、MT を用いた CLIR において検索要求翻訳のアプローチと文書翻訳のアプローチを比較する際には、MT の非対称性を考慮する必要がある。我々の実験では、EJD において日英 ASTRANSAC を、EJQ において英日 ASTRANSAC を用いたが、現状

¹⁷ Q_j の検索要求が完全に復元されたものは Q_{jX} で 13 件、 Q_{jY} で 10 件あった。よって同じ K, b の値を用いればこれらの初期検索精度は J-MIR と同一 (100%) となる。

では日英に比べ英日のほうが翻訳品質がよいことがわかっている¹⁸。よって、もしも日英と英日の品質が同程度の MT により EJD と EJQ の比較実験を行えば、BMIR-J2 規模でも有意な差が検出できる可能性はある。

5 日本語検索要求による擬似英語文書の検索 (Je-CLIR)

本章では、日本語検索要求を MT により英訳し、擬似英語文書を検索する Je-CLIR (JeQ と呼ぶ) の実験について述べる¹⁹。これは、1999 年 9 月に開催予定である NTCIR の JE-CLIR タスクの予備実験として行ったものである。本来ならば予備実験でも英語文書 (D_E と呼ぶ) を用いて JE-CLIR の実験を行い、J-MIR や EJ-CLIR との比較を行うべきである。しかし、我々には D_E の持ち合わせがなく、かつこれまでに NEAT の英語文書に対する検索精度評価を行ったことがなかったので、代わりに以下の手順で Je-CLIR の実験を行うことにした。

1. Q_J を日英 MT により擬似英語文 Q_e に変換し、 Q_{EX} および Q_{EY} の場合 (4.1 節) と同様に初期検索条件を作成する。
2. Q_e を用いて、4.2 節で作成した D_e に対して J-MIR と同様に初期検索および LF を行う。

表 3 の (2) に、同表の (1) に対応する Q_e の例を示す。

上記 JeQ の実験は、検索要求の翻訳による JE-CLIR のほば上限を与えられとされる。なぜなら、前者が Q_e による D_e の検索、すなわち同一の日英 MT が出力した擬似英語同士のマッチングを意味するのに対し、後者は Q_e による自然言語文書 D_E の検索を意味し、一般には前者のほうが容易であると推測されるからである。

さらに本実験では、通常の LF に加えて、我々が独自に作成した別の日本語コレクション TCIR-N1 [20][22] を用いて翻訳前の LF (pre-translation local feedback) [1] も試みた。CLIR における通常の LF が検索要求の翻訳後 (post-translation) に検索対象そのものを用いて行うのに対し、翻訳前の LF は、検索要求と言語が同じかつ検索対象とは独立な monolingual corpora を用いる²⁰。これは、コーパスを用いた CLIR においては、翻訳精度をあげるための文脈を補う効果がある。本実験での翻訳前の LF の手順は以下の通りである。

1. Q_J より TCIR-N1 の文書 5,048 件 (D'_j と呼ぶ) に対して初期検索を行う。 K, b には表 1 の BMIR-J2 における最適値 0.4 および 0.5 を流用する。

¹⁸ 例えば一般用語辞書の規模を比較すると、英日は約 24 万語、日英は約 14 万語である。

¹⁹ 自然言語の英語文書を用いていないため、文書翻訳によるアプローチ JeD を考慮することはできない。

²⁰ よってこの実験では 1 章で述べた (b) のリソースも扱っていることになる。

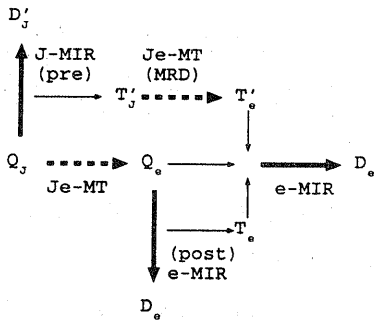


図 3: JeQ with pre/post-translation feedback.

