平成１４年度特許情報に関するデータ蓄積等補助事業

報告書

日本特許文献の機械翻訳の性能・精度向上に関する調査・研究

―機械翻訳辞書構築支援システムに関する基礎的研究に係る開発―

平成１５年3月３１日

財団法人 日本特許情報機構

この事業は、競輪の補助金を受けて実施したものです。

1．概要・はじめに

本報告書は、財団法人日本特許情報機構（以下、Ｊａｐｉｏという。）が日本自転車振興会より「平成１４年度自転車等機械振興事業に係る補助金」の交付を得て行なった機械翻訳用辞書構築支援システムの原型システム開発およびそのシステムの効果についてまとめたものである。

1.1. 調査の目的

広く工業所有権情報の利用者に対して、国内外の工業所有権情報の利便性を向上し、翻訳コストの低減を図るため、工業所有権に特化した機械翻訳システムを研究し、もって機械関連産業の振興に寄与することを目的とする。

1.2. 事業の背景

日本の国際特許戦略の重要性が認識され、海外特許出願のニーズが高まっているが、その出願コストは高い。海外出願費用の半分近くが人手による翻訳費用であり、これを効率化することで費用を低減することが求められている。

加えて日本の特許情報を外国においても広く利用してもらうことが重要である。この観点では翻訳費用の低減に加えて用語・表現の統一、翻訳期間の短縮が求められている。

1.3. 翻訳の課題

翻訳を効率的に行う手段として機械翻訳の導入が試行されており、外国文献の概要理解支援など有効に利用されはじめている。特許文献の翻訳に関しても特許庁の特許電子図書館において特許明細書の機械翻訳サービスが始められ、日本語がわからない人への支援ツールとして役立てられている。しかし、機械翻訳された文章だけで技術内容が理解できる程に機械翻訳が十分な精度を有しているとは言いがたい。各社の専門用語辞書は特許の技術分野とは無関係に作られているため、専門用語の訳語の選択がうまく行かないことがある。特許文献の翻訳には、分野毎に特有の用語の理解および適切な用語の選定能力および特許特有の表現・記述方法の解析が不可欠である。

1.4. 事業の内容

Ｊａｐｉｏでは長年にわたって特許文献の抄録作成業務（日本語－英語の翻訳業務）を行なっており、その過程で分野毎の適訳を調査し翻訳している。この成果物を活用することにより、特許特有の用語とその適切な対訳を選択して、効率良く特許特有の分野別専門用語辞書を構築するための支援システムに関する基礎的な調査・研究を行なった。

1.5. 成果概要

概念検索と対訳候補の抽出機能を組み合わせることで翻訳に迷った用語の３割程度が過去の文献から容易に探し出せることが分かった。このことは翻訳者の能率をあげるだけでなく、翻訳と同時に辞書構築を行なえる可能性を示している。

本研究の成果を活用することにより、外国語出願の多いユーザーは、日英対訳文書から効率的に固有の特許用語辞書を構築できる可能性がある。

1.6. 将来計画

この成果をさらに発展させ、効率良く特許特有の分野別専門用語辞書を構築出来るシステムを実用に供すべく改良を重ねる予定。

1.7. 謝意

本システム開発および調査は「平成１４年度自転車等機械振興事業に係る補助金」を得て行なった。また、概念検索についてはＧＥＴＡ（注１）と茶筌（注２）を利用させていただいた。有益なソフトを開発公開して下さった関係各位に感謝いたします。

注１：汎用連想計算エンジンＧＥＴＡは情報処理振興協会（ＩＰＡ）が実施した「独創的情報技術育成事業」の研究成果です。

注２：奈良先端科学技術大学院大学自然言語処理学講座 からリリースされた、 フリーの日本語形態素解析器です。

2．構築システム概要

2.1. 目的

Ｊａｐｉｏで行なっている特許文献の抄録作成（日本語－英語の翻訳）において特許文献特有の翻訳が困難な用語は、適訳を調査し、的確な翻訳を実施している。この作成行程において過去の成果物を活用することにより効率良く対訳を調査すること及び、その結果を簡単に効率良く特許特有の分野別専門用語辞書として蓄積することを目的とする。

2.2. 実現方法の検討

特許特有の分野別専門用語辞書構築システムを検討するにあたって３つの手法を検討した結果、次の、「(1)の翻訳に迷う用語を過去の対訳で会話的に解決する方法」を採用した。

検討にあたって先ず、特定の機械翻訳システムに依存する方法は避けることとした。次に概念検索および対訳候補の抽出などに関して評価の定まっていない手法を導入することから、リスクの大きい一括処理的な方法を避けることとした。結果として会話的にシステムを操作して評価する上記の方法を選定した。以下に検討した 3 つの手法の考え方を示す。

2.2.1. (1) 翻訳に迷う用語を過去の対訳で対話的に解決する方法

翻訳者が翻訳に迷った場合、過去の対訳を利用して迷いを解決し、同時に辞書を構築する方法。

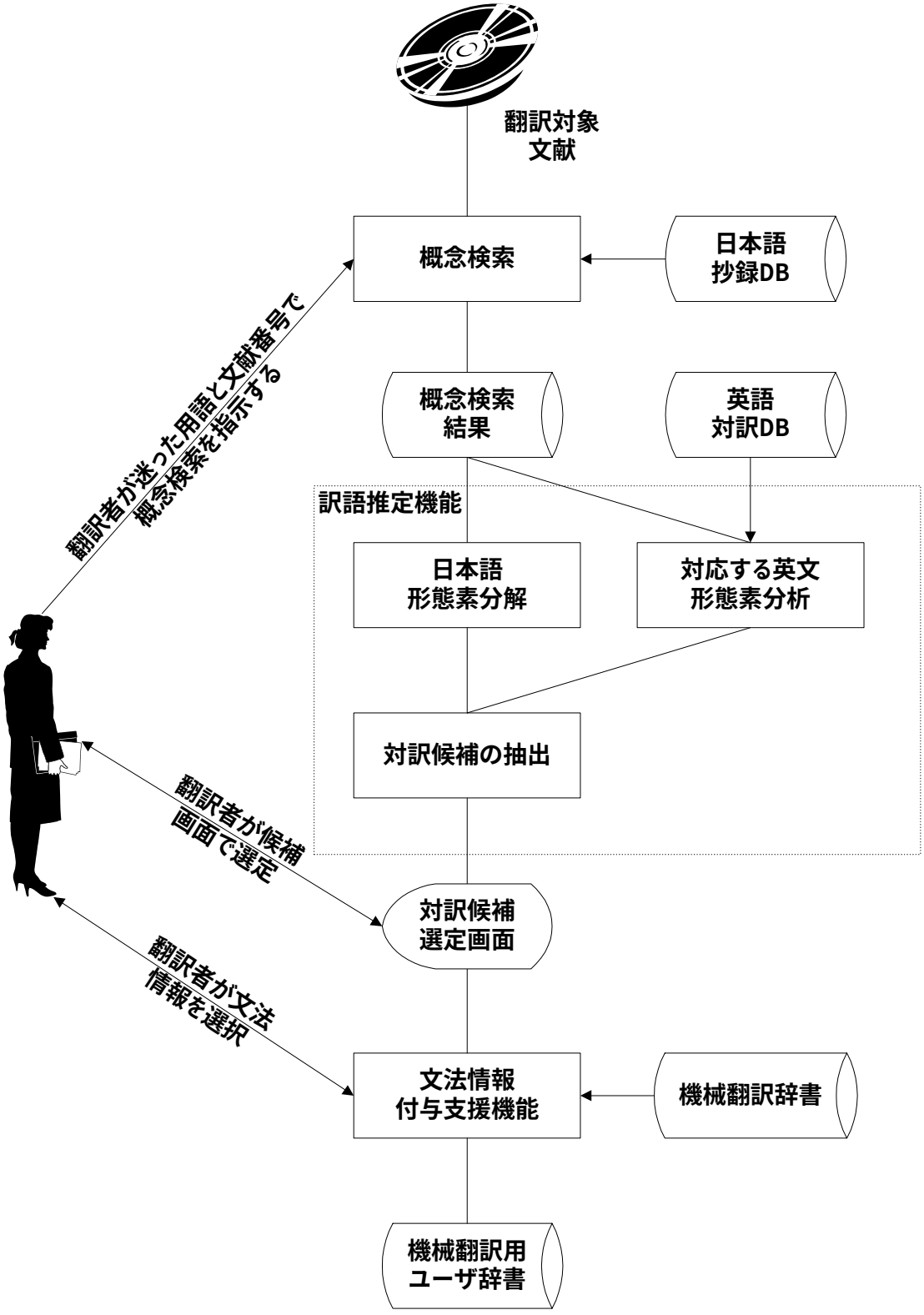
従来の単純な用語検索では翻訳に迷った用語を検索して多数の先例を見つけても、翻訳者が実際に参照するのは先頭近くの数件である。もし、参照した文献が別の分野であった場合は参考にならないことが多い。類似文献があった場合でも、訳語の選定には翻訳者による差異があり、少数の対訳を参考にして訳語を決めるのは品質確保の面で問題がある。

