

産業日本語における日本語の諸相

—文書の作成・利用を高度化する新たな文書工学の仕組—

一般財団法人日本特許情報機構 特許情報研究所顧問 **横井 俊夫**

PROFILE

1966年に電気試験所（現在：産業技術総合研究所）。1982年より第五世代コンピュータプロジェクトの推進に従事。1987年より電子化辞書プロジェクトの推進、運営に従事。1995年よりフィリピン国にてODAプロジェクトの推進、指導に従事。1997年より東京工科大学、2008年より現職。工学博士、東京工科大学名誉教授。

✉ yokoi@media.teu.ac.jp

☎ 03-3615-5513

産業日本語は、明晰な文書を効率良く作成し、蓄積された文書を効果的に利用できるようにする文書環境を作り出す。文書には、多様な言語が係わり、多様なメディアが係わる。文書環境の構築に求められる日本語の諸相を明らかにし、文書の作成や利用を高度に支援する仕組を概観する。新たな文書工学・ドキュメント工学の展開である。Web情報環境の格段の整備と関連技術の飛躍的な進展、文書処理技術・言語処理技術の進展、認知言語学や対照言語学などの言語学における知見の蓄積、これらに基づく新たな文書工学の展開である。

1 和文文書を作成する

ソフトウェア工学は、ソフトウェアの開発・管理のプロセスを工学化するために、そのプロセスを工学的に割り切ってモデル化した。文書工学においても、文書の作成・利用のプロセスに対する工学的なモデル化が必要である。まず、和文（日本文）文書を作成するプロセスをとりあげる。そのプロセスのモデル化が図1である。

日本文化（日本の技術文化）とA分野（ある技術分野）という文脈を背景にして、文書は作成される。日本の技術文書は、日本社会の諸制度や日本の企業慣習や日本の生活環境などという日本の技術文化の中で作成され利用される。技術文書といえども、日本的なもの影響下にある。

先立って、文書工学の基本概念である「文書（document）」と「文章（text）」を定義しておく。文書は、

主となる表現メディアに言語を用いるコンテンツ形式である。文章は、言語という表現メディアにおける情報の表現形式である。ソフトウェア工学においては、「ソフトウェア」や「プログラム」が、主となる表現メディアにプログラム言語や仕様記述言語を用いるコンテンツ形式となる。そして、「プロシージャ」とか「クラス」と呼ばれるものが、プログラム言語という表現メディアにおける計算手順の表現形式となる。

「試みる日本語」

文書作成は、「書き手（著者）」の「着想」を端緒に始まるとする。着想を生み出す段階は、一応、完了していきましょう。着想は、全体の概略や部分々々の要点を書きとめた「〇（ゼロ）次文書（着想メモ）」の形式にまとめられる。この〇次文書を書くための日本語が「試みる日本語」である。

言葉の基本機能には、思考のツールとしての機能と伝達のツールとしての機能がある。言葉を媒介にして、思考、すなわち、試行錯誤が行われる。「試みる日本語」は、思考のツールとしての機能に重きを置いた日本語である。思考のツールとしての「試みる日本語」には、思考のダイナミズムを加速する柔軟さ、時には、いい加減さが求められる。思考が集団による協調作業として行われる場合も、作業者どうしの了解が得られていると仮定できるなら、用語の用法や文の構文や文章の構成などに関する細かな規則に煩わされる必要はない。ある意味では、言語を大雑把に使いこなすセンスが必要である。

