社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2003－DBS－131 (14)
2003／7／16

多言語情報源を対象とした意味的連想検索実現のためのメタデータ自動翻訳方式

大橋英博[†] 清木康[††] 石黒晶子[†††]

本稿では，翻訳辞書群を用いた信頼度依存の多数決処理によって，専門分野のメタデータを自動翻訳する方式を示す．本方式は，複数の翻訳辞書に信頼度を設定し，信頼度を反映させた多数決処理により翻訳語を決定することにより，翻訳精度を向上させる方式である．また，本方式を多言語情報源を対象とした意味的連想検索に適用する方式を示す．本方式を意味的連想検索に適用することにより，意味空間の言語的制約を取り除くことができ，専門家の知識の利用範囲を大幅に拡大させることが可能となる．

An Automatic Metadata Translation Method for Semantic Associative Search in Multilingual Information Resources

HIDEHIRO OHASHI,[†] YASUSHI KIYOKI[††] and AKIKO ISHIGURO[†††]

We present a method of automatic metadata translation by majority decision reflecting reliability of several translation dictionaries. In this method, we set reliability to several translation dictionaries, and select a translated word through the majority decision process. By using this method, we can improve precision of translation. We also present an application of the automatic metadata translation method to semantic associative search. We apply this method to extend the scope for using the semantic space which is originally created depending on the specific language. By this extension, we obtain a semantic associative search environment supporting multiple language for representing queries and retrieval-candidate information resources.

1. はじめに

大量の文書情報を格納したドキュメントデータベースから情報を検索する手法として，パターンマッチングによる検索手法が普及している．しかし，検索者の求める情報を取得するためには，パターンマッチングによる検索手法では不十分である．大量のドキュメントの中から検索者にとって価値のある情報を見つけ出すためには，検索者の文脈を解釈する機構を有する検索手法が必要となる．検索者の文脈を解釈する機構を有する検索手法として，意味的連想検索方式 [5,6,7] が提案されている．この方式は，専門家の知識を意味空間として体系化し，検索者が入力した検索キーワードと各ドキュメントとの意味的な類似性を計量する空間として使用する点に特徴がある．この方式により，検索者の文脈を専門家の知識体系に基づいて解釈し，検索キーワードと各ドキュメントとの意味的な類似性に基づいて，ドキュメント検索を行うことが可能となる．

意味的連想検索機構は専門知識を意味空間として体系化する．専門知識を表現するために，ある一つの言語で記述する必要がある．この意味空間を用いた意味的連想検索を行う時には，専門知識を表現するために使用した言語を使用しなければいけないという制約を持つ．意味空間は専門家の知識を数学的手法を用いて体系化された高度な計量空間である．そのため，この空間は本来特定の言語環境よらず，あらゆる言語環境のドキュメント検索に適用できる計量空間である．この意味空間が一つの言語環境においてのみしか利用できない点が，多言語環境において高度なドキュメント検索を行う場合の制約となっている．

本研究では，多言語環境において，意味的連想検索を実現することを目的とする．そのために多言語情報源

† 慶應義塾大学 政策・メディア研究科
Graduate School of Media and Governance, Keio University
†† 慶應義塾大学 環境情報学部
Faculty of Environmental Information, Keio University
††† 株式会社レクサー・リサーチ
LEXER RESEARCH Inc.

を対象としたメタデータ自動翻訳方式を示す.

2. 本 方 式

本研究の本方式を以下に示す. 本方式では, 次の手順でメタデータの翻訳を行う.

(1) メタデータ翻訳に用いる辞書の選択および信頼度設定

(2) 対訳表作成

(3) 対訳表を用いたメタデータ自動翻訳

以下に, 各手順の詳細を説明する.

2.1 メタデータ翻訳に用いる辞書の選択方法および信頼度設定方式

メタデータの翻訳を行うための翻訳辞書を選択する. ここで, 意味空間を記述するために使用した言語を L_m, 検索キーワードおよび検索対象ドキュメントを記述している言語を L_l とする. 以下の方針に基づいて,L_m から L_l への翻訳を行う 3 種類の辞書を選択する.