これを解決するために、検索の専門家でない翻訳者が過去の類似文献を容易に検索できる概念検索機能と、検索結果から、システムが統計的に訳語を選択する対訳候補の抽出機能を実現する。両機能は翻訳者が会話的に利用できるようにする。

他の方法に比べて会話的な要素が多く、一人の翻訳者が入力する件数は少ないが、日々の翻訳に利用してもらうことで翻訳者全員に機械翻訳辞書の構築への協力を得ることができる。

具体的な処理の流れは、「３．３システムの流れ」を参照されたい。

(1)翻訳に迷う用語を過去の対訳で会話的に解決する方法



2.2.2. (2) 過去の対訳から一括して機械翻訳辞書を作成する方法

過去の文献を分野別に分け、それを統計的に処理し、訳語を推定し、リストを作成し、それらを一括して辞書登録する。前の方法が会話的な手法なのに対し、この方法はバッチ的な方法である。

用語の出現頻度を利用して集計するので、過去の案件を能率的に辞書に登録できる。もし、対訳候補の抽出機能が十分な精度を持つならば辞書構築に非常に有効なシステムになる。

しかし、概念検索機能および対訳候補の抽出機能とも未だ確立された技術とは言えないことから、リスクを避ける意味で、今回は採用を見送った。

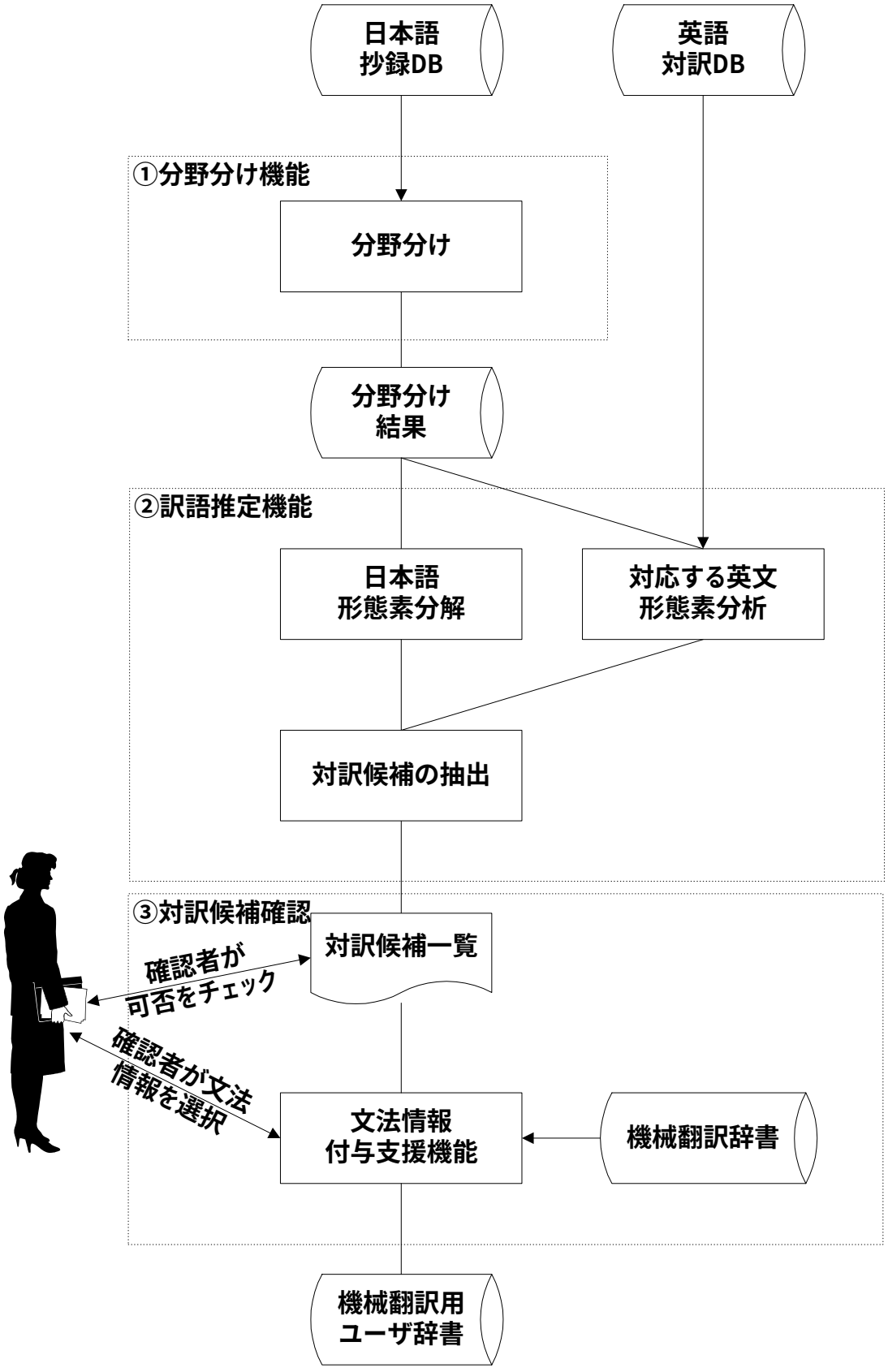
処理のポイントを記す。

① 分野分け機能は過去の日本語抄録文献を特許分類毎に分ける。ユーザ辞書は分野別に作成するので、以降の訳語推定機能および文法情報付与の処理は分割した分野毎に行う。

② 訳語推定機能は日本語抄録文献に対応する英語対訳を取得し、形態素解析によりそれぞれを単語に分解し、辞書に登録する対訳候補を統計的な手法で自動抽出して、対訳候補一覧を作成する。対訳候補一覧には、分野別辞書にふさわしくない一般的な用語を載せない。

③ 対訳候補確認は人手で行う。確認者は対訳候補一覧から分野別辞書にふさわしい用語を選定し、機械翻訳用のユーザ辞書に登録する。確認者は登録と同時に、誤訳の修正と文法情報を行う。