思考には、いわゆる言語だけではなく、思考に適した

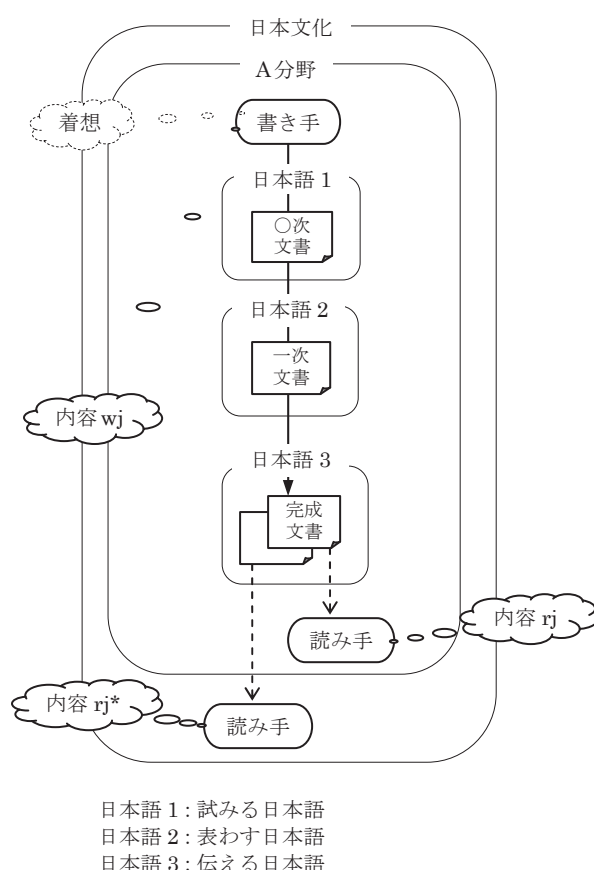


図 1. 和文文書を作成する

図表等の様々な表現メディアが用いられる。いわゆる言語には、線状構造という束縛が課せられる。この束縛から逃れるために、マインドマップのような図式言語などを有効に活用することもできる。ただし、いかなる表現メディアを用いるにしても、日本人は日本語で考えるという鉄則に則ることが重要である。

言語を思考のツールとして使うプロセスと言語を伝達のツールとして使うプロセス、この2つのプロセスを同時に行うことはできない。制限英語をはじめとする制限言語に生じた多くの不具合は、この2つのプロセスの基本的な違いを混同したことが原因である。

「表わす日本語」

次に、思考から伝達に向けたプロセスが始まる。このプロセスは、大きく2つに分けられる。表現対象に注視するプロセスと表現メディアに注視するプロセスに分

けられる。

まず、表現対象に注視し、表現対象を適切に確定し、表現対象を正確に表現する。このプロセスは、思考の精密化というプロセスを伴って進行していく。精密な思考によって文書の内容が確定され、〇次文書は、「一次文書」に仕上げられる。この一次文書を書く日本語が「表わす日本語」である。

「表わす日本語」には、表現に向けて以下のような対応が求められる。すなわち、

- ① 表現対象の論理的な構成や構造を過不足なく正確に捉え表わすこと
- ② 文書の使用目的や文書項目の記述要請を過不足なく満たすように表わすこと
- ③ 図表や引用資料など、付属情報との整合性が保てるように表わすこと



「表わす日本語」は、文書の構成形式にしたがって対象内容を表現する。したがって、「表わす日本語」は、文書の特性や対象の特性に強く依存する日本語である。表現対象の個々の特性は、共通となる特性に整理され、それに文書の特性が加味され、それぞれの文書形式に特有の文章特性が定まる。この文章特性を体現するのが「表わす日本語」である。

「伝える日本語」

表現メディアに注視し、「読み手（読者）」が効率良く間違いなく読み取れるようにすることを目的に、一次文書は、「完成文書」に仕上げられる。内容としては、一次文書と完成文書の間に大きな違いはない。原則として、同一である。この完成文書を書く日本語が「伝える日本語」である。

「伝える日本語」は、産業日本語という表現メディアに共通となる表現特性に基づく日本語である。すなわち、すべての産業技術文書に共通となる文章特性に基づく日本語である。「伝える日本語」は、個々の文書形式には依存せずに、文章一般の形式¹⁾に基づいて定義される。

「伝える日本語」は、書き手が意図する「内容 wj」を読み手に正確に効率良く伝達するための日本語である。「内容 wj」と「内容 rj」あるいは「内容 rj*」とは、等

1) 文章の定義：まとまりのある情報を表現する基本単位である。言語という表現メディアにおける表現のモジュールである。読み手は、それぞれの文章からまとまった情報を読み取る。文章 (text) は、ひとつ以上の段 (paragraph) からなり、段は、ひとつ以上の文 (sentence) からなり、文は、ひとつ以上の語 (word) からなる。モジュールとしての文章は、結束表現に関して、以下のような条件を課せられる。すなわち、①接続表現（接続詞や接続表現語句など）は、文章の先頭や末尾には用いない。②通常の参照表現（指示詞を用いた照応詞など）は、文章内の表現を参照するだけである。③参照できる文章外の表現は、文章文脈外に共通に設定されている物事や参照先を明示的に指定できる物事に限られる。ただし、すべての文章が文章・段・文・語という正規の構成を取るわけではない。文章を定義付ける基準は、モジュール性である。タイトル文は、モジュール性の高い表現単位である。したがって、タイトル文は、これだけで文章である。一文だけから構成される文章、あるいは、複数の語から構成される文章である。要約文も、高いモジュール性を持つ。したがって、要約文を表現するものもひとつの文章である。この文章は、ひとつの段、あるいは、高々数個の段から構成される。