- $Dict_s$: 当該分野において最も専門性の高く信頼性の高い翻訳辞書
- $Dict_b$: 当該分野において基本的な翻訳辞書
- $Dict_g$: 分野によらない一般的な翻訳辞書

ここで,$Dict_s$ は当該分野の専門用語を翻訳するために使用する辞書である.$Dict_b$ は当該専門分野の基本語を翻訳するための辞書である.$Dict_g$ は分野によらない一般的な単語を翻訳するための辞書である. 一般に, 単語の収録数は多い順に $Dict_g$, $Dict_b$, $Dict_s$ の順となる.

これらの $Dict_s$, $Dict_b$, $Dict_g$ を使用してメタデータの翻訳を行う. ここで,3 辞書の信頼度設定を行う. 信頼度設定はその辞書が当該分野においてどの程度の専門性をもつかによって行う. ここでは翻訳辞書 D の信頼度 r(D) を以下のように定義する.

$$r(D) = \begin{cases} 3 & \text{D=}Dict_s \text{ のとき} \\ 2 & \text{D=}Dict_b \text{ のとき} \\ 1 & \text{D=}Dict_g \text{ のとき} \end{cases}$$

当該分野において,$Dict_s$ は専門性の高い単語を翻訳できる辞書であるため,r($Dict_s$) には最も高い信頼度 3 を設定する.$Dict_b$ は, 当該分野において, 基本的な単語を翻訳できる辞書であるため,r($Dict_b$)=2 を設定する.$Dict_g$ は分野によらない一般的な単語を翻訳する辞書であるため,r($Dict_g$)=3 を設定する.

2.2 対訳表作成方式

本方式では, 翻訳を行うための対訳表を作成する. 対訳表は表 2.2 に示す形式となっている. 対訳表は次の手順で作成する.

表 2.2 対訳表の形式

日本語	英語	優先度	辞書情報ビット列

(1) 翻訳辞書群の信頼度を用いて優先度計算表を作成する. 優先度計算表は図 1 に示す形式によって構成する. 優先度は次のように計算する.

優先度計算表

優先度	参照辞書	辞書情報ビット列		
7	A,B,C	1	1	1
6	A,B	1	1	0
5	A,C	1	0	1
4	A	1	0	0
3	B,C	0	1	1
2	B	0	1	0
1	C	0	0	1

辞書の信頼度
辞書A>辞書B>辞書C

図 1 優先度計算表

(a) 翻訳辞書ビット列辞書の信頼度の値により, 各辞書を各ビット列に対応させる. 信頼度 n の辞書を第 n ビットに対応させる. 例えば, 信頼度 3 の辞書は第 3 ビットに対応する. このビット列を翻訳辞書ビット列 DictBits と呼ぶことにする.

(b) W_m と W_l とが互いに対訳であるとき,$W_m \leftrightarrow W_l$ と書くことにする. 翻訳辞書ビット列を設定する. 翻訳元言語 L_m の中の単語 W_m を各辞書で調べ,L_l の単語 W_l を得たとする. このとき,W_m に対する翻訳語 W_l の翻訳辞書ビット列 DictBits は以下のように設定する.DictBits の第 n ビットの値 bit(n) を次のように定義する. ただし, 辞書 D において W_m が W_l の訳であることを $(W_m \leftrightarrow W_l) \in$ D と表す.

$(W_m \leftrightarrow W_l) \in D$ であることを A と表す.

$$bit(n) = \begin{cases} 1 & A \cap r(D) = n \\ 0 & \text{それ以外のとき} \end{cases}$$

この bit(n) によって DictBits の各ビットの値を設定する.

(c) 翻訳辞書ビット列 DictBits を 10 進数に変換した値を対訳 $W_m \leftrightarrow W_l$ の優先度 p($W_m \leftrightarrow W_l$) とする.

優先度 p($W_m \leftrightarrow W_l$) は次のように定義する.

$$p(W_m \leftrightarrow W_l) = priority(DictBits, W_m, W_l)$$

ここで,priority(DictBits, W_m, W_l) は,W_m と W_l, および優先度計算表によって定まる DictBits(2 進数) を,10 進数に変換する関数とする. このように設定した優先度計算表を図 1 に示す.

(2) 言語 L_m のメタデータ群 W_{mi}(i=1...n) を用意する. この W_{mi} を対訳表の L_m の単語群として設定する.

(3) 翻訳辞書群から翻訳元言語 L_m の中の単語 W_m の訳である言語 L_l の単語 W_l を調べ, 対訳表中の W_l の欄に設定する.