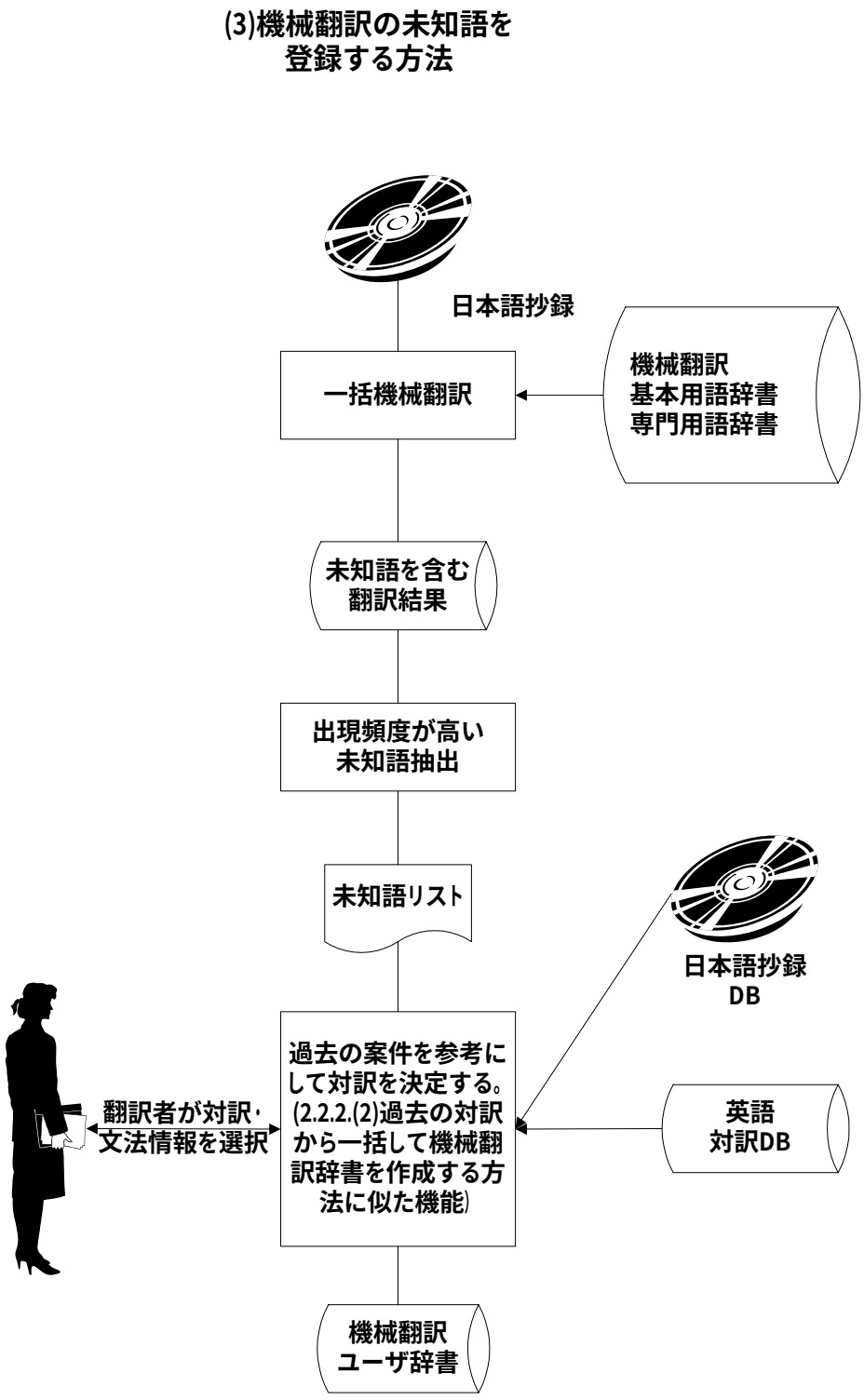
(2)過去の対訳から一括して機械翻訳辞書を作成する方法



2.2.3. (3) 過去の文献を機械翻訳して未知語を対訳から抽出する方法

過去の文献を一括で機械翻訳し、未知語および誤訳に対する正しい対訳を一括登録する方法。この方法は前にあげた二つの方法と考え方が異なり、特定の機械翻訳における辞書の追加で翻訳の精度を上げるようにチューニングする方法である。

この方法はバッチ的な手法を用いることで、能率的に辞書構築を行うことができるが、選定した機械翻訳が十分実用的である必要がある。また、選定した機械翻訳に特化した機械翻訳辞書を作ることになってしまう。従ってこの方法は避けることとした。



3．システム設計

翻訳者が過去の対訳を使って翻訳に迷う用語を解決し効率的に翻訳を進めると同時に辞書を構築する手法を設計・検証した。「2．2．1 翻訳に迷う用語を過去の対訳で対話的に解決する方法」参照。

3.1. 現状分析と解決案

翻訳者の作業を見ると、翻訳者は適切な訳語を選ぶため、翻訳時間のかなりの部分を翻訳に迷った用語の訳語調査に費やしているのが実情である。この事実をもとに問題点そして解決策を検討した。

3.1.1．従来の流れ

以下の様に翻訳に迷う用語で過去の文献を検索してその対訳を参照して翻訳のヒントとしていた。

① 用語検索によって迷った用語を検索。数十～数百件の回答を得る。
② タイトルを見て翻訳している文献の類似文献を選択。
③ 数件の対訳を参照して目的の用語の対訳を見つける。
④ 数件で参考になる文献が見つからなければ、専門書等、別の資料を探す。

3.1.2．解決すべき課題

従来の流れの問題点をあげる。

① 用語検索で類似文献を見つけるのが難しい。（翻訳者は検索の専門家でないため。）
② 参照した文献が別の分野であった場合は参考にならない。
③ 少数の対訳で訳語を決めるのは品質確保の面で問題がある。
（訳語の選定は翻訳者による差異がある）
④ 正しい対訳を見つけた場合でも、辞書登録に結び付かない。

3.1.3．解決案

これらの問題点を解決するために本システムは以下のシステムを装備する。

① 類似文献を容易に検索できる概念検索機能を備える。
② 統計的に訳語を選択する対訳候補の抽出機能を備える。
③ 対訳を簡単に辞書登録できる機能を備える。

翻訳に迷った用語を検索するということで、一人の翻訳者が入力する件数は少ないと思われるが、会話的で翻訳者が手軽に使えるようにすることで、翻訳者の利用頻度を上げ、翻訳者全員に機械翻訳辞書の構築に協力してもらうことを目指す。

3.2. システム機能概要

本システムに装備した 3 つの機能概要を説明する。

3.2.1．概念検索機能

調べたい用語とその用語を含む文献の番号を指定することにより調査対象文献の発明の名称と要約文を質問とし、過去の文献を検索して、指定された用語を含みかつ翻訳中の文献と似ている文献を検索する。翻訳者は検索の専門家ではないため、通常の用語検索では翻訳中の文献を効率良くみつけることができない。翻訳者を補助するため、概念検索機能を装備した。

概念検索エンジンとしてＧＥＴＡおよび用語の認定にあたって茶筌を利用した。調査を目的とする本システムの中核機能である概念検索は特定メーカに依存することを避けるため、ソースコードが公開されているＧＥＴＡおよび茶筌を利用した。

3.2.2．対訳候補の抽出機能

日本語と対訳の言葉の並びが一致しないこともあり、複数の概念検索結果を見比べて日本語要約中の目的の用語および対応する英文の中から対応する用語の対訳を探すのは効率的でない。これを効率的に行うため、検索結果の日本語と英語の対訳を統計的に処理して、目的の用語の対訳を推定する機能を装備した。

本機能のエンジンには東芝の機械翻訳システムの対訳候補の抽出機能に一部機能を追加したものを利用した。追加した機能は専用ＡＰＩ(Application Program Interface)および日本語１語対英語２語の対訳候補の抽出機能である。