しくなければならない。ただし、ここでの読み手が読み取る内容とは、一次情報としての内容である。この一次情報に読み手の意図が加味され、最終的な読み取り内容が定まる。読み手が書き手と同じ分野に属するか否かによって、「伝える日本語」は異なってくる。分野特有の用語や言い回しに関しては、分野ごとの配慮が必要である。ただし、閉鎖的なジャーゴン化した分野特有の日本語という現象には、改善すべき点が多々ある。「伝える日本語」は、諸分野に渡って開かれた日本語であることを心掛けるべきである。

次の2つが「伝える日本語」に求められる基本要件である。すなわち、

- (1) 表現単位に関して、構造的にも解釈的にも確実性が高く、その内部構成の容認性が高いこと
- (2) 表現単位間の関係に関して、構造的にも解釈的にも確実性が高く、その関係性の容認性が高いこと

この基本要件にしたがって、仕様策定を進める。その際の観点としては、①状況・事象・事物とその表現、②客観表現と主観表現、③表現と表現の連携構造、④オブジェクト表現とメタ表現、⑤具象事象と抽象事象、⑥難解表現と平易表現および冗長表現と簡潔表現、等々を挙げることができる。これらの観点に基づいて、「伝える日本語」の仕様は、「表わす日本語」の非明晰さを明晰化する言い換え規則と共通辞書の管理規則の形式にまとめられる。以下のような規則が含まれることになる。すなわち、

[事象表現の明晰化規則]

- 単事象表現（単文）の明晰化規則
- 複事象表現（連文、複文）の明晰化規則
- メタ表現の推奨表現規則

[事物表現の明晰化規則]

- 事物基表現の明晰化規則
- 事物基修飾の明晰化規則

[構造表現の明晰化規則]

- 符号（読点や括弧など）による構造表現規則
- 並立表現の明晰化規則

[共通辞書の管理規則]

- 未登録表現（未登録複合語、未登録複語）の登録規則

〈作成を支援する〉

着想から完成文書に至る和文書作成を支援する仕組みを概観する（図2）。着想を〇次文書にまとめるプロセスを支援する機能は、多くが発想支援として議論されている機能に含まれる。具体的には、①「試みる日本語」で発想の断片や断片間の関係付けを書いたり修正したりすることを自由闊達に行える場を提供すること、②適切な発想事例や発想パターンを探し出し発想を促進させる示唆を提供すること、等々である。

〇次文書を一次文書にまとめ上げるプロセスを支援する機能は、文書形式ごとに用意される文書作成支援機能となる。具体的には、①「表わす日本語」で書かれた部分々々の整合性を対象世界のオントロジーや文書の記述要件に基づいて検証すること、②適切な文書事例や文書パターンを探し出し作成を誘導すること、等々である。

一次文書を完成文書に仕上げるプロセスを支援する機能は、「表わす日本語」から「伝える日本語」への翻訳支援機能となる。一次文書の「内容 wj」と完成文書の「内容 wj'」には、原則として、差異はない。「表わす日本語」から「伝える日本語」への言い換え規則は、機械化できるものから人間の判断に多くを委ねるものまで多岐に渡

る。したがって、日本語から日本語へのインタラクティブ方式翻訳支援機能として実現されることになる。

2 X 文書を（翻訳によって）作成する

X 文書（X 語、例えば、英語、中国語などで書かれた文書）を作成する方法には、①書き手が X 語で書き下ろす、②和文書をその書き手が X 語に翻訳する、③和文書の X 語への翻訳を翻訳者に依頼するという 3 つの場合が考えられる。ただし、ここでの X 文書は、簡単なメールや報告書の類ではなく、対外的に然るべき責任を伴うような文書であるとする。したがって、①の場合は、書き手に母語話者と同等の X 語能力が求められる。この場合の作成プロセスは、図1の「日本」を「X」に置き換えたものとなる。しかし、このような場合は、稀であり、多くは、②か③の場合となる。ここでは、③の場合を取り上げる。

X 文書を和文書から翻訳によって作成するプロセス、すなわち、翻訳者が行う日 X 翻訳のプロセスをモデル化しなければならない。そのモデル化が図3である。通常の翻訳では、翻訳者は、「和文依頼文書」から直接「X 文完成文書」を作成する。これを 3 ステップのプロセスにモデル化する。和文依頼文書を「訳せる日本語」で書かれた「和文翻訳原稿」に翻訳するステップ、