(4) W_m の訳である W_l を参照した辞書から, 辞書情報ビット列 DictBits を作成する. 例えば,W_m に対する訳 W_l が $Dict_s$ の辞書から得られたとする. このとき,$W_m \leftrightarrow W_l$ の DictBits の r($Dict_s$) 列に 1 を立てる.

(5) 辞書情報ビット列 DictBits から優先度 p($W_m \leftrightarrow W_l$) を計算する

上記の (1)〜(5) の手順により, 対訳表を作ることができる. 表 2.2 にこの手順によって作成された対訳表の例を示す.

表 2.2 対訳表の例

日本語	英語	優先度	辞書情報ビット列
アルコール依存症	alcohol dependency	1	001
アルコール中毒	alcoholism	6	110
アルツハイマー病	Alzheimer disease	7	111
アレルギー	allergy	7	111
アレルギー性鼻炎	allergy rhinitis	1	001
イタイイタイ病	Itai-Itai disease	4	100
インターフェロン	interferon	1	111
インフルエンザ	influenza	1	111

2.3 対訳表を用いたメタデータ自動翻訳方式

対訳表の優先度を用いたメタデータの自動翻訳方式を示す. メタデータの翻訳においては, 言語 L_l を翻訳元言語とし, 言語 L_m を翻訳先言語とする. 対訳表中の各エントリーには翻訳元言語 L_l の単語 W_l と, 翻訳先言語 L_m の単語 Wm の対訳と翻訳辞書ビット列, および優先度が設定されている. この優先度の最も大きい訳を使用して, メタデータの自動翻訳を行う.W_l の訳 W_{sm} を次のように決定する.

(1) 対訳表の中で, 言語 L_l の中の単語 W_l に対応する言語 L_m 中の単語が n 個存在したとする. このとき言語 L_m 中の単語群を W_{mi}(i=1...n) とする.

(2) $W_l \leftrightarrow W_{mi}$ の翻訳に関する優先度 p は次のように表す.

$$p(W_l \leftrightarrow W_{mi})(i = 1...n)$$

(3) W_{mi}(i=1...n) の中で優先度 p($W_l \leftrightarrow W_m$) が最大となる単語 W_m を W_l に対する訳 W_{sm} とする.

$$W_{sm} := W_{mi} \quad in \quad max(p(W_l \leftrightarrow W_{mi}))$$
$$(i = 1...n)$$

図 2.3 に対訳表の優先度を使用したメタデータの自動翻訳の例を示す. ここでは,L_m を日本語,L_l を英語としている.L_l 中の単語「anorexia」に対する日本語訳を, 優先度に従って決定している. ここでは, 英単語「anorexia」に対して日本語「食欲不振」が最も大きな優先度 3 となっている. これにより,「anorexia」は「食欲不振」と翻訳される.

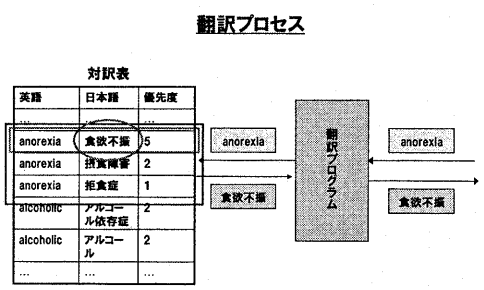

図 2.3 翻訳の例

3. 本方式を適用した多言語情報源を対象とした意味的連想検索システム

本方式を適用した意味的連想検索システムの実現例を示す. この例では, 医療分野の専門辞書群を使用して医療意味空間を作成した. この意味空間は日本語で記述されている. また, 検索対象ドキュメントおよび検索キーワードは英語によって表現されている. つまり,L_m=日本語,L_l=英語となる. このような条件で, 日

本語で記述された専門知識である医療意味空間を使用して, 英語で記述されたドキュメントの検索を実現可能とした. 図 3.1 にシステム構成図を示す.

