

同一トピックの日英ブログにおける文化間差異の発見支援

中崎 寛之[†] 川場真理子[†] 山崎小有里^{††} 宇津呂武仁[†] 福原 知宏^{†††}

[†] 筑波大学大学院システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 筑波大学第三学群工学システム学類 〒 305-8573 茨城県つくば市天王台 1-1-1

^{†††} 東京大学 人工物工学研究センター 〒 277-8568 千葉県柏市柏の葉 5-1-5

あらまし 本研究では、ある同一のトピックについてまとまった規模の記述が書かれたブログサイトを、日英各言語について検索し、その記述内容を二言語間で対照分析する方式を提案する。あるトピックの日英二言語表現を得る際には、Wikipedia の日英二言語エントリを用いる。ブログサイトの検索においては、特定トピックを表すキーワードを用いて商用検索エンジン API により上位のブログサイトを収集し、これを、特定トピックを表すキーワード、および Wikipedia から収集した関連語の出現数順にリランキングする方法を用いる。この方法により、そのトピックについての記述が多く含まれる有用なブログサイト、および、それらのブログサイト中における有用な記事を上位にランキングすることが可能となる。さらに、これまでに行った評価実験では、それらのブログサイトの内容を日英二言語間で対照分析することにより、ブログ特有の個人レベルの情報や意見における国間差異が多数観測されている。

キーワード ブログ, トピック分析, 文化間差異, Wikipedia, 言語横断情報検索

Semi-Automatic Discovery of Cross-Cultural Gaps through Japanese/English Blog Distillation

Hiroyuki NAKASAKI[†], Mariko KAWABA[†], Sayuri YAMAZAKI^{††}, Takehito UTSURO[†], and
Tomohiro FUKUHARA^{†††}

[†] Grad. Sch. Systems and Information Engineering, University of Tsukuba, Tsukuba, 305-8573, Japan

^{††} College of Eng. Sys., Third Cluster of Colleges, University of Tsukuba, Tsukuba, 305-8573, Japan

^{†††} Research into Artifacts, Center for Engineering, University of Tokyo Kashiwa, Chiba 277-8568, Japan

Abstract The goal of this paper is to cross-lingually analyze multilingual blogs collected with a topic keyword. The framework of collecting multilingual blogs with a topic keyword is designed as the blog feed retrieval procedure. Multilingual queries for retrieving blog feeds are created from *Wikipedia* entries. Finally, we cross-lingually and cross-culturally compare less well known facts and opinions that are closely related to a given topic. Preliminary evaluation results support the effectiveness of the proposed framework.

Key words blog, topic analysis, cultural gaps, Wikipedia, cross-lingual IR

1. はじめに

近年、世界中でブログサービスやブログツールが普及し、各地域の人々がそれぞれインターネット上で個人の意見や評判を発信することが可能になった。それに伴い、様々な情報がブログに記載され、商用ブログ検索サービスを利用することでそれらの情報を取得することができるようになった。具体的なサービスの例として、*Technorati*^(注1)、*BlogPulse*^(注2)、*kizasi.jp*^(注3)、

blogWatcher^(注4)などが挙げられる。これらの検索サービスは、巨大なブログ空間の索引付けという観点から見ると、キーワードや評判、時系列変化や人手によって作成されたカテゴリ情報などを索引として用いて、利用者の求めるブログ記事やブログサイトを検索する。また、多言語ブログサービスとしては、*Globe of Blogs*^(注5)が言語横断ブログ記事検索機能を提供している。他にも、アジア言語ブログの検索機能を提供している

(注1) : <http://technorati.com/>

(注2) : <http://www.blogpulse.com/>

(注3) : <http://kizasi.jp/>(日本語のみ)

(注4) : <http://blogwatcher.pi.titech.ac.jp/>(日本語のみ)

(注5) : <http://www.globeofblogs.com/>

Best Blogs in Asia Directory ^(注6) や、多言語ブログ記事の分析を行っている *Blogwise* ^(注7) がある。

上記の現状を踏まえた上で、本研究では、ある同一のトピックについてまとめた規模の記述が書かれたブログサイトを、日英各言語について検索し、その記述内容を二言語間で対照分析する方式を提案する。あるトピックの日英二言語表現を得る際には、Wikipedia の日英二言語エントリを用いる。ブログサイトの検索においては、特定トピックを表すキーワードを用いて商用検索エンジン API により上位のブログサイトを収集し、これを、特定トピックを表すキーワード、および Wikipedia から収集した関連語の出現数順にリランキングする方法を用いる [1]~[3]。この方法により、そのトピックについての記述が多く含まれる有用なブログサイト、および、それらのブログサイト中における有用な記事を上位にランキングすることが可能となる。さらに、これまでに行った評価実験では、それらのブログサイトの内容を日英二言語間で対照分析することにより、ブログ特有の個人レベルの情報や意見における国間差異が多数観測されている。

本研究の全体的枠組みを図 1 に示す。手順としては、まずトピック名である日英 Wikipedia エントリのタイトルを検索語として、日英ブログサイトを検索する。次に、トピックの日英 Wikipedia エントリから関連語を抽出し、関連語を用いて検索したブログサイト集合からトピックに関わりのあるブログ記事を検索する。そして、検索した日英ブログ記事集合から共起語を抽出することで、共起語単位で日英ブログの文化間差異の発見を支援することができる。また、Wikipedia 関連語を用いてブログサイト集合とブログ記事集合をランキングすることによって、トピックについて詳細な記述を載せているブログサイトおよびブログ記事の発見が容易になり、ブログサイト・ブログ記事単位で日英ブログの文化間差異の発見支援となる。