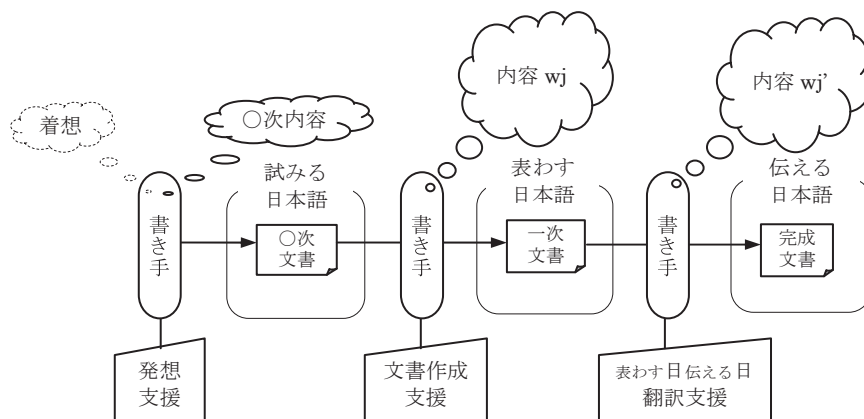
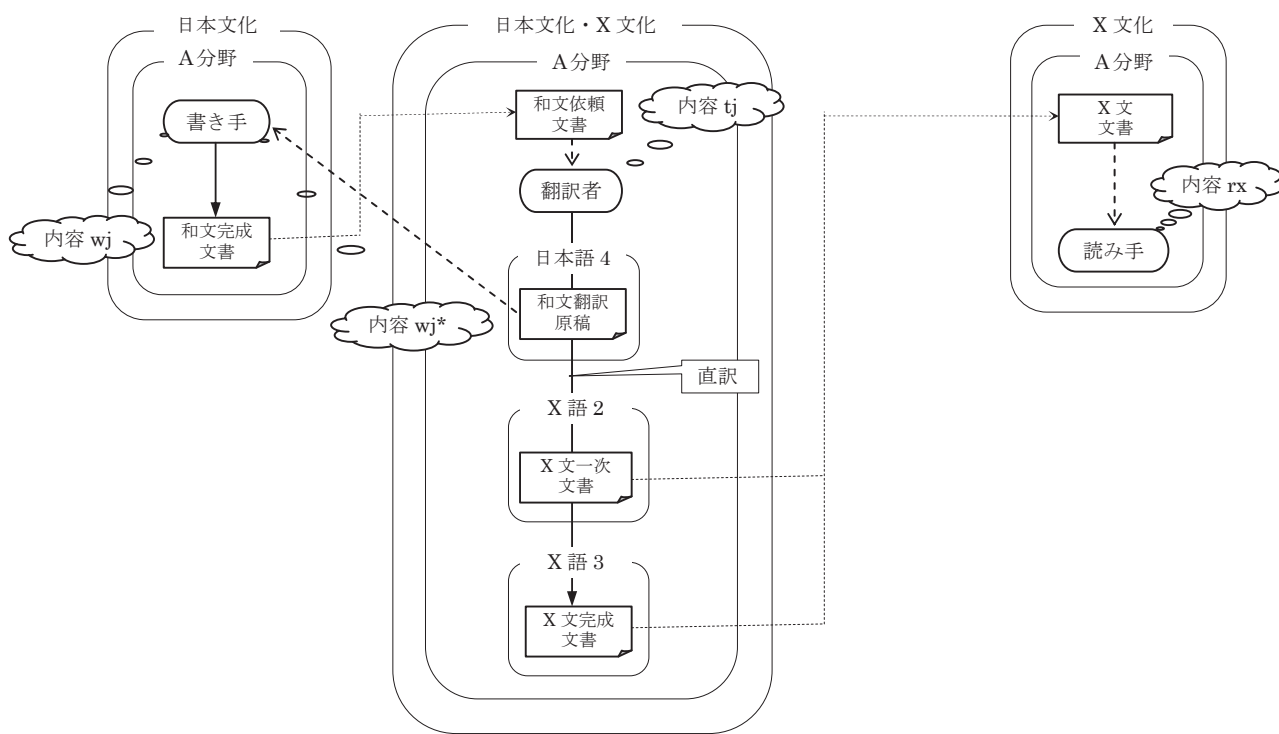


図2. 和文書作成を支援する



日本語 4：訳せる日本語（付注つき日本語）
 X語 2：表わす X語
 X語 3：伝える X語

図 3. X 文文書を（翻訳によって）作成する

和文翻訳原稿を直訳し「表わす X 語」で書かれていることになる「X 文一次文書」を作成するステップ、X 文一次文書を「伝える X 語」で書かれた「X 文完成文書」に言い換えるステップ、これらの 3 ステップである。「伝える X 語」によって、X 語を母語とする読み手は、違和感なく効果的に「内容 rx」を読み取ることができるようになる。