3.1 システム構成図

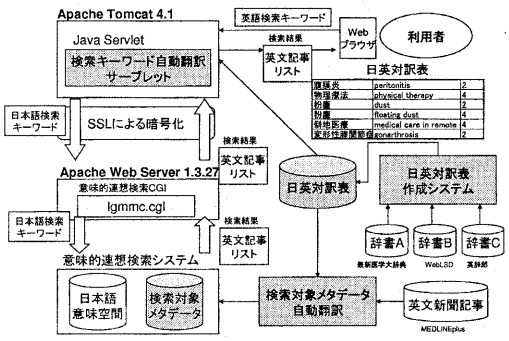

図 3.1 多言語情報源を対象とした意味的連想検索システム

3.2 対訳表作成機能

実現システムで使用した対訳表の作成には, 医学大辞典 [2], WebLSD [3], 英辞郎 [1] の 3 辞書を使用した. ここでは, 医学大辞典を医療分野の専門辞書 ($Dict_s$) とした.WebLSD を医療分野の基本辞書 ($Dict_b$) とした. 英辞郎を分野によらない一般的な翻訳辞書 ($Dict_g$) とした. これらの辞書を使用して, 医療分野における対訳表を作成した. 対訳表の言語 L_m (日本語) の単語群として, 医学意味空間の基本語群 691 語を使用した. この結果,2506 エントリーの医学分野の対訳表を作成した. 前述の表 2.2 は今回作成した医療分野の対訳表の例となっている.

3.3 検索対象メタデータ自動翻訳機能

検索対象ドキュメントは MEDLINplus [4] の医療に関する英文で書かれた新聞記事 578 件を使用した. この新聞記事から以下の方法で検索対象メタデータを抽出し, 自動翻訳を行った.

(1) 対訳表の言語 L_l 欄にある単語 (英単語) の集合を W_{li} (i=1...n) とする.

(2) 英文記事 a_i の中の単語を調べる.$w_l \in W_{li}$ (i=1...n) を含んでいた場合,a_i のメタデータとして w_l を割当てる.

(3) w_l を言語 L_m (日本語) の単語 W_m に翻訳する. 本方式を適用し, 対訳表の中で w_l に対応する w_{mi} (i=1,..,n) の中で最も優先度の大きい単語を w_l に対する訳とする.

この方法を全ての英文記事に対して行う. これにより, 各英文記事に対して, 日本語の検索対象メタデータが付与される.

3.4 検索キーワード自動翻訳機能

検索キーワードの自動翻訳方法について説明する. この実現例では言語 L_l (英語) を使用して検索を行う. 言語 L_m (日本語) で表現された医療意味空間上で検索を行うために, 言語 L_l (英語) で表現された検索キーワードを, 言語 L_m (日本語) に翻訳する必要がある. 英語の検索キーワードを以下のように日本語検索キーワードに自動翻訳する.

(1) 検索キーワード K_s は対訳表の言語 L_l (英語) の単語群 W_l に含まれる単語とする.

(2) 対訳表の中に K_s に対応する言語 L_m (日本語) の単語が n 個存在したとする. この単語群を W_{mi}(i=1...n) とする. この W_{mi}(i=1...n) の中で, 最も優先度の大きい単語 W_{sm} を K_s に対する訳とする.

これにより, 英語の検索キーワードが日本語意味空間上の検索に使用できる日本語検索キーワードに翻訳される. 以下に, 実現システムによる検索例を示す. 図 3.4.1, 図 3.4.2 は本方式を意味的連想検索システムを適用したシステムの検索画面である. このシステムでは日本語の医療意味空間による意味的連想検索システムを使用して, 英語検索キーワードによる英文記事の検索が可能となっている. 図 3.4.1 では、検索キーワード入力画面である. 例では、検索キーワードとして「dust」を入力している. 図 3.4.2 では「dust」を本方式により自動翻訳を行い,「粉塵」に変換されている. この例では「dust」に対して「粉塵」が最も大きな優先度とを持っている. この結果、検索キーワード「粉塵」を使用して, 日本語の医療意味空間上で意味的連想検索を実行し,「粉塵」に意味的に近い英文ドキュメントを得ている. これらの英文記事には, 日本語メタデータが付与されている. この日本語メタデータは英文記事中に含まれる英語メタデータを本方式によって自動翻訳されたメタデータである. 本方式により, メタデータを自動翻訳することにより, 多言語情報源を対象とした意味的連想検索システムが実現可能となっている.