2. 評価用トピック

まず、評価用トピック候補として、Wikipedia の中から 50 トピックほど選定した。選定したトピックはいずれも日英 Wikipedia エントリが存在し、かつトピックに関わる日英ブログサイトが存在すると思われるトピックである。このトピック候補の中から、評価用トピックとして、「捕鯨」、「臓器移植」、「喫煙」、「サブプライムローン」の社会系トピック 4 つを選定した。これらの評価用トピックの要約と日英ブログにおける評価用トピックに対する主な意見を表 1 に示す。

3. 二言語対照ブログ分析

3.1 ブログサイト検索

本研究ではまず、Wikipedia の中のある特定のトピックから、そのトピックについての意見や評判などの情報が書かれているブログサイトを探し、対応づける。しかし、現在のブログ検索サービスでは、被リンク数の多い人気ブログサイトの記事から

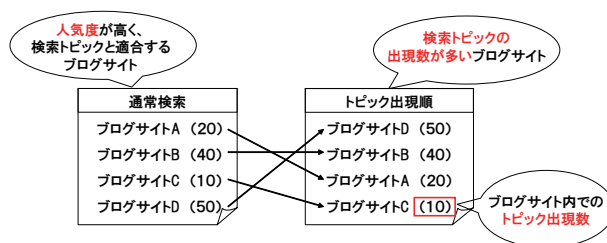


図 2 特定トピックに一致するブログサイトの検索手法

優先的に検索されるために、被リンク数は多くないが、特定トピックについて詳細な記述を載せているブログサイトが検索されにくい。本研究の目的を達成するためには、トピックについて詳細な記述を載せているブログサイトの集合を得る必要がある。そこで、本稿では、検索トピックがブログサイトにどれだけ出現しているかで検索トピックについて述べられているブログサイトかどうかを判定するという手法 [1] を用いる。具体的には、図 2 に示すように、

エントリ名を検索クエリとした通常の方法でブログサイトを検索した後、検索されたブログサイト集合をエントリ名の出現数が多い順に並び替える。

さらに、各ブログサイトにおいて、エントリ名のヒット数を求め、ヒット数が下限未満 (本論文では、10) のブログサイトを削除した。

ブログサイトを検索するために、本研究では日本語ブログの検索には、Yahoo!Japan 検索 API を、英語ブログの検索には米 Yahoo!検索 API を利用し、日本語ブログでは大手 11 社 ^(注8)、英語ブログでは大手 12 社 ^(注9) のブログ会社のドメインに限定して検索を行った。

3.2 ブログ記事検索

次に、検索した日英ブログサイト集合の中から、トピックについて詳しく書かれたブログ記事を検索する。手法としては、トピック名がタイトルである各言語の Wikipedia エントリのリダイレクト、さらに Wikipedia エントリの本文から太字、他エントリリンクをブログ記事検索のための関連語として抽出する。そして、抽出した関連語のいずれかが出現する各言語のブログ記事をブログサイト集合内からそれぞれ検索する。

本研究で評価した各トピックの Wikipedia 関連語数、各トピックで検索したブログサイト数、検索したブログサイト中で Wikipedia 関連語のいずれかが出現したブログ記事数、検索したブログ記事本文に含まれる総形態素数および総単語数を表 2 に示す。

3.3 ブログ記事からの共起語抽出

本研究では、対照分析の方法として、各言語のブログに出現する共起語を用いる。まず、検索した日本語ブログ記事からは名詞句を抽出し、検索した英語ブログ記事からは一単語、二単

(注6) : <http://www.misohoni.com/bba/>

(注7) : <http://www.blogwise.com/>

(注8) : FC2.com, yahoo.co.jp, rakuten.ne.jp, ameblo.jp, goo.ne.jp, live-door.jp, Seesaa.net, jugem.jp, yaplog.jp, webry.info.jp, hatena.ne.jp

(注9) : blogspot.com, msnblogs.net, spaces.live.com, livejournal.com, vox.com, multiply.com, typepad.com, aol.com, blogsme.com, wordpress.com, blog-king.net, blogster.com