2. 通常の LF と同様に、新しい検索語 (T_j' と呼ぶ) を抽出する。ただし、簡単のために擬似正解文書数 n' は表 1 の最適値 4 で固定し、展開語数 m' のみを最適化する。
3. T_j' の各語を日英 MT により擬似英語に変換し (T_e' と呼ぶ)、前述の Q_e に付加する。

図 3 に JeQ の実験の流れを示す。図中で、e-MIR は擬似英語の MIR を、Je-MT は日英 MT を表し、上半分が翻訳前の LF に、下半分が翻訳語の LF に対応する。ここで注意すべきは、コーパス利用の CLIR とは違い、翻訳前の LF は Q_j から Q_e への翻訳精度になんら寄与しないという点である。すなわち、 Q_j および T_j' はそれぞれ独立に日英 MT により Q_e および T_e' に変換される。また、 T_j' は文ではなく語であるので、MT は MRD としてしか機能しない。

表 7 に JeQ の結果を示す。 Q_e については、 $K = 0.2, b = 0.9, n = 10, m = 20$ とした。PRE は翻訳前の LF の結果 ($m' = 10$)、PRE+GLF は GLF と PRE の検索条件をマージした場合の結果を表している²¹。すなわち、図 3 の上半分の処理のみ施したものが PRE、下半分の処理のみ施したものが LF および GLF、そして両方の処理を施したものが PRE+GLF である。符号検定で有意差が見られたのは JeQ-INIT と JeQ-PRE / JeQ-LF の間 ($\alpha = 0.05$) および JeQ-INIT と JeQ-GLF / JeQ-PRE+GLF の間 ($\alpha = 0.01$) であった。全体的に精度が高いのは、前述の通り D_E のかわりに D_e を用いているためである。Je-CLIR により MIR と同水準の精度が実現できたことから、もしも AS-TRANSAC の対称性がほぼ成り立つならば、実際の JE-CLIR の精度は EJ-CLIR と同程度になると考えられる。

JeQ-LF には及ばないものの、JeQ-PRE が有効であることは興味深い。翻訳前の LF の効果は、利用するコーパス D_j' が検索対象 D_e とどの程度似た性質をもつか、および展開語の翻訳精度に依存すると考えられるが、本実験の D_j' (TCIR-N1) は 1995 年の日本経済新聞の記事、 D_e は 1994 年の毎日新聞の記事を日英 MT により変換したものであ

²¹ 今回は行わなかったが、翻訳前の LF においても GLF と同様に n, m の値を場合分けすることは可能である。

り、 D_j' と D_e との関連性はそれほど高くはないと考えられる。さらに、前述のとおり、展開語の翻訳は単に MRD の訳語の第一候補を用いていることに相当するため、それほど高精度であるとは考えられない。以上にも関わらず JeQ-PRE が有効であったことは、日本語を扱う CLIR において、また MT 利用の CLIR において、翻訳前の LF が有望であることを示していると考えられる。また、翻訳前後の LF の併用 JeQ-PRE+GLF と JeQ-GLF との間には有意差がなかったが、翻訳前の LF においても n の最適化や GLF と同様の最適化を行えば、併用の効果が得られる可能性はある。

6 まとめ

本論文では、機械翻訳システムとローカルフィードバックを利用した日・英間の CLIR の検索精度評価を行い、その有効性を示した。英語検索要求による日本語文書の検索実験では、文書翻訳のアプローチにより MIR の 90% 以上の精度を、検索要求翻訳のアプローチにより MIR の 80% 以上の精度を実現し、またこれらの評価値は検索要求の翻訳者に依存することを示した。日本語検索要求による擬似英語文書の検索実験では、検索要求の翻訳前後のローカルフィードバックを扱い、MIR と同程度の精度を実現した。これは、日本語検索要求と英語文書を扱う CLIR における我々のアプローチの上限を示すものと考えられる。いずれの実験においても、ローカルフィードバックが検索精度向上のための有効な手段であることがわかった。

図 4 は、J-GLF、JeQ-PRE+GLF、EJD-GLF (Q_{EX})、EJQ-GLF (Q_{jX}) の再現率-適合率曲線 [25] である。また表 8 は、これらの CLIR の結果を検索要求毎に J-GLF と比較した結果である。この表から明らかのように、全体の 3 分の 1 程度の検索要求については CLIR のほうが MIR よりも精度が高い。これは 2 章で述べた下記の主張を裏付けるものであると考えられる。

「現状の MT は人が読んで満足するレベルには達していないが、CLIR には十分有用である」

例えば、表 3 の (2) は、翻訳結果としてはいま一步であるが、これを初期検索条件に変換すると、表 3 の (3) の人手で作成した検索要求から作成した初期検索条件と同一になる。すなわち、初期検索語 “agenc, depress, government, measur” を含む図 1 の 1 番目の条件に変換される。現状の MIR の性能自体が必ずしも充分であるわけではないが、少なくとも MT と CLIR との親和性は高いと言える。

CLIR では MIR に増して検索結果の提示方法が重要となる。例えば、1 章で述べた B 氏のようなユーザーに対しては、言語障壁を隠蔽するユーザーインタフェースを考慮する必要があると考えられる [12] [26]。今後は、検索精度・翻訳精度の向上にとどまらず、ユーザーが検索結果から実際に情報を入手するまでのインタラクションを含めた研究を進める予定である。

表 3: Example search requests.