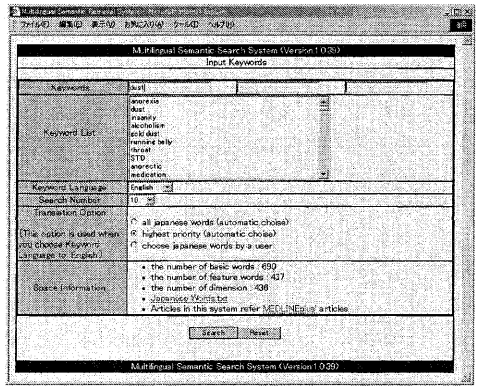

図 3.4.1 検索キーワード入力画面

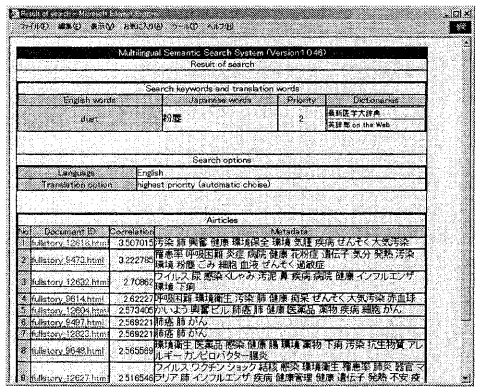

図 3.4.2 検索結果画面

4. 実　　験

ここでは本方式の翻訳精度を検証するための実験について述べる.

4.1 実験の概要

この実験では, 本方式によるメタデータの自動翻訳実験を実施した. あらかじめ人手により準備したメタデータの正解データと, 本方式によって作成された翻訳結果とを比較して, 翻訳精度を検証した.

4.2 実験に使用した翻訳辞書

実験には, 上述の意味的連想検索システム実現例の同様に, 医学大辞典 [2], WebLSD [3], 英辞郎 [1] の 3 辞書を使用した. 同様に, 医学大辞典を医療分野の専門辞書 ($Dict_s$) とした.WebLSD を医療分野の基本辞書 ($Dict_b$) とした. 英辞郎を分野によらない一般的な翻訳辞書 ($Dict_g$) とした. これらの辞書を使用して, 医療分野における対訳表を作成した. 対訳表の言語 L_m (日本語) の単語群として, 医学意味空間の基本語群 691 語を使用した. この結果,2506 エントリーの医学分野の対訳表を作成した.

4.3 正解データ

医療に関する 100 個の英語メタデータを用意し, それらに対して人手により日本語訳を割り振った. この 100 個のメタデータの組を正解データとして使用した. 正解データ例を表 4.3 に示す.

表 4.3 正解データの例

No.	英語	日本語
1	AD	アトピー性皮膚炎
2	AIDS	エイズ
...	...	...
100	alcoholism	アルコール中毒

4.4 翻訳精度の計算方法

翻訳精度の計算を以下のように行った. 自動翻訳によって得られた単語のうち, 正解データと一致した単語の数を, 正解データ数 (=100) で割った値を翻訳精度とした.

c: 正解データと一致した単語数
n: 全正解単語数 (=100)

翻訳精度 p = c / n * 100

4.5 実 験 方 法

次の 5 項目について実験を行った.

(1) 実験 1 : 分野によらない一般翻訳辞書 ($Dict_g$) のみによる自動翻訳
一般翻訳辞書 ($Dict_g$) のみを使用して作成した対訳表を用いて自動翻訳を行った. 今回の実験では一般翻訳辞書として英辞郎を使用した.

(2) 実験 2 : 当該分野の基本辞書 ($Dict_b$) のみを使用した自動翻訳
当該分野の基本辞書 ($Dict_b$) のみを使用して作成した対訳表を用いて自動翻訳を行った. 今回の実験では当該分野の基本辞書として WebLSD を使用した.

(3) 実験 3 : 当該分野の専門辞書 ($Dict_s$) のみを使用した自動翻訳
当該分野の専門辞書 ($Dict_s$) のみを使用して作成した対訳表を用いて自動翻訳を行った. 今回の実験では, 当該分野の専門辞書として医学大辞典を使用した.

(4) 実験 4 : 上記の 3 辞書を使用した多数決による自動翻訳
分野によらない一般翻訳辞書, 当該分野の基本

辞書, および当該分野の専門辞書の 3 辞書を使用して対訳表を作成し, 自動翻訳を行った. この実験では,3 辞書の多数決によって訳の優先度を決定し, その優先度の最も大きい訳を選択して自動翻訳を行った. 実験 4 の優先度は, 当該対訳を得ることができた辞書の数を設定した. つまり, 対訳 $W_m \leftrightarrow W_l$ が 3 辞書から得られたとき, 優先度を 3 に設定した. 同様に対訳 $W_m \leftrightarrow W_l$ が 2 辞書から得られたときは, 優先度を 2 に設定した. 対訳 $W_m \leftrightarrow W_l$ が 1 辞書からのみ得られたときは, 優先度を 1 に設定した.