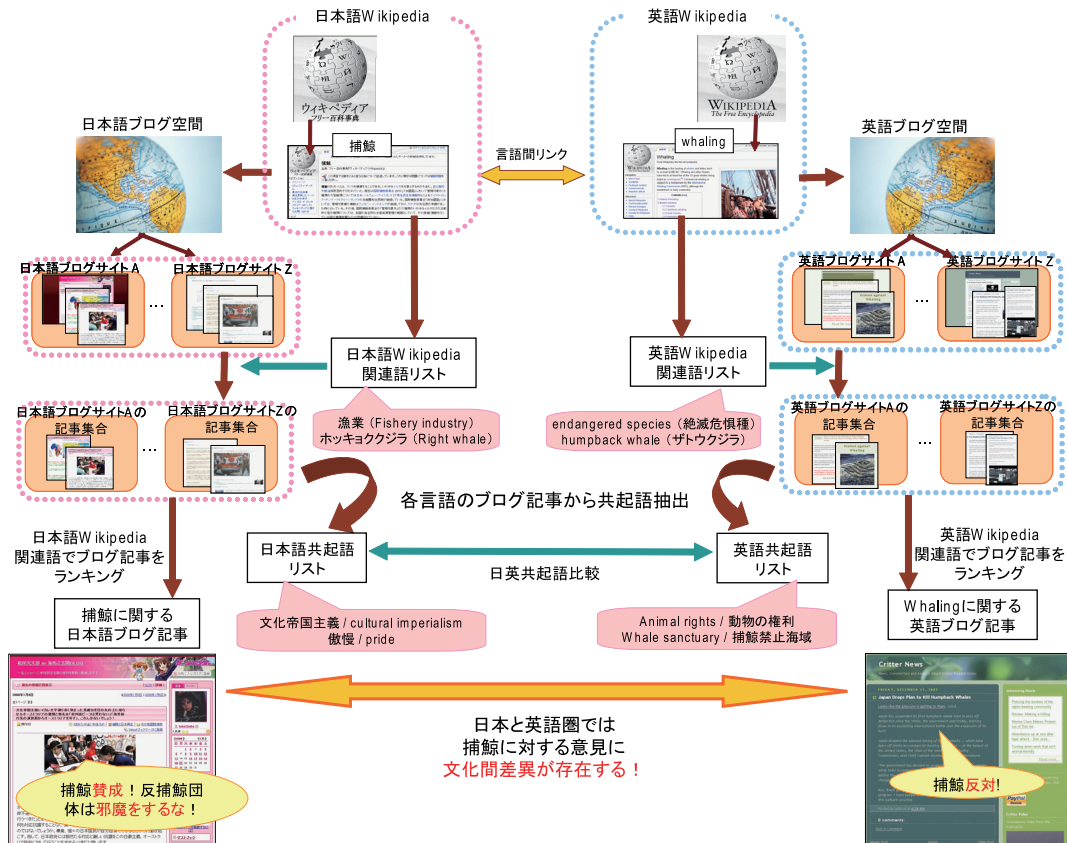


図 1 二言語対照ブログ分析の全体的枠組み

表 1 評価トピックの日英ブログにおける意見の要約

評価トピック — 概要	
日英ブログ間の意見の差異	
(日本語ブログ)	(英語ブログ)
捕鯨 (Whaling) — 捕鯨問題において、捕鯨賛成派と捕鯨反対派が対立している。 多くのブログが捕鯨賛成派、反捕鯨団体を激しく非難している。また、捕鯨について書いているブロガーには、右寄りの考えを持つ人が多くみられた。	多くのブログが捕鯨反対派、特に日本の捕鯨を激しく非難している。また、いくつかのブロガーはホエールウォッチングについて書いている。
臓器移植 (Organ transplant) — 治療のために、多くのブログは日本の臓器移植法改正の必要性を訴えている。また、いくつかのブログでは、日本の医者によって行われた病気腎移植問題のことに注目している。	提供されたドナーの臓器を患者に移植する医療法 多くのブログで、臓器不足という現状から、臓器移植のドナー登録を強く推奨している。また、いくつかのブログでは中国の違法臓器摘出を非難している。
喫煙 (Tobacco smoking) — 喫煙することで、多くのブログで、健康や喫煙マナーの悪さを理由に喫煙に反対しているが、一部のブログは喫煙賛成派である喫煙者のブロガーであった。	人の健康を損なうということ知られている。 多くのブログで、肺がんの原因である喫煙に反対している。
サブプライムローン (Subprime lending) — 近年発生した世界金融危機の大きな原因の一つ 多くのブログで、米国のサブプライム問題による影響で日本経済が悪化したと指摘。	多くのブロガーが経済学者で、サブプライムローンによって発生した住宅バブルや、現在の金融危機や経営危機の発生原因など考察している。

語連語、三単語連語を抽出し、それぞれの頻度統計と出現確率を求める。日本語名詞句 X_J の日本語ブログにおける出現確率 $P_J(X_J)$ と、英語一単語・二単語連語・三単語連語 Y_E の英語ブログにおける出現確率 $P_E(Y_E)$ を以下のようにそれぞれ定義する。

$$P_J(X_J) = \frac{X_J \text{ の出現頻度}}{\text{対象日本語ブログサイト集合内の総形態素数}}$$

$$P_E(Y_E) = \frac{Y_E \text{ の出現頻度}}{\text{対象英語ブログサイト集合内の総単語数}}$$

また、抽出した語句の訳語が相手言語ブログに出現するか調べ

表 2 (日本語/英語) ブログ記事検索に用いた Wikipedia 関連語統計情報

評価トピック	Wikipedia 関連語数	ブログサイト数	ブログ記事数	総形態素数/ 総単語数
捕鯨	162 / 174	121 / 239	2232 / 6532	5024966 / 2611942
臓器移植	100 / 231	89 / 206	696 / 1301	995927 / 781476
喫煙	399 / 276	86 / 252	1481 / 400	1323767 / 492727
サブプライムローン	39 / 68	134 / 205	1088 / 1216	980552 / 883450

るために、Wikipedia の言語間リンクを使用して語句の訳語を求め、Wikipedia で語句の対訳を取得できない場合は、英辞郎^(注10)で語句の対訳を取得する。さらに、抽出した語句の出現率と対訳語句の出現率から、相手言語ブログと比較した出現確率比を求める。本研究では、抽出した日本語名詞句 X_J と X_J の英訳 X_E の出現確率比 $R_J(X_J, X_E)$ と、英語単語・二単語連語・三単語連語 Y_E と Y_E の和訳 Y_J の出現確率比 $R_E(Y_E, Y_J)$ を以下のように定義した。

$$R_J(X_J, X_E) = \frac{P_J(X_J)}{P_E(X_E)}, \quad R_E(Y_E, Y_J) = \frac{P_E(Y_E)}{P_J(Y_J)}$$