モデル化の要となるのは、「訳せる日本語」で書かれた「和文翻訳原稿」である。翻訳原稿というものを顕在化させる。この顕在化によって得られる利点は以下である。

- ① 書き手の意図と翻訳者の解釈を正確にすり合わせることができる。

書き手（翻訳依頼者）が翻訳原稿から読み取る内容（内容 wj*）に基づいて、書き手の意図（内容 wj）と翻訳者の解釈（内容 tj）をすり合わせることができる。

- ② 「訳せる日本語」を中継（中間）言語とし、多言語翻訳に対応することができる。

「訳せる日本語」を様々な X 語に対応できる中継言語となるように設計する。ただし、当初は、英語に対する「訳せる日本語」を中継言語の近似解とする。

- ③ インタラクティブな翻訳支援を実現できる。

機械翻訳システムを現在のブラックボックス方式から、インタラクティブ方式へと進展させることができる。和文翻訳原稿作成までのステップは、日本語能力だけで対応できるからである。

- ④ 翻訳知識を集積化し、翻訳作業を高度化・効率化していくことができる。

翻訳原稿を中心にして一連の文書が構成するコーパスは、体系的に外在化された翻訳知識である。大量に集積された翻訳知識を用いて、非常に高度な翻訳支援システムを開発することができる。

「訳せる日本語」

×語へ直訳できる日本語が「訳せる日本語」である。そこで、まず、直訳できるということを定義する。すなわち、

- (1) 文章、および、その構成成分（段、文、語等）は、線形翻訳可能である。すなわち、ある構成成分の翻訳は、それを構成する要素の翻訳結果を線形合成することによって得られる。
- (2) 非線形となる成分、すなわち、非線形複語表現や非線形複合語（線形複合語である臨時一語を除く）などは、翻訳開始の前には、すべて共通辞書に登録されている。
- (3) 曖昧性が生じる部分（多義語の訳語選択、参照・修飾先等の確定）に対しては、曖昧性解消の手順が明示されている。あるいは、確定するための情報が付注によって提示されている。

この直訳の定義に従えば、「訳せる日本語」は、詳細には訳言語である×語によって異なってくる。そこで、英語に対する「訳せる日本語」を共通基準として採用することから始める。さらに、何をもって非線形成分と認定するのか、何をもって多義を認定するのか、言語学的に厳密な議論をすると、結論の得られない事柄が多くなる。そこで、工学的な基準を設ける。すなわち、共通辞書に非線形成分として登録されているか否かが、非線形性を認定する基準であるとする。また、辞書に登録されている語義をもって、多義を認定する基準とする。そのために、共通辞書の初期状態と新規項目の辞書登録手順を定めておく。後は、然るべき辞書管理者のもとで、利用者達の協調知としての共通辞書が作られていくことになる。新規複合語を除く非線形成分は、いずれ、すべてが共通辞書によって把握されることになる。

「訳せる日本語」の仕様は、「伝える日本語」を直訳できるようにするための規則によって定義される。規則は、言い換え規則と付注規則からなる。規則を、依頼文書の原文を書き換えてしまう言い換え規則と原文をそのままにして注記とする付注規則に分ける。言い換えが、書き換えを行っても日本語としての自然さが保たれる場合

は、言い換え規則とし、不自然となる場合は、付注規則とする。このように規則を分けることによって、翻訳者が直観的に混乱なく判断でき、機械処理が余計な雑音によって誤処理することを避けることができる。日本には、漢文訓読という付注文化の伝統がある。

言い換え規則と付注規則を定める際の要点を以下に挙げておく。

[言い換え規則で対応する日本語特有構文]

- 主題 - 題述による情報構造表現
- 主観的把握による事態把握に基づく事象表現
- 主格成分や対格成分が明示化されない事象表現
- 主格成分が長く複雑となる事象表現
- 間接受身や自動詞受身の構文
- 二重ガ格構文や八 - ガ構文
- 複数動詞述語による構文
- 数量詞分離構文、等々

[付注規則で対応する表現特性]