3.2.3．文法情報付与支援機能

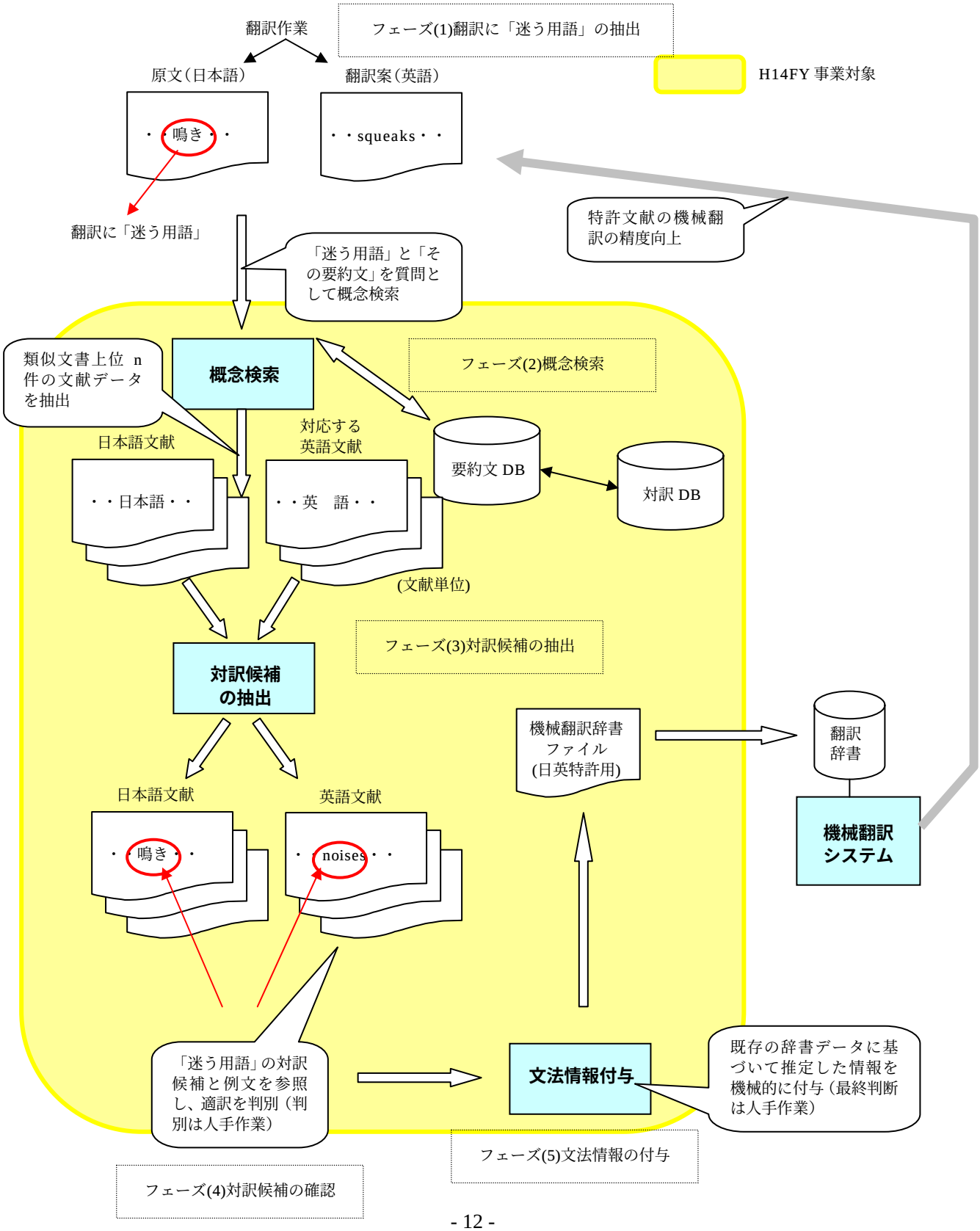
目的の対訳が見つかった時点で、すでに用語と対訳が選ばれているので、最小限のキー操作で辞書を構築出来れば多くの翻訳者の協力を得ることができる。これを実現するため、文法情報付与機能を装備した。

機械翻訳辞書に必要な文法情報の推定は東芝の機械翻訳エンジンに一部機能を追加したものを利用した。追加した機能は専用ＡＰＩである。

対訳候補の抽出および文法情報付与支援機能についてはソースコードが公開されているシステムが存在しない。今回採用した東芝の機械翻訳システムは特許庁の特許電子図書館で公開公報の日英機械翻訳ソフトウェアとして採用されており、特許向きと判断した。

3.3. システム全体の流れ

翻訳作業において発生した翻訳に「迷う用語」と、その用語を含む要約文で日本語要約文を概念検索して類似文献を抽出し、対応する英語対訳を利用して訳語を推定する。正しい訳語が選択できたら文法情報を付与して機械翻訳システムに取り込み可能な辞書形式ファイルを出力する。一連の行程を人的作業を含めて５つのフェーズに分けて説明する。以下に全体像を示す。



（1） 翻訳に「迷う用語」の抽出フェーズ

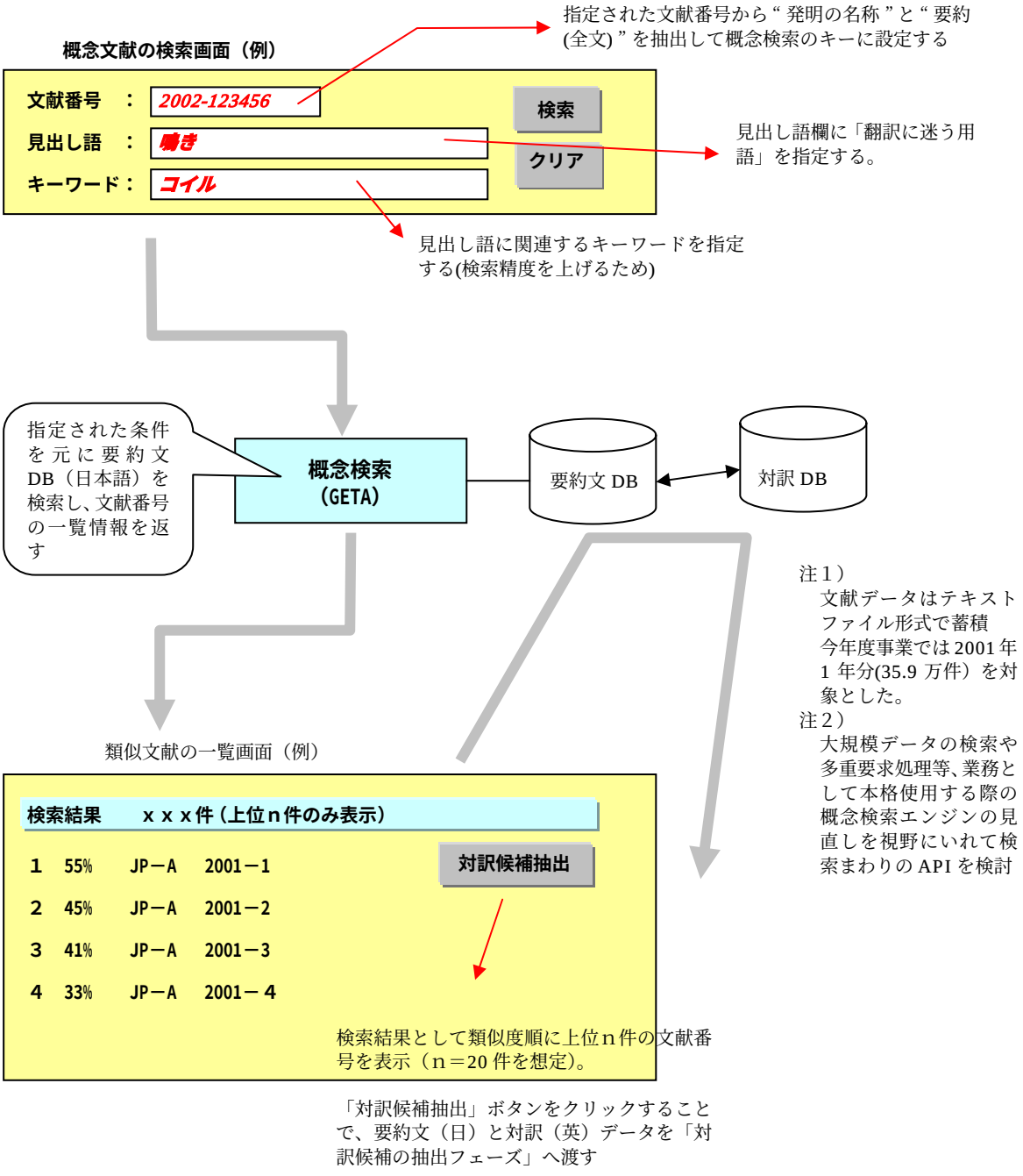
翻訳業務において翻訳者が翻訳文原稿の翻訳に迷う用語を抽出する。

今年度事業においては実際の翻訳で「迷う用語」を抽出せず、2001 年度の翻訳作業で発生した訳語の質問（用語調査票）および翻訳指導の対象となった用語（翻訳指導票）を合わせた約 1000 件を使って効果を検証した。