そして、定義した出現確率比で各言語の共起語をランキングし、それぞれの言語で高い出現確率比の共起語を比較することで、共起語単位でブログ空間におけるトピックの文化間ギャップ発見を支援することができる。

各トピックの日英ブログ記事から抽出した共起語例を表 3 に示す。表 3 から、片言語で特徴的な共起語もあれば、両言語で多く出現する共起語も存在することがわかる。ここで定義した各共起語の出現確率と出現確率比を 3.5 節の共起語マップの座標として用いる。

3.4 ブログサイト・ブログ記事の順位付け

本研究では、よりトピックについて詳しく書いてあるブログサイトおよびブログ記事を得るために、検索した各言語のブログサイト群とブログ記事群をそれぞれ順位付けする。順位付けには、3.2 節で抽出した Wikipedia 関連語を用いる。

ブログ記事の順位付けには、以下の式を用いる。

$$PostScore(p) = \sum_t (weight(type(t)) \times freq(t))$$

$weight(type(t))$ は、Wikipedia 関連語 t の種類 $type(t)$ に付与する重みで、 $freq(t)$ は、ブログ記事 p 内における Wikipedia 関連語 t の出現頻度である。また、Wikipedia 関連語 t の種類 $type(t)$ がリダイレクトの場合は重みを 3、太字の場合は重みを 2、他エントリリンクの場合は重みを 0.5 とする。上記の式を各ブログ記事に対して適用し、 $PostScore(p)$ の高いブログ記事がトピックについてより詳しいブログ記事である可能性が高い。

ブログサイトの順位付けには、以下の式を用いる。

$$SiteScore(s) = \sum_p PostScore(p)$$

ブログ記事 p は、ブログサイト s に含まれるブログ記事である。 $SiteScore(s)$ の高いブログサイトは、トピックについてより詳

しいブログサイトである可能性が高い。

各評価トピックのブログ記事上位 10 件の内容を表 4 と表 5 に示す。多くのブログ記事はトピックについて詳しく書いてあるブログ記事であり、いくつかのブログ記事ではそのトピックに対してブログ著者の賛成意見や反対意見が述べられていることがわかる。

3.5 文化間ギャップ発見支援システム

本研究では、同一トピックの日英ブログにおける文化間差異をより発見しやすくするために、文化間ギャップ発見支援システムを作成した。システムの使用方法を図 3 に示す。

まず、トピックがカテゴリ別に分類されているので、調べたいカテゴリを選択する。カテゴリ情報は、階層構造である Wikipedia の上位カテゴリ約 300 個を用いる。次に、選択したカテゴリのトピックリストが表示されるので、調べたいトピックを選択することで、そのトピックの日英ブログから抽出した共起語を提示する共起語マップを表示することができる。また、検索窓からトピックを検索することで、共起語マップを表示することも可能である。

共起語マップでは、日本語に特徴的な共起語群の代表語、英語に特徴的な共起語群の代表語、両言語共通の共起語群の代表語が提示される。提示された代表語を選択することで、その代表語と関連のある共起語群がマップに表示される。さらに、マップに表示された共起語を選択することで、選択した共起語と関連するブログ記事ランキングが提示される。これによって提示されたブログ記事を分析することで、日英ブログの文化間差異発見の足掛かりとなる。

文化間ギャップ発見支援システムを用いて表示した、トピック「捕鯨」と「エア・ギター」の共起語マップ例を図 4 に示す。共起語マップの横軸は、各共起語の出現確率比を表し、縦軸は各共起語の単言語における出現確率を表す。日本語ブログから抽出した日本語共起語 X_J は、座標 $(-R_J(X_J, X_E), P_J(X_J))$ に表示される。このとき、日本語共起語 X_J の英訳 X_E が英語ブログに出現しない場合は、日本語ブログで特徴的な共起語として最も左に表示される。また、英語ブログから抽出した英語共起語 X_E は、座標 $(R_E(X_E, X_J), P_E(X_E))$ に表示される。そして、英語共起語 X_E の和訳 X_J が日本語ブログに出現しない場合は、英語ブログで特徴的な共起語として最も右に表示される。

また、いくつかの共起語は相互的に関係が強く、それらは共通の話題から抽出された共起語群ということがわかる。さらに、片言語のみで特徴的である話題から抽出された共起語群は、縦軸から大きく離れている座標に表示される傾向にある。逆に、

(注10) : <http://www.eijiro.jp/>

表 3 各評価トピックの日英ブログから抽出した共起語例

(a) 捕鯨									
日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_J(X_J, X_E)$	日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_E(X_E, X_J)$
反捕鯨国	136	antiwhaling country	0	∞	動物の権利	0	animal rights	233	∞
調査捕鯨	620	scientific whaling	116	2.78	捕鯨禁止海域	2	whale sanctuary	101	97.15
—	—	—	—	—	絶滅危惧種	47	endangered species	303	12.40
—	—	—	—	—	ザトウクジラ	98	humpback whales	266	5.22
—	—	—	—	—	絶滅	236	extinction	299	2.44