(1) Q_J	行政機関が関係する不況対策
(2) Q_e	Measure against depression relating with a governmental agency
(3) Q_{EX}	Anti-depression measures related to governmental agencies
(4) Q_{EY}	Specific measures taken by government administrators to combat the recession
(5) Q_{jX}	反低下法案は政治の機関に関係があった
(6) Q_{jY}	後退と格闘するために政府行政官によって得られた特定の法案

表 4: EJD results.

	Q_{EX}		Q_{EY}	
	11pt	%base	11pt	%base
EJD-INIT	0.364	79/73	0.323	70/65
EJD-LF	0.416	90/83	0.353	76/71
EJD-GLF	0.427	92/85	0.361	78/72

表 5: Per-query comparisons between X and Y.

11pt	(1)EJD-		(2)EJQ-	
	INIT	GLF	INIT	GLF
X > Y	26	27	22	28
X = Y	4	0	10	3
X < Y	20	23	18	19
計	50	50	50	50

表 6: EJQ results.

	Q_{jX}		Q_{jY}	
	11pt	%base	11pt	%base
EJQ-INIT	0.335	72/67	0.262	57/52
EJQ-LF	0.379	82/76	0.313	68/63
EJQ-GLF	0.387	84/77	0.326	70/65

表 7: JeQ results.

	Q_e	
	11pt	%base
JeQ-INIT	0.409	88/82
JeQ-PRE	0.426	92/85
JeQ-LF	0.457	99/91
JeQ-GLF	0.461	100/92
JeQ-PRE+GLF	0.462	100/92

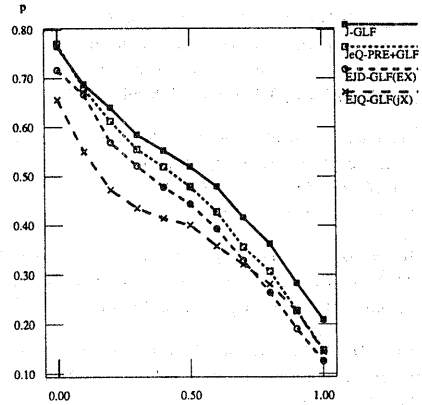


図 4: Recall/Precision Curves.

表 8: Per-query comparisons with J-GLF.

J-GLF と比べ 11pt が*	(1)EJD-GLF		(2)EJQ-GLF		(3)JeQ- PRE+ GLF
	Q_{EX}	Q_{EY}	Q_{jX}	Q_{jY}	Q_e
高い	13	11	16	13	16
同じ	0	1	2	1	0
低い	37	38	32	36	34
計	50	50	50	50	50