（2） 概念検索フェーズ

「迷う用語」を含んだ翻訳対象案件と類似した過去の対訳済み文献（日本語抄録と対の英文抄録）を、概念検索システムを用いて類似文献として抽出する。

検索対象データは日本の公開特許の要約文とその対訳英文とする。



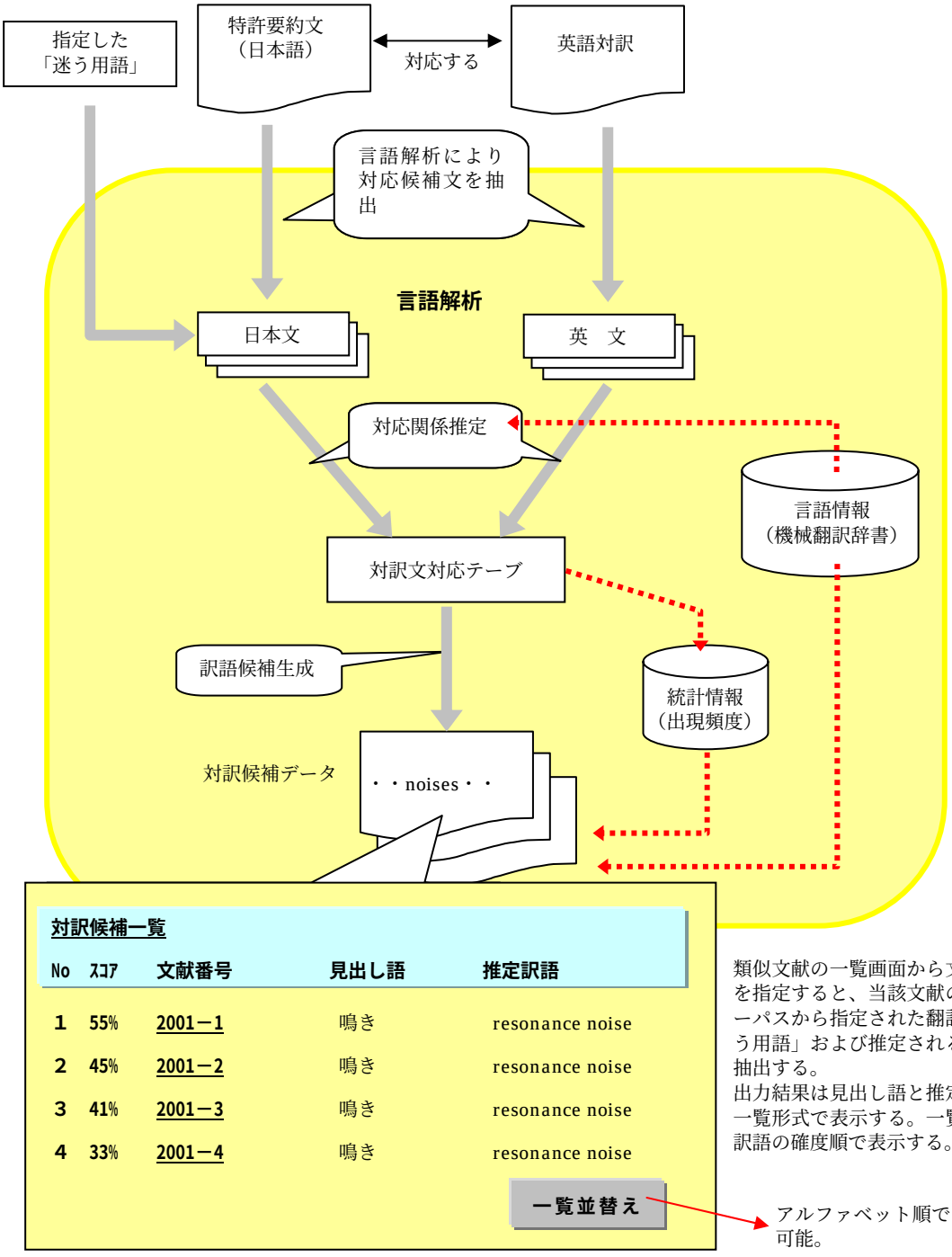
（３） 対訳候補の抽出フェーズ

抽出した類似文献と英語の対訳を対象に言語情報と統計情報により指定した翻訳に「迷う用語」の訳語候補を抽出する。

言語情報による判断は、既存対訳辞書を参照して語句の対応度に基づく対訳類似度から対訳を推定するものである。

統計情報による判断は、文書中の対応候補文に出現する頻度を基に対訳を推定するものである。

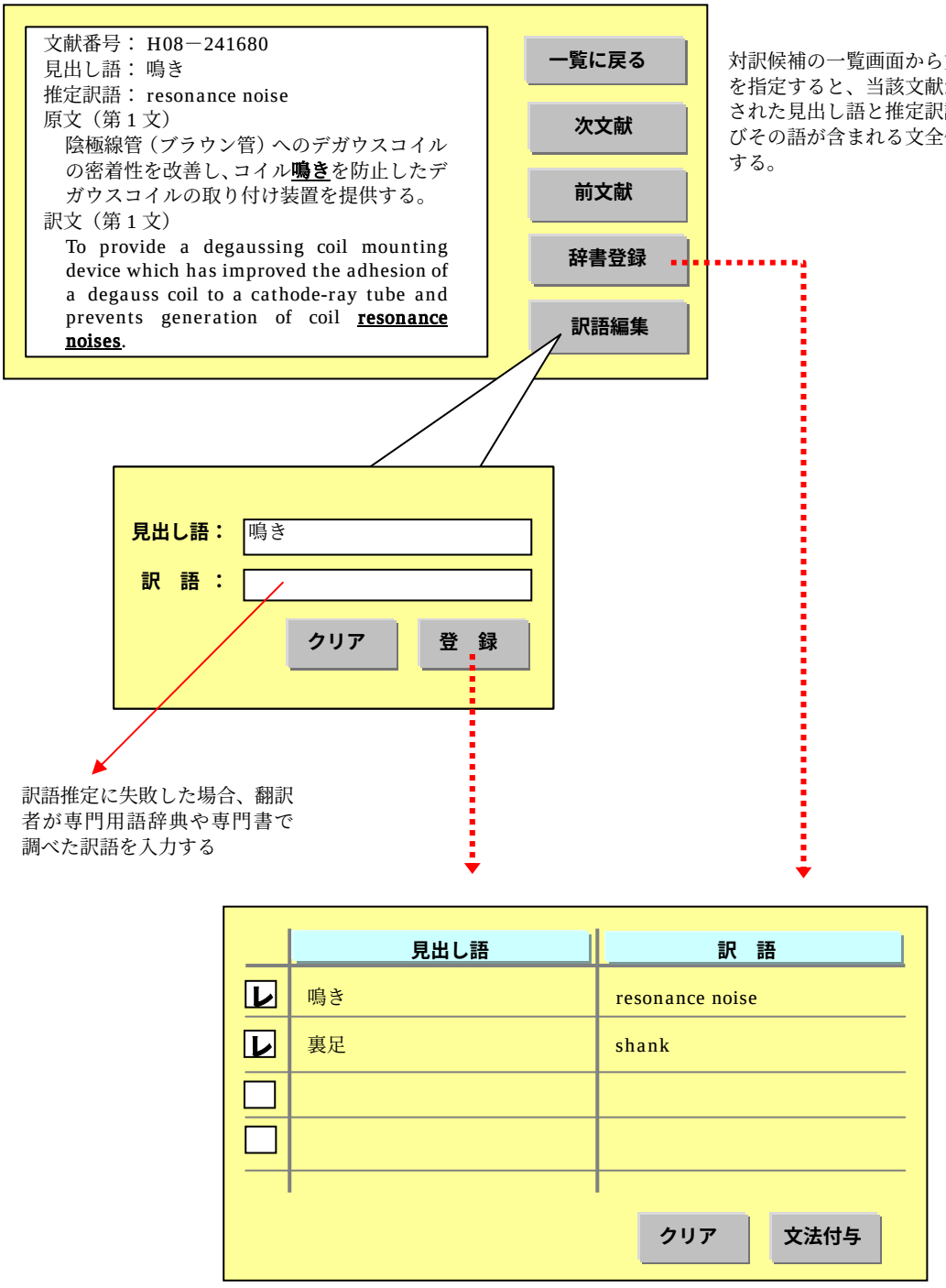
いずれの判断も対象文献の言語解析が必要となるが、この処理には東芝の機械翻訳システムの言語解析（形態素解析等）の機構を一部機能追加の上で利用した。