(5) 実験 5 : 本方式-信頼度依存の多数決による自動翻訳

本方式による自動翻訳を行った.3 辞書に信頼度を設定し, その信頼度と多数決によって優先度を決定し, 優先度が最も大きい訳を選択して自動翻訳を行った.

4.6 実験結果 1-翻訳精度の比較

実験結果を図 4.6 に示す. 図 4.6 はそれぞれの実験の翻訳精度を示している. この図から, 分野によらない一般辞書だけを使用した翻訳の場合, 翻訳精度は 40 となっている. 医療分野の基本辞書だけを使用した場合は 36, 医療分野の専門辞書を使用した場合では 37 となっている. これらの結果から, いずれの種類の辞書も単独で使用した場合では十分な翻訳精度が得られないことが明らかになった.3 辞書の多数決によって, 翻訳精度は 80 と大幅に向上している. さらに本方式である辞書に信頼度を設定し, その上で多数決を行う方式では, 翻訳精度は 85 と最も高い値を示している.

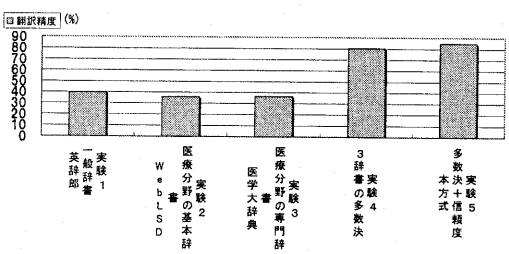

図 4.6 実験結果　翻訳精度の比較

4.7 実験結 2-不正解データの内容の分析

次に各実験で不正解となった単語の内容について調べた. この結果を図 4.7 に示す. 図 4.7 では不正解となった単語を, 誤訳した単語と対訳表で発見できなかった単語とに分け, それぞれの全正解単語数に対する割合を「誤訳率」と「未発見率」として示している. 分野によらない一般辞書だけを使用した場合, 誤訳率は 20, 未発見率は 40 となっている. 医療分野の基本辞書では, 誤訳率 6, 未発見率 58 となっている. 医療分野の専門辞書では, 誤訳率 2, 未発見率 61 となっている. このように, 辞書の専門性が高くなるにつれ, 誤訳率が低くなる一方で, 未発見率が高くなっている.3 辞書の多数決により翻訳を行った場合では, 未発見率は 0 になった. 一方で誤訳率は 20 となり, 分野によらない一般辞書の誤訳率と同じ値になっている. 本方式では, 未発見率は同様に 0 となっており, 誤訳率が 15 となっている. 本方式では, 多数決処理のみを行う場合に比べて, 誤訳率が減少している.

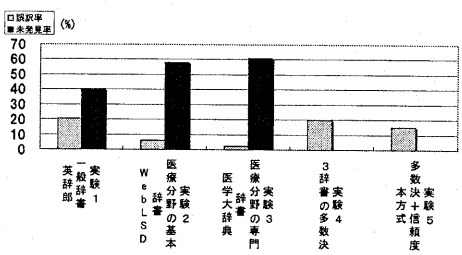

図 4.7 実験結果　不正解データの内容の分析

5. 考　　察

実験結果についての考察を述べる. 実験結果より, それぞれの 1 翻訳辞書だけ使用した場合では, 十分な翻訳精度が得られていない.3 辞書を使用した多数決による翻訳方法では, それぞれ 1 辞書だけを使用した場合より, 翻訳精度が大きく向上している. 本方式による翻訳により, 多数決のみよる翻訳に比べてさらに翻訳精度が向上している. この結果から, 本方式による自動翻訳の有効性が確認できた.

不正解データの内容を分析した結果から, 単独の辞書を使用した場合では, 使用する辞書の性質により単語の誤訳率と未発見率が異なることが確認された.3 辞書の多数決による翻訳では, 単語の未発見率は 0 になったものの, 単語の誤訳率が上昇している. これは,3 辞書