(b) 臓器移植									
日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_J(X_J, X_E)$	日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_E(X_E, X_J)$
日本移植学会	75	対訳なし (The Japan Society for Transplantation)	0	∞	対訳なし (臓器摘出)	0	organ harvesting	270	∞
病気腎移植	442	対訳なし (transplant using diseased kidney)	0	∞	人権	9	human rights	508	71.93
脳死移植	366	brain-dead transplant	0	∞	ドナーカード	27	donor card	124	5.85
臓器移植法	123	organ transplant law	3	32.17	臓器提供	200	organ donation	673	4.29

(c) 喫煙									
日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_J(X_J, X_E)$	日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_E(X_E, X_J)$
マナー	127	manners	1	47.27	肺癌	18	lung cancer	637	95.08
禁煙	520	smoking cessation	59	3.28	乳癌	6	breast cancer	339	151.79
—	—	—	—	—	受動喫煙	136	secondhand smoke	95	1.88

(d) サブプライムローン									
日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_J(X_J, X_E)$	日本語共起語	出現頻度 (日本語)	英語共起語	出現頻度 (英語)	$R_E(X_E, X_J)$
日本経済	96	economy of Japan	0	∞	ウォールストリート	11	Wall Street	474	47.83
日経平均	100	Nikkei 225	1	90.10	経営危機	3	financial crisis	55	20.35
株価	451	stock prices	30	13.54	不動産	100	real estate	637	7.07
—	—	—	—	—	住宅バブル	32	housing bubble	189	6.56

両言語で共通している話題から抽出された共起語群は、日英共起語群がそれぞれ縦軸から近い座標に表示されることが多い。

トピック「捕鯨」では、英語ブログで特徴的な共起語は、捕鯨反対の意見をあらわすものが多い。逆に、日本語ブログで特徴的な共起語は、反捕鯨を表明している国を非難している意見をあらわすものが多かった。また、トピック「エア・ギター」では、英語ブログ特有で出現した共起語には、エアギター世界選手権大会で結果を残したプロエアギターリストを賞賛しているものが多かった。逆に、日本語ブログで特徴的な共起語は、エアギターの世界大会で活躍している日本人の話題や、日本の有名なあるキャラクターがエアギターの分野でも活躍していることに驚いているものが多く見られた。このことから、日英ブログから抽出した共起語が日英ブログの文化間差異の発見支援となることがわかった。

4. 関連研究

多言語でのブログ分析を対象とした研究として、日韓中英のブログ内で、キーワードのバーストの時系列の変化を各言語間で調べるものがある [4]。本稿で行った日英ブログの比較対照実験では、ブログの内容を見ており、キーワードのバーストの時系列の変化を調べるといったことは行っていない。また、同じ事象について、複数の情報源の情報の伝え方の異なりかたを分析する方式についての研究 [5] がある。この研究では複数の国の代表的なメディアが発信するニュースを情報源として、各々の国の世論がどのように事象を分析しているかの理解を図ろうとしている。

5. おわりに

本稿では、Wikipedia エントリを用いてトピックに関連する日英ブログサイトを検索し、その記述内容を二言語間で対照分

表 4 日英ブログ記事ランキング上位 10 件の詳細 (「捕鯨」,「臓器移植」)

トピック	
日本語	英語
記事ランキング順位/ サイトランキング順位/ 詳細	記事ランキング順位/ サイトランキング順位/ 詳細
捕鯨 (Whaling)	
(記事) 1 位, 3 位, 5 位, 6 位, 8 位/ (サイト) 1 位/ 捕鯨賛成. 反捕鯨団体を批判.	(記事) 1 位, 3 位, 4 位, 5 位, 10 位/ (サイト) 2 位/ 捕鯨に関して中立的立場. 日本に 30 年以上在住しているブロガー.
(記事) 4 位/ (サイト) 6 位/ 捕鯨賛成. 米国在住 12 年のブロガー.	(記事) 6 位/ (サイト) 3 位/ 捕鯨反対. 動物愛護運動家のブロガー.
(記事) 7 位, 9 位/ (サイト) 4 位/ 捕鯨賛成. ブログサイトランキング 1 位のブログと同じ著者.	(記事) 8 位/ (サイト) 10 位/ 捕鯨反対. シーシェパード派, 反グリーンピース派.
臓器移植 (Organ transplant)	
(記事) 2 位, 3 位, 4 位, 6 位, 9 位/ (サイト) 1 位/ 病気が腎移植のニュースを取り上げている. 病気が腎移植に反対する日本移植学会を批判.	(記事) 2 位, 6 位/ (サイト) 8 位/ 中国の違法臓器摘出を批判しているニュースを紹介
(記事) 7 位/ (サイト) 14 位/ 脳死移植に反対. 臓器移植法の改正は慎重に行うべきと主張.	(記事) 3 位/ (サイト) 2 位/ 違法臓器摘出を批判しているニュースを紹介
(記事) 8 位, 10 位/ (サイト) 7 位/ 病気が腎移植に反対. 患者が完治するとは思えないと主張.	(記事) 8 位/ (サイト) 7 位/ 臓器提供に関するニュース記事を紹介. ドナー登録することを強く推奨.