（４） 対訳候補の確認フェーズ

自動抽出された対訳候補を画面上で確認し評価する。確認作業（訳語選定の判断）は人手によるものとする。対訳を文脈の中で理解できるようにさらに対応する文全体を表示する。見出し語と推定した対訳をハイライト表示することで作業能率をあげる。対訳が合っていたら辞書登録ボタンを選択する。

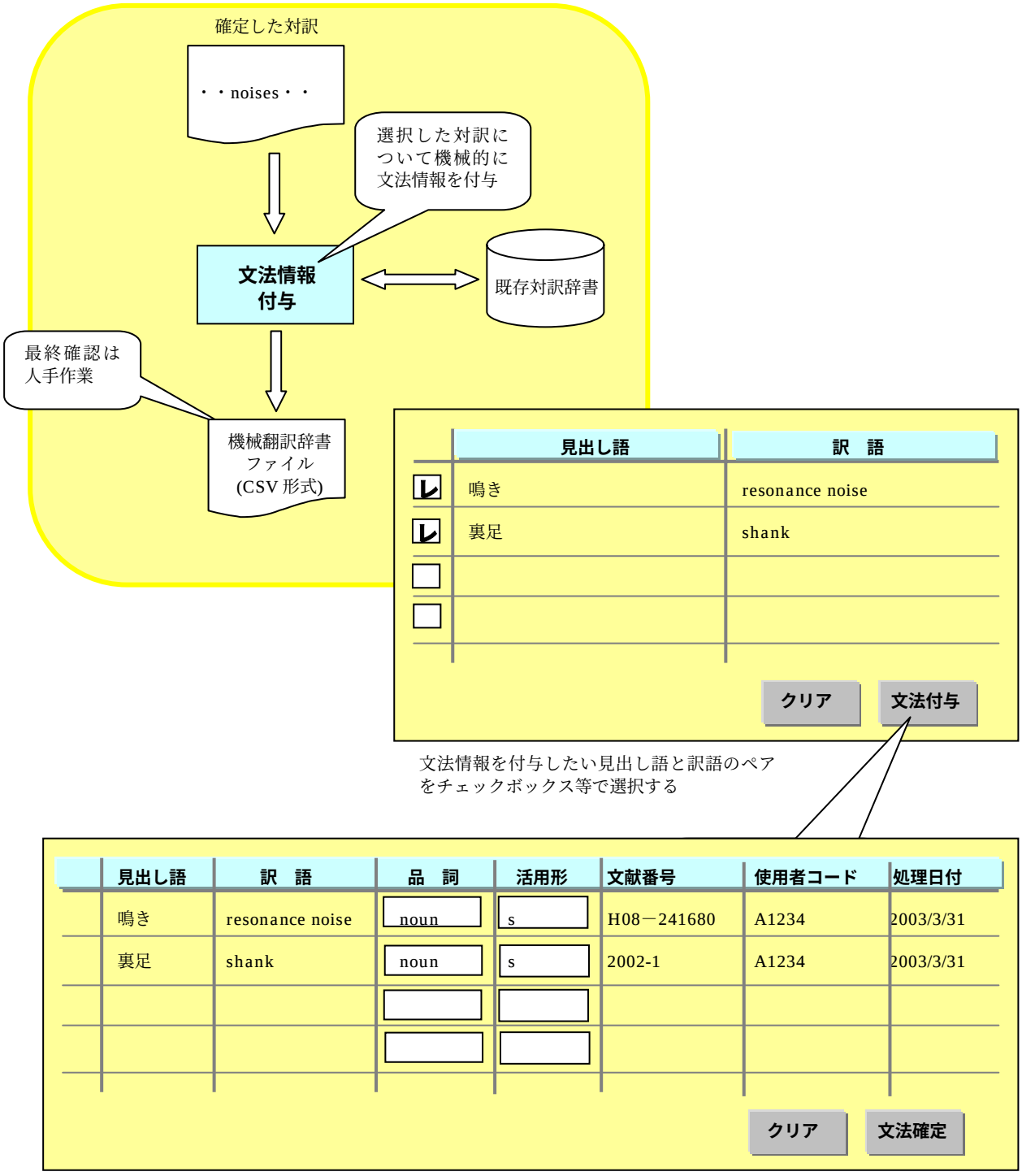
的確な訳語がない場合は機械的な推定に失敗したので、訳語編集ボタンを選択し、従前の様に人手で専門用語辞典や専門書等から対訳を入力する。