- 事物表現（名詞）の表現特性
- 事象表現（動詞）の表現特性、等々

〈日本語から×語への翻訳を支援する〉

和文依頼文書を×文完成文書へと翻訳するプロセスを支援する仕組みを概観する（図4）。和文依頼文書から和文翻訳原稿を仕上げるプロセスを支援する機能は、「伝える日本語」から「訳せる日本語」への翻訳支援機能となる。依頼文書の「内容 tj」と翻訳原稿の「内容 tj'」とは、原則として、同じである。「伝える日本語」から「訳せる日本語」への言い換え規則は、機械化できるものから人間の判断に多くを委ねるものまで多岐に渡る。したがって、翻訳支援システムは、インタラクティブ方式のシステムとなる。支援システムが実現されるまでは、ライティングマニュアルが用意されることになる。日本語を母語とする翻訳者であれば、短期間のトレーニングでスキル獲得ができるようになるであろう。

「訳せる日本語」から「表わす×語」への翻訳は、自動化度の高い日×機械翻訳システムに委ねることができる。ただし、日本語の解析に関しては、翻訳者の介入が必要になる。「訳せる日本語」は、翻訳者が共通常識

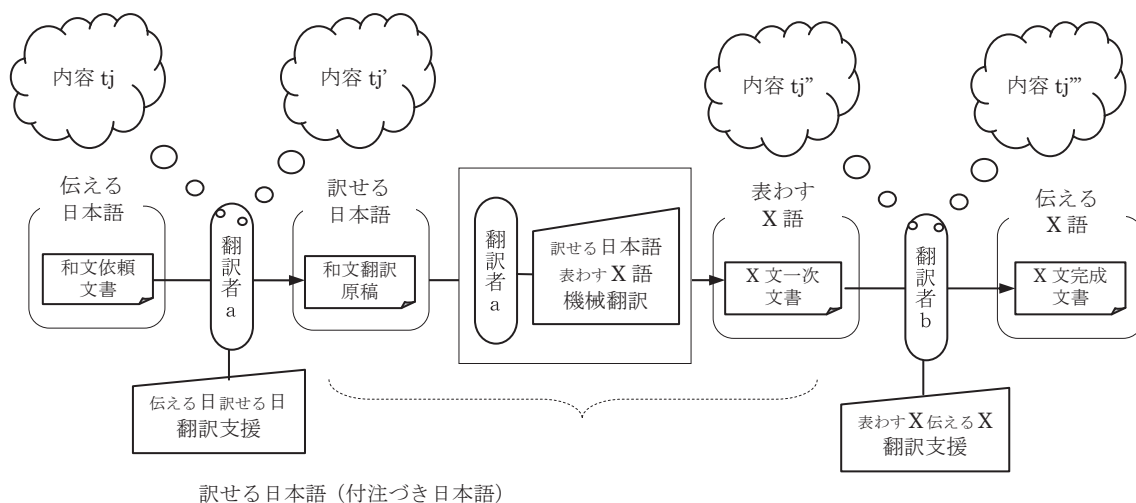


図 4. 日本語から X 語への翻訳を支援する

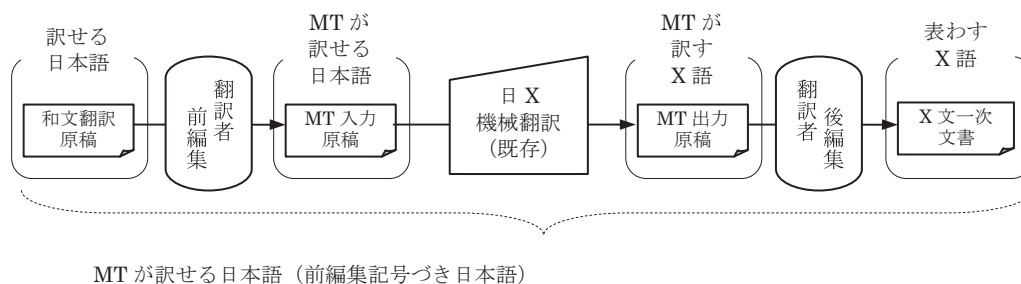
として持つ意味的・文脈的理解を前提にした日本語である。そのような意味的・文脈的理解機能を持たないシステムには、翻訳者の判断が介入することが必要となる。

X 文一次文書を X 文完成文書に仕上げるプロセスを支援する機能は、X 語を専門とする翻訳者に対する「表わす X 語」から「伝える X 語」への翻訳支援機能となる。一次文書の「内容 tj'」と完成文書の「内容 tj''」には、原則として、差異はない。この翻訳支援機能も、インタラクティブ方式の翻訳支援として実現されることになる。

〈既存の日 X 機械翻訳を活用する〉

図 4 で点線の中カッコで示された和文翻訳原稿を直訳し X 文一次文書を作成するステップには、既存の日 X 機械翻訳システムを利用する便宜的な方法を用いることもできる (図 5)。図 4 の機械翻訳に比べて、既存の機械翻訳には、いくつかの機能が不足している。これらの不足した機能を補うための工夫が必要になる。不足している機能は、次の 2 点である。