表 5 日英ブログ記事ランキング上位 10 件の詳細 (「喫煙」,「サブプライムローン」)

トピック	
日本語	英語
記事ランキング順位/ サイトランキング順位/ 詳細	記事ランキング順位/ サイトランキング順位/ 詳細
喫煙 (Tobacco smoking)	
(記事) 1 位/ (サイト) 7 位/ 喫煙者と非喫煙者の間で対立が起きていることを指摘. そもそも喫煙者の文化と非喫煙者の文化は違う.	(記事) 2 位/ (サイト) 6 位/ 米国北部と米国南部の喫煙率を比較. また, タバコは米国の主要な農産物の一つだと主張.
(記事) 2 位, 5 位, 10 位/ (サイト) 10 位/ 喫煙反対. 喫煙者をもっと喫煙マナーを守る必要があると主張.	(記事) 5 位/ (サイト) 1 位/ 禁煙を強く推奨. 喫煙は人体に悪影響を及ぼすだけであると主張.
(記事) 9 位/ (サイト) 14 位/ 喫煙は認知症の発症率を上げる可能性があることを喫煙者に警告.	(記事) 9 位/ (サイト) 9 位/ 喫煙しないことが最も肺がんになりにくい方法であると主張.
サブプライムローン (Subprime lending)	
(記事) 1 位, 4 位, 6 位, 9 位/ (サイト) 2 位/ 日本の大学の経営学教授のブログ. 日本市場はサブプライム問題に対して迅速な対応ができなかったことを指摘.	(記事) 1 位, 4 位/ (サイト) 3 位/ 連邦準備銀行は当初, サブプライム問題を深刻な問題として受け止めていなかったことを指摘.
(記事) 3 位/ (サイト) 17 位/ 日本の経済アナリストのブログ. 誰も不動産の価格が下落するとは思っていなかったため, サブプライム問題の影響がより拡大したと指摘.	(記事) 3 位/ (サイト) 1 位/ サブプライムローン利用者は対策のしようがなかった. 貸手側に大きな責任があると指摘.
(記事) 10 位/ (サイト) 18 位/ 近年の金融危機やサブプライム問題を引き起こした連邦準備制度理事会を批判.	(記事) 5 位/ (サイト) 19 位/ いつか住宅バブルは弾けるとわかっていながらも, 住宅バブルの影響でサブプライムローンを利用して家を購入した人が増加したことを指摘.

析する方式を提案した. 評価トピックの日英ブログサイト集合を検索し, 検索した日英ブログの対照分析を行ったことで, 日英ブログの文化間差異を発見することができた.

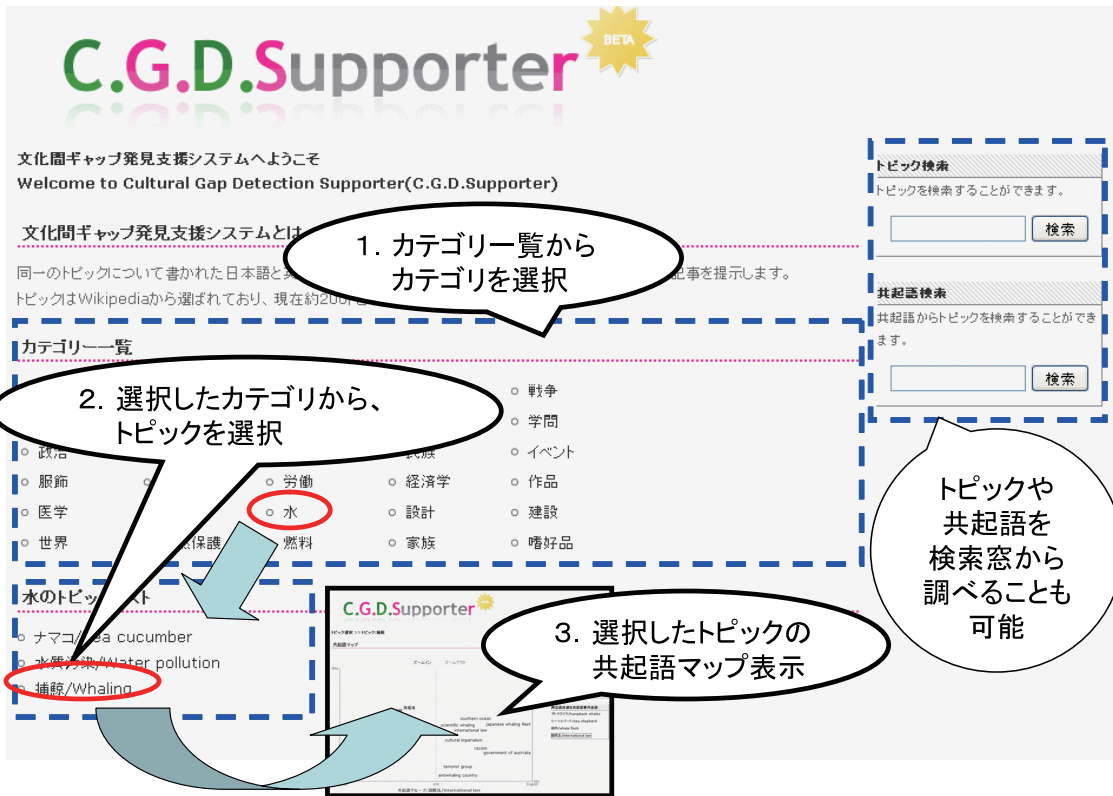
今後の課題として, 日英ブログから主観情報を多く含む箇所を抽出し, 日英ブログの文化間差異の新たな発見支援としたいと考えている.