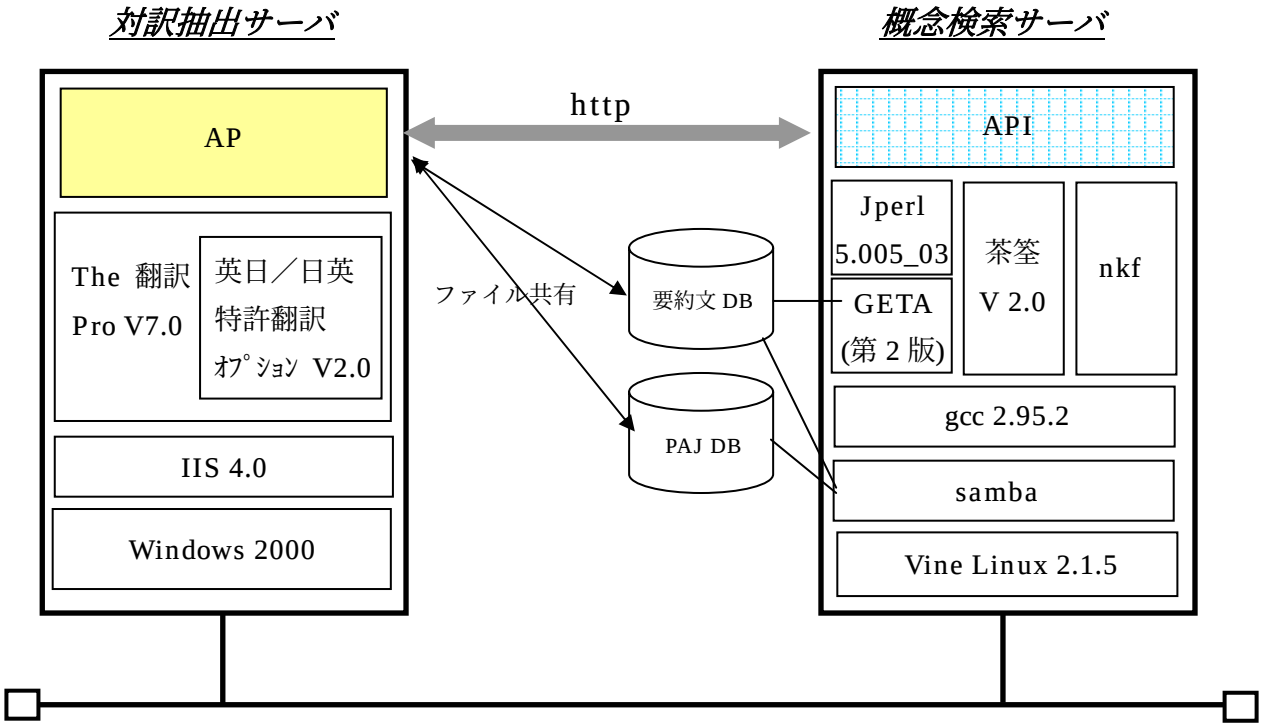
（５） 文法情報の付与フェーズ

選定した訳語に機械翻訳で使用する文法情報と特許分類情報を機械的に付与する。文法情報は東芝の機械翻訳システムを利用した簡易的推理法および既存の辞書データに基づく情報によって機械的に付与する。機械的に付与された文法情報は画面上で確認し修正できる。文法情報の確認・修正作業は翻訳者が行う。

確認が完了したものは汎用的に機械翻訳システムで利用可能なファイル形式（CSV）として出力する。



3.4. システム構成



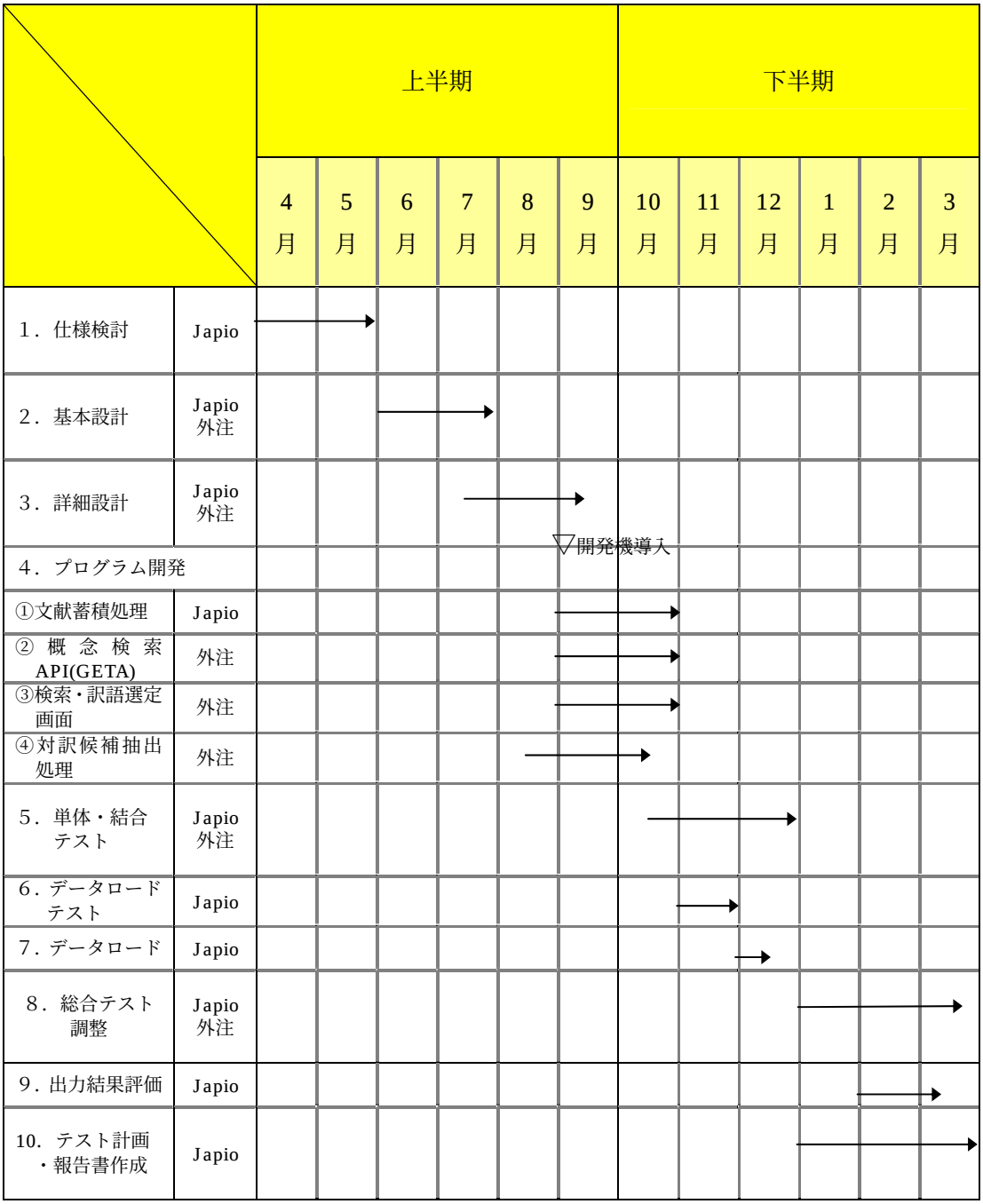
対訳抽出サーバ：ハードウェア構成

本 体　：東芝製 MAGNIA LiTE30 (W)
CPU　：PentiumⅢ(1GHz)×2
ﾒﾓﾘ　：2GB (512MB×4)
HDD　：240GB（80GB×3）
ﾓﾆﾀｰ　：12.1 型ｶﾗｰ液晶ﾓﾆﾀｰ

概念検索サーバ：ハードウェア構成

本 体　：Dell 社製 PE 500SC
CPU　：PentiumⅢ(1.13GHz)
ﾒﾓﾘ　：2GB (512MB×4)
HDD　：360GB（120GB×3）
ﾓﾆﾀｰ　：15 型ｶﾗｰ液晶ﾓﾆﾀｰ

4．平成１４年度スケジュール



補足） ２．基本設計で外注が関与したのは機械翻訳システムの改造部分。
４．①文献蓄積処理：　概念検索サーバへの文献実データ（対訳コーパス）の蓄積処理（CD-ROM 公報ライクの文献管理方式）。及び概念検索 DB（GETA）の検索データ（インデックス）蓄積処理。
４．②概念検索 API：　本システムで使用する概念検索（GETA）機能の API 作成。（Perl ベースのインタフェースを想定）
４．③検索・訳語選定画面：　概念検索画面、訳語選定画面、文法情報付与画面などの操作画面作成。（Web 画面ベースを想定）
４．④対訳候補抽出処理：　言語解析による対訳コーパスからの対訳候補抽出モジュールの作成。
６．データロードテスト：　概念検索 DB の検索データ作成テスト
７．データロード：　概念検索 DB の検索データの蓄積作業（要約文と対訳 2001 年分）

5．検証

5.1. 検証の方法

本来なら、実際の翻訳作業で効果を検証するのが理想的であるが、厳密な納期が定められた実作業の中で、不確定な要素を含む原型システムを利用して検証を行うことは困難である。また、人的資源を新たに投入することも容易ではないので、これに代る方法としてＪａｐｉｏが蓄積している用語調査票と翻訳指導票を利用して効果の検証を行なった。

Ｊａｐｉｏでは翻訳者が翻訳を行ない、校閲者が翻訳結果をチェックして品質を保っている。用語調査票は、翻訳者が翻訳時に迷った用語を質問用紙に記録し、校閲者が調査回答したデータシートである。翻訳指導票は、翻訳者の翻訳を校閲者がチェックして、不適切な翻訳を正した上で、翻訳者を指導するために作成したデータシートである。Ｊａｐｉｏではこれらを合わせて「難語シート」と呼び、広く利用している。