- ① ユーザーとインタラクティブに対話しながら日本語の正しい解析結果を得る機能
- ② 付注を処理する機能



MT が訳せる日本語 (前編集記号つき日本語)

図 5. 既存の日 X 機械翻訳システムを活用する

既存の機械翻訳に対しては、前編集と後編集によってこれらの不足した機能を補う。すなわち、翻訳原稿に以下の前編集を施すことによって MT 入力原稿を作成し、MT 出力原稿に以下の後編集を施すことによって X 文一次文書を仕上げる。

前編集：「訳せる日本語」を「MT が訳せる日本語」に言い換える

- ㊤ 付注を取り去り、付注無しの文章にする
- ㊦ システムの日本語解析特性を補正するための情報を前編集記号によって注記する

後編集：「MT が訳す X 語」を「表わす X 語」に言い換える

- ㊧ 前編集で取り去った付注を反映するように X 文文章を編集する

既存の機械翻訳における通常の前編集や後編集は、いったってアドホックなものである。用意されているマニュアル相当のものも、翻訳者が使うに値するようなものにはなっていない。ここでは、前編集と後編集を規則にしたがって行う手順立ったものとする。こうすることによって、既存の機械翻訳も透明性の高いシステムに成長してことになる。

3 X 文文書を(翻訳し)利用する

X 文文書の利用は、まず、X 文文書を読むことから始まる。X 文文書の読み方には、(1) 絞り込んだ X 文文書を詳細に読み解く読み方と (2) 多量の X 文文書を通読しながら必要な文書を探し出したり、必要な情報を抽出したりする読み方がある。すなわち、少量精読と多量通読である。読む方法には、①読み手が X 語を読みこなす、② X 文文書の日本語への翻訳を翻訳者に依頼する、③ X 文文書の日本語への翻訳に機械翻訳を利用するという 3 つの方法が考えられる。少量精読に対しては、①か②の方法である。X 語が英語の場合は、①の比重が高くなり、英語以外の場合には、②の比重が高くなる。多量通読に対しては、③の方法を利用できるようにしたいとい

うのが図 6 である。

日 X 翻訳・通訳であろうと X 日翻訳・通訳であろうと、翻訳者や通訳者は、日本語と X 語の間に立ち、翻訳結果に対して両者に均等の責任を負う。しかしながら、翻訳内容に然るべき責任が伴う文書翻訳の翻訳結果に関しては、翻訳する側に一方的な責任が負わされる。したがって、書き手が日 X 翻訳を行う場合は、書き手は、翻訳結果に十全の責任を負わねばならない。しかし、読み手が X 日翻訳を行う場合は、翻訳結果は、読み手の責任の範囲内でそのニーズが満たされる程度で十分である。X 日翻訳に関しては、あるレベルの翻訳精度が得られれば、機械翻訳も十分に実用的となる。

図 5 において、「日本語」と「X 語」との入れ替え、さらに、その入れ替えに付随する変更を施す。そうすると、図 5 は、「既存の X 日機械翻訳システムを活用する」図となる。ただし、「訳せる X 語」とか「MT が訳せる X 語」とかは存在しない。「X 文完成文書」が、直接、機械翻訳システムの「MT 入力原稿」となる。

X 語が英語の場合は、英語圏におけるテクニカルライティング教育の徹底と英語を第二言語とする読み手への配慮の徹底により、「伝える英語」の完成度は高い。したがって、「英文完成文書」は、「英文翻訳原稿」の要件をかなり満たしていると考えてよい。ただし、技術英文にとっては、Basic English のような語彙の平易化は望ましいものではない。また、X 語が同じ語族に属する韓国語のような場合は、「X 語完成文書」は、日本語向け「X 語翻訳原稿」としてほぼ通用すると考えてよい。

既存の X 日機械翻訳システムの実用性の判定は、「MT が訳す日本語」の程度によることになる。「MT が訳す日本語」が「表わす日本語」レベルに達するのは無理としても、内容の概略把握に支障のない程度になれば、十分に役立つことになる。