文 献

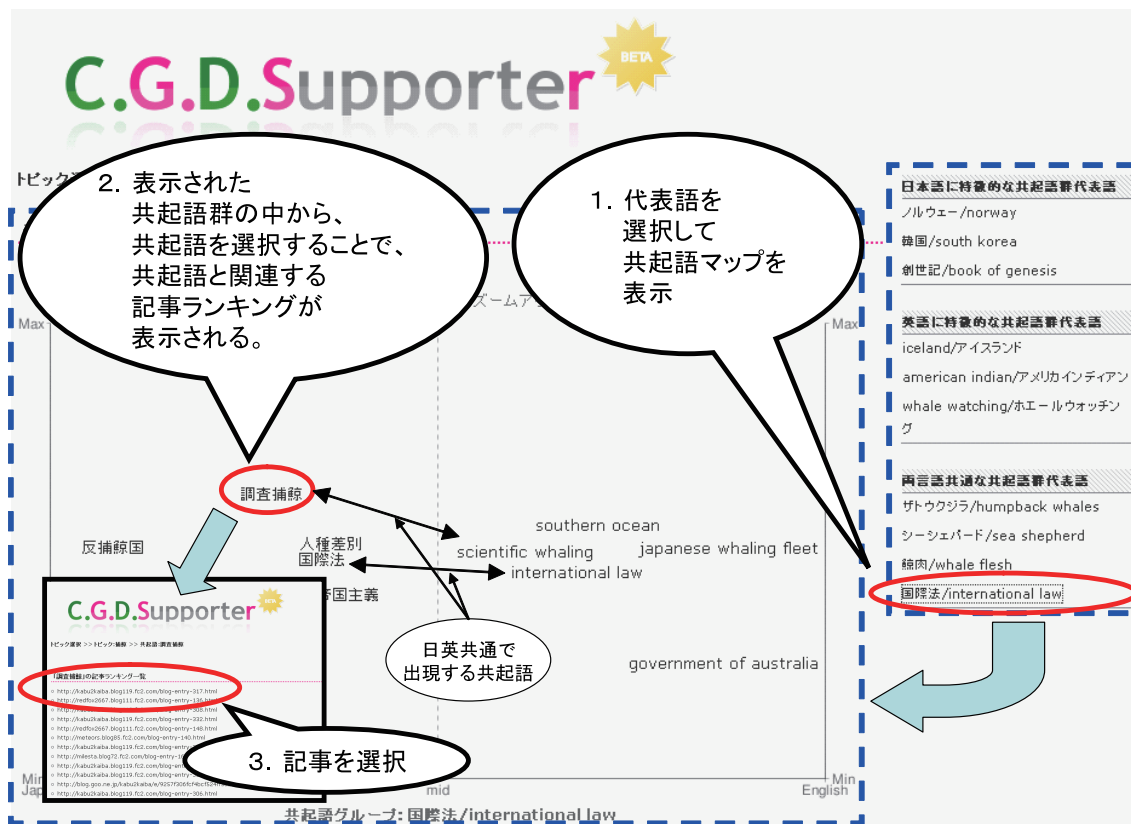
- [1] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. Wikipedia エントリとブログサイトの対応付けのための特定トピックのブログサイト検索. 電子情報通信学会第 19 回データ工学ワークショップ, 第 6 回日本データベース学会年次大会 (DEWS2008) 論文集, 2008.
- [2] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. 多言語 Wikipedia エントリを用いた特定トピックブログサイト検索と日英対照ブログ分析. 第 22 回人工知能学会全国大会論文集, 2008.
- [3] 川場真理子, 中崎寛之, 宇津呂武仁, 福原知宏. Wikipedia エントリとブログサイトの対応付けによる日本語ブログ空間の

トピック分布推定. 情報処理学会研究報告, Vol. 2008, No. (2008-NL-187), pp. 83-90, 2008.

- [4] 福原知宏, 宇津呂武仁, 中川裕志. 複数言語間の語彙出現傾向比較による言語横断型ウェブログ関心解析システムの開発. 言語処理学会第 13 回年次大会「大規模 Web 研究基盤上での自然言語処理・情報検索研究」ワークショップ論文集, pp. 40-43, 2007.
- [5] 吉岡真治. 複数のニュース源の差異を考慮したニュース分析の研究. 言語処理学会第 13 回年次大会「大規模 Web 研究基盤上での自然言語処理・情報検索研究」ワークショップ論文集, pp. 27-20, 2007.