これらのシートには日本語見出し・対訳・文献番号が記載されているので、「日本語見出し」を「翻訳に迷った語」、「対訳」を「正解」として、本システムが想定する翻訳時に「迷う用語」が解決できることを検証する。

5.2. データ範囲

データの範囲は 2001 年に公開された公開特許公報の要約文(359,103 件)とその英訳とする。

5.3. テストデータ

テストデータは 2001 年の公開公報を英訳する際に、翻訳者・校閲者の間で交された用語調査票・翻訳指導票の内、1,245 件を利用した。（以下「テストデータ」という。）

テストデータは以下の項目を持つ。

5.3.1．翻訳に迷う用語（見出し語）

日本語の単語または複合語。翻訳者が校閲者に問い合わせた用語または校閲者が翻訳者を指導した用語。

5.3.2．対訳

見出し語に対応する英語表現。単語、複合語もしくはフレーズの場合がある。

5.3.3．出願番号

見出し語が存在した特許文献の公開番号。（以下この文献を「注目案件」という。）

5.4. 調査項目

調査した項目と調査結果を分析する上での注意点をあげる。

5.4.1. 概念検索

概念検索の状況を調査する。本調査の母集団は 2001 年の公開特許要約であり、テストデータは 2001 年の文献から発生しているため、検索が成功していれば必ず、注目案件がヒットすることになる。この点は新しく公開された文献の翻訳に「迷う用語」を検索する場合と異なるところである。

(1)ヒット件数

概念検索でヒットした文献総数。ヒット件数が２０件を越えるものは、類似度上位２０件と表示する。

(2)案件位置

注目案件の概念検索のリスト内の位置。リストは類似度順に並ぶため、案件位置が 1 番であることが望ましい。注目案件がリスト内に存在しない場合は何らかの理由で概念検索が失敗している。

5.4.2. 対訳候補の抽出

ヒット件数が１件以上の場合、対訳候補の抽出機能を調査した。ヒット件数が２０件を越える場合は類似度上位２０件で対訳候補抽出処理を行なった。

(1)対訳候補の正解位置

対訳候補の抽出に成功した場合、対訳候補が先頭から何番目に存在したか記録した。尚、対訳候補は確度順に並ぶため、1 番が最も望ましい値である。

尚、文法情報付与支援機能の調査に関しては、テストデータの情報に含まれない文法情報を必要とするため、今回の評価を見送った。

6．検証結果

概念検索と対訳候補の抽出の成功率の調査を行なった。

6.1. 概念検索結果の分析

本システム概念検索で似た用語を持つ文献を検索する成功率は、次の結果を参酌すると６５%（補正後）と８８%（補正前）の間と考察される。今回のデータ範囲は 2001 年の公開特許だけであるが、さらにデータ範囲を広げることで概念検索結果の成功率を８８%に近づけることを目指す。

6.1.1．補正前の概念検索結果（概念検索ヒット１件以上）

項番	内訳	語数	%	備考
1	調査した語数	1,245 語		
2	概念検索に成功した語数	1,092 語	88%	ヒット件数 1 を含む
3	概念検索に失敗した語数	153 語	12%	注

注：概念検索に失敗した原因は以下の様なものがある。尚、定量的分析は行っていない。
(1) 分かち書きミス。分かち書きシステム茶筌の辞書チューニングが必要。
(2) 表記の振れ。「見出し」の表記が要約文の表記と異なる。
(3) シート記入ミス。公開番号が間違って付けられている。

6.1.2．補正後の概念検索結果（概念検索ヒット２件以上）

前述の「補正前の概念検索結果」はヒット件数 1 を含んでいる。ヒット件数 1 は注目案件、すなわち翻訳に迷う用語を採取した案件自身であるため、これを除いた件数を知る必要がある。

項番	内訳	語数	%	備考
1	調査した語数	1,245 語		
2	概念検索のヒット件数１件	286 語	23%	注目案件自身だけがヒット
3	概念検索のヒット件数 2 件以上	806 語	65%	補正後の概念検索結果
4	概念検索に失敗した語数	153 語	12%	

6.2. 概念検索結果の信頼性の確認

概念検索結果の結果が概念的に近いかどうか定量的に分析した結果、概念検索のヒット件数が 2 件以上のなかで、注目案件の位置が１番目のものが９０%あるので、用語的な類似文献が上位に来ることが確かめられた。

項番	内訳	語数	%	備考
1	調査した語数	1,245 語		
2	概念検索のヒット件数１件	286 語		注目案件自身だけがヒット
3	概念検索のヒット件数 2 件以上	806 語	100%	注目案件以外もヒット
4	概念検索で２件以上ヒットの内注目案件の位置が１番目	726 語	90%	概念検索成功

リストの上位には類似文献が集まっていたがその評価には、本当に概念的に近いかどうかの判定基準を定めて、定量的に検証する必要がある。

6.3. 対訳候補の抽出結果

翻訳に迷う用語に対して対訳候補を抽出できた成功率は３０%台である。これは技術文献の翻訳に慣れた翻訳者であっても翻訳が難しかった語について正解となる用語を発見出来たことを意味する。

6.3.1. 対訳候補の抽出結果（全体）

概念検索のヒット件数が１件の場合は注目案件、すなわち翻訳に迷う用語のあった案件自身がヒットしているため、対訳候補抽出結果を分析する場合は区別して考える必要がある。下表の値は参考値である。

項番	内訳	語数	100%は分母	100%は分母	備考
1	調査した語数	1,245 語		100%	
2	概念検索のヒット件数１件以上	1,092 語	100%		
3	概念検索のヒット件数 1 件以上の場合の対訳候補の抽出成功	402 語	36% (注１)	32% (注２)	注 1：402/1092 注２：402/1245

6.3.2. 対訳候補の抽出結果（ヒット件数別）

上記の結果は注目案件自身だけがヒットした概念検索のヒット件数１件を含んでいるため、ヒット件数と抽出結果の変化を調査した。概念検索のヒット件数は対訳候補の抽出の元となる対訳文献数のことである。
注目案件自身つまり１件だけで対訳候補の抽出を行なった場合が最も成功率が低く、ヒット件数２～３件の場合、成功率がピーク３７～３９%に達する。つまり、概念検索で多くの件数がヒットした場合、類似度の上位３件を対訳抽出の対象にすることで、対訳候補発見の成功率が高くなる可能性がある。

項番	内訳	（A）概念検索該当語数	（B）対訳抽出成功	B／A %	備考
1	概念検索のヒット件数１件	286 語	89 語	31%	注目案件自身
2	概念検索のヒット件数 1 件以上	1,092 語	402 語	37%	
3	概念検索のヒット件数 2 件以上	806 語	313 語	39%	
4	概念検索のヒット件数 3 件以上	654 語	239 語	37%	
5	概念検索のヒット件数 4 件以上	562 語	200 語	36%	
6	概念検索のヒット件数 5 件以上	499 語	164 語	33%	
7	概念検索のヒット件数 10 件以上	307 語	94 語	31%	
8	概念検索のヒット件数 20 件以上	121 語	42 語	35%	２０件で打切り

6.4. 調査の課題および問題点

概念検索の精度をあげる必要があることは当然であるが、より効果が見込まれるのは対訳候補の抽出機能の向上である。

対訳候補の抽出機能の成功率は３０%台であるが、正解が対訳候補リストの上位に現れないものも多かった。これは先頭画面に入る１０番目程度までに入ることが望ましい。

フレーズなど３語以上の英語に対応する用語の対訳候補の抽出は困難である。但し、推定訳語が正解に含まれる用語を示すことが多いのでヒントとして利用できる可能性がある。

言語処理は辞書など、チューニングが必要な個所が多くある。チューニングを施しながら、さらに使いやすいシステムにして実際の翻訳支援に使うことを目指したい。

7．付録

テストデータおよびテスト結果一覧