参考文献

- [1] Ballesteros, L. et al.: Resolving Ambiguity for Cross-language Retrieval, *ACM SIGIR '98 Proceedings*, pp. 64-71 (1998).
- [2] Carbonell, J.G. et al.: Translingual Information Retrieval: A Comparative Evaluation, *IJCAI'97*, pp.708-714 (1997).
- [3] Dumais, S.T. et al.: Automatic Cross-Linguistic Information Retrieval using Latent Semantic Indexing, *ACM SIGIR Workshop on Cross-Linguistic Information Retrieval*, <http://www.rxrc.xerox.com/research/mltt/DMHead/CLIR/SIGIR96CLIR.html> (1996).
- [4] 藤井ほか: 多言語情報検索のための複合語翻訳, 情処学会情報学基礎研究会 98-FI-51, pp. 55-62 (1998).
- [5] Gachot, D.A. et al.: The SYSTRAN NLP Browser - An Application of Machine Translation Technology in Multilingual Information Retrieval, *ACM SIGIR Workshop on Cross-Linguistic Information Retrieval*, <http://www.rxrc.xerox.com/research/mltt/DMHead/CLIR/SIGIR96CLIR.html> (1996).
- [6] Gaussier, E. et al.: Xerox TREC-6 Site Report: Cross Language Text Retrieval, TREC6, <http://trec.nist.gov/pubs/trec6/t6.proceedings.html> (1997).
- [7] Hirakawa, H. et al.: EJ/JE Machine Translation System ASTRANSAC - Extensions towards Personalization, *MT Summit III Proceedings*, pp. 73-80 (1991).
- [8] Hull, D.: Using Statistical Testing in the Evaluation of Retrieval Experiments, *ACM SIGIR '93 Proceedings*, pp. 329-338 (1993).
- [9] Hull, D. et al.: Querying Across Languages: A Dictionary-Based Approach to Multilingual Information Retrieval, *ACM SIGIR '96 Proceedings*, pp. 49-57 (1996).
- [10] Jones, G. et al.: Cross-Language Information Access: a case study for English and Japanese, 情処学会情報学基礎研究会 98-FI-51, pp. 47-54 (1998).
- [11] Kando, N. et al.: Cross-lingual Information Retrieval using Automatically Generated Multilingual Keyword Clusters, *Proc. of The 3rd International Workshop on Information Retrieval with Asian Languages*, pp. 86-94 (1998).
- [12] 菊井ほか: インターネット情報ナビゲーションにおける多言語機能, 情処学会 自然言語処理の応用に関するシンポジウム, <http://titan.mcnet.net.jp/titan-help/jp/3.html> (1995).
- [13] Kinoshita, S. et al.: ASTRANSAC - Toshiba Machine Translation System, *MT Summit VI Proceedings*, p. 284 (1997).
- [14] Mateev, B. et al.: ETH TREC-6: Routing, Chinese, Cross-Language and Spoken Document Retrieval, TREC6, <http://trec.nist.gov/pubs/trec6/t6.proceedings.html> (1997).
- [15] NTCIR(NACSIS Test Collection for Information Retrieval Systems): <http://www.rd.nacsis.ac.jp/~ntcadm>
- [16] Oard, D.W.: Alternative Approaches for Cross-Language Text Retrieval, *AAAI Spring Symposium on Cross-Language Text and Speech Retrieval Electronic Working Notes*, <http://www.ee.umd.edu/medlab/filter/sss/papers/> (1997).
- [17] Oard, D.W. et al.: Document Translation for Cross-Language Text Retrieval at the University of Maryland, TREC6, <http://trec.nist.gov/pubs/trec6/t6.proceedings.html> (1997).
- [18] Pirkola, A.: The Effects of Query Structure and Dictionary Setups in Dictionary-Based Cross-language Information Retrieval, *ACM SIGIR '98 Proceedings*, pp. 55-63 (1998).
- [19] Robertson, S. E. et al.: Simple, Proven Approaches to Text Retrieval, Computer Laboratory, University of Cambridge (1994).
- [20] Sakai, T. et al.: Application of Query Expansion Techniques in Probabilistic Japanese News Filtering, *Proc. of The 3rd International Workshop on Information Retrieval with Asian Languages*, pp. 46-55 (1998).
- [21] 酒井ほか: 情報検索システム評価のためのテストコレクション, *Computer Today 1998年9月号* (No.87), pp. 31-35, サイエンス社 (1998).
- [22] 酒井ほか: 情報フィルタリングのためのプール式と文書構造を利用した検索条件生成と検索精度評価, 情処学会論文誌, Vol. 39, No. 11, pp. 3076-3083 (1998).
- [23] Schauble, P. et al.: Cross-Language Information Retrieval(CLIR) Track Overview, TREC6, <http://trec.nist.gov/pubs/trec6/t6.proceedings.html> (1997).
- [24] Sheridan, P. et al.: Experiments in Multilingual Information Retrieval using the SPIDER system, *ACM SIGIR '96 Proceedings*, pp. 58-65 (1996).
- [25] Witten, I. H. et al.: *Managing Gigabytes: Compressing and Indexing Documents and Images*, Van Nostrand Reinhold, pp. 148-151 (1994).
- [26] Yamabana K. et al.: A Language Conversion Front-End for Cross-Linguistic Information Retrieval, *ACM SIGIR Workshop on Cross-Linguistic Information Retrieval*, <http://www.rxrc.xerox.com/research/mltt/DMHead/CLIR/SIGIR96CLIR.html> (1996).