(a) トピック選択～共起語マップ表示



(b) 共起語マップ～関連ブログ記事 URL リスト表示

図 3 文化間ギャップ発見支援システムの使用法

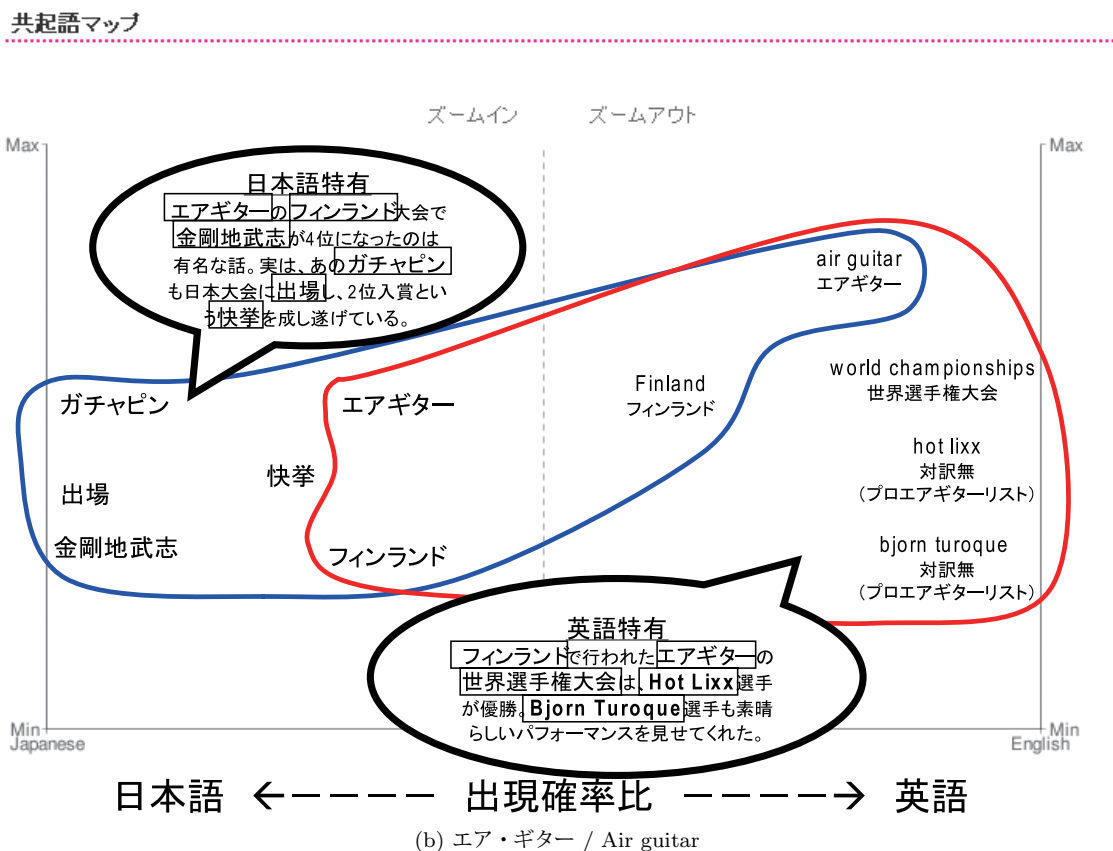
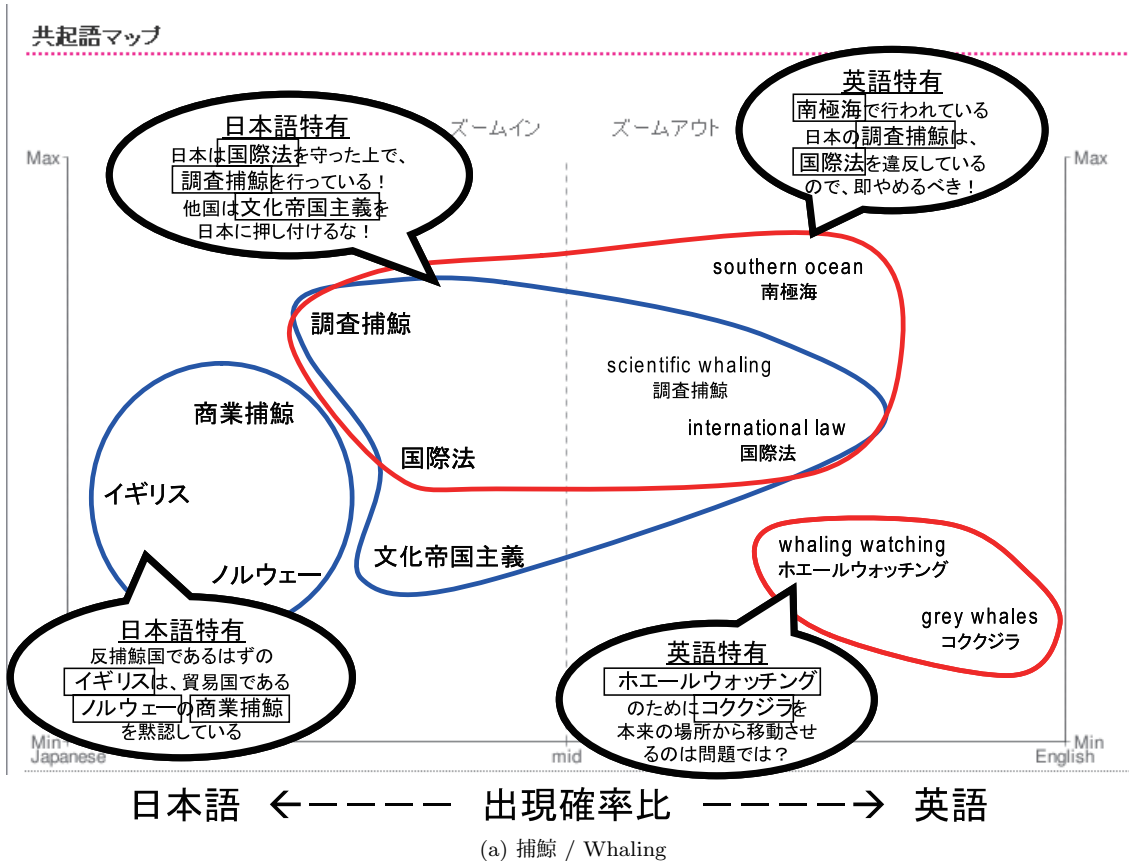


図 4 日英ブログから抽出した共起語例を用いた共起語マップ (「捕鯨」, 「エア・ギター」)