を合わせて使用したことにより, 分野によらない一般辞書の影響を受けたためと思われる. このことから, 複数の辞書を単純に合わせて使用するだけでは, 誤訳率を上昇させてしまうことが明らかになった. 複数の辞書の合わせて単語の未発見率を下げつつ, 誤訳率の上昇を抑えるには, 本方式による翻訳方式が有効である. 本方式では, 翻訳に使用する辞書を, 分野によらない一般辞書, 当該分野の基本辞書, 当該分野の専門辞書の 3 種類に分類し, それぞれ適切な信頼度を設定する方式となっている. この信頼度を考慮した優先度計算表を作成し, これを使用して単語の翻訳を行っている. このため, 単純に多数決によって翻訳語を決定する方法と比べ, きめ細かい優先度の設定が可能となっている. 例えば, 対訳表の中には, 英単語「pulsy」に対して日本語訳「脈」,「気分」が 2 語存在している.「pulsy ↔ 脈」の優先度が 4 (辞書情報ビット列 100),「pulsy ↔ 気分」の優先度が 1 (辞書情報ビット列 001) となっている. 正解が「pulsy ↔ 気分」としたとき, 本方式では, 最も大きな優先度 4 を持つ「pulsy ↔ 気分」が選択されるので必ず正解が得られる. 多数決のみによる翻訳ではどちらの対訳も辞書情報ビット列中の 1 の数は 1 であるため, どちらも優先度が 1 となる. この場合, どちらの訳が適切か判断するための情報がないため,2 組の対訳の中から 1 組の対訳が非決定的に選択される. これらの理由から, 多数決による翻訳に比べ, 本方式は高い翻訳精度を実現できる.

本方式による翻訳でも正しい訳が得られない場合がある. 今回の実験では本方式による自動翻訳を行ったとき,15 個の英単語に対して正しい訳を得ることができなかった. 表 5 に例を示す. 英単語「anorectic」には「食欲不振」と「摂食障害」の 2 個の日本語訳が存在している. この例では辞書 WebLSD からのみ訳が見つかっており, 辞書情報ビット列も同じ値となっているので, 同じ優先度となっている. この場合, 優先度が同じであるため, 本方式では訳を 1 語に決定できず, どちらかの訳を非決定的に選択することになる. このように, 辞書情報ビット列が同じ値を持つ場合には, 同じ優先度を持つため, 本方式でも正確に翻訳が出来ない. この問題を解決するためには, 各辞書内での使用頻度を考慮した優先度の設定を行う必要がある. 辞書によっては, 訳を使用頻度順に並べて返すものがある. このような辞書を利用すれば, 同じ辞書から複数の訳を取得した場合でも, 使用頻度に応じて異なる優先度を設定することが可能となる.

表 5 本方式でも正確に翻訳できない単語の例

日本語	英語	優先度	辞書情報ビット列
食欲不振	anorectic	2	010
摂食障害	anorectic	2	010

6. ま と め

本稿では, 専門分野におけるメタデータの自動翻訳方式を示した. 本方式では信頼度を設定した複数の辞書の多数決により自動翻訳を行う. この方式により, 複数の辞書の多数決のみによって自動翻訳を行う方法に比べて精度の高い自動翻訳を実現した. また, 本方式を適用した多言語情報源を対象とした意味的連想検索システムを実現した. 本方式を意味的連想検索システムに適用することで, 意味空間の言語的制約を取り除き, 専門家の知識の利用範囲を拡大することが可能となる.

参 考 文 献

1) 英辞郎, アルク,2002
2) 最新医学大辞典 CD-ROM 第 2 版, 医歯薬出版, 1997
3) ライフサイエンス辞書 WebLSD , http://lsd.pharm.kyoto-u.ac.jp/WebLSD/ , 京都大学大学院薬学研究科 医療薬理学分野, 2003
4) MEDLINEplus, http://www.nlm.nih.gov/medlineplus/, A service of the U.S. National Library of Medicine and the National Institutes of Health, 2003
5) Kiyoki, Y., Kitagawa, T. and Hayama, T.: "A metadatabase system for semantic image search by a mathematical model of meaning", ACM SIGMOD Record, vol. 23, no. 4, pp.34–41, 1994.
6) Kiyoki, Y.,Kitagawa, T. and Hitomi, Y.: "A fundamental framework for realizing semantic interoperability in a multidatabase environment", Journal of Integrated Computer-Aided Engineering, Vol.2, No.1, pp.3–20, John Wiley & Sons, Jan. 1995.
7) 清木 康, 金子 昌史, 北川 高嗣: "意味の数学モデルによる画像データベース探索方式とその学習機構", 電子情報通信学会論文誌,D-II,Vol.J79-D-II,No.4,pp.509–519, 1996.