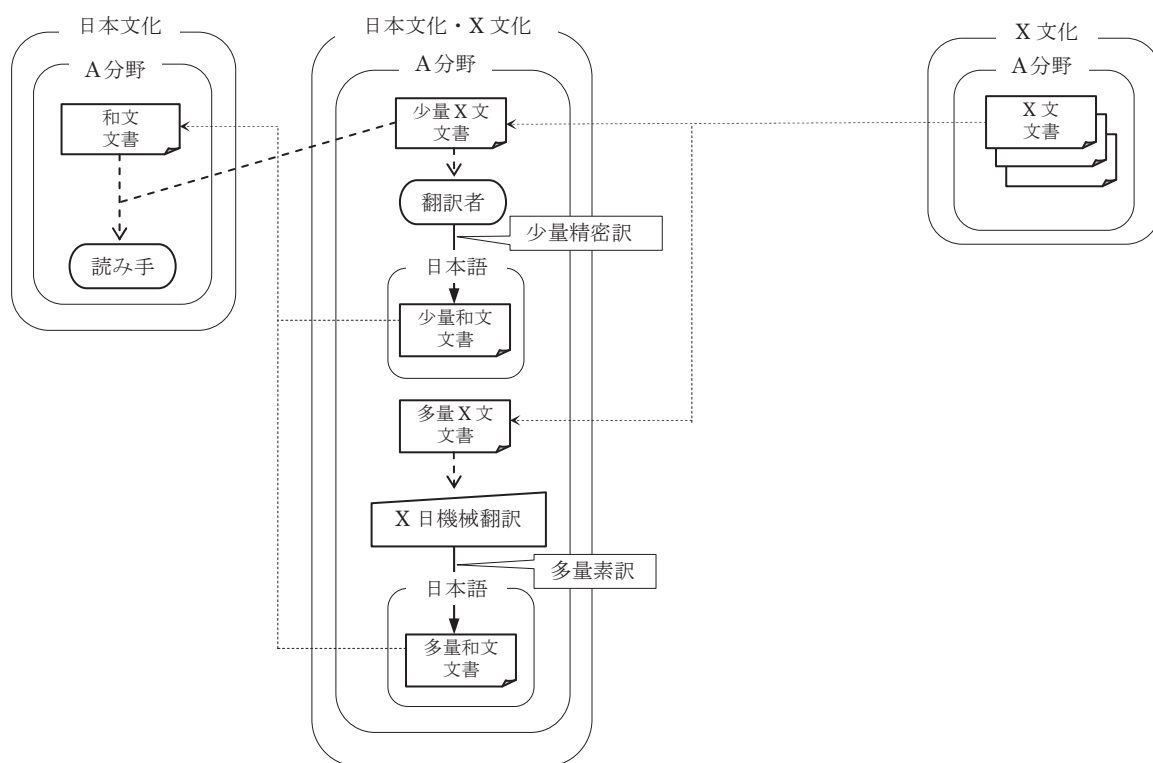


図6. X文書を（翻訳し）利用する

4 和文書を（検索し、縮約し）利用する

和文書の利用には、読むことへの配慮は不要である。基本操作として、必要な文書を探し出せる（検索できる）ことと文書を簡約し情報を抽出できる（縮約できる）ことが求められることになる。文書を検索するためには、大量の文書を文書の特徴を捉えた簡潔な「検索文書」に変換し、この検索文書を対象に検索操作を行う。この検索文書を記述する日本語が「探せる日本語」である。文書を縮約するためには、文書から要点となる部分を抽出し、抽出した情報を「要約文書」にまとめることになる。要点の抽出を容易に正確に行えるようにする日本語が「約せる日本語」である。また、要約文書を記述する日本語にも特徴を持たせることができる。

「探せる日本語」

文書内容の特徴点を最も簡潔に表現する方法は、文書の内容語の連鎖で近似する方法である。「探せる日本語」は、キーワード（内容語）が線状に並んだ超簡易日本語である。キーワードに重みをつけたり、表記のゆれを吸収したり、同義や類義の処理ができるようにしたり、翻訳によって多言語キーワードを扱えるようにしたり、等々の工夫がなされる。さらに、Web上では、検索精度を上げるために、Webページ間のリンク情報や検索履歴に関する情報など、いわば、文脈情報を利用する仕組みが活用されている。

「探せる日本語」をキーワードの連鎖から、キーフレーズの連鎖やキーセンテンスの連鎖へとレベルアップさせようという試みもある。しかしながら、膨大な文書を変換するコストとの兼ね合いという条件を満たさなくてはならない。この条件との見合いの中で、色々な工夫がな

されていくであろう。

「約せる日本語」

文書の内容を縮約し、簡約化された情報を取り出す。さらには、取り出された情報を「要約文書」にまとめる。文書の内容を特徴的に表現する要点部分を同定することが基本となる。「約せる日本語」とは、要点部分（要点となる段、要点となる文、要点となる句、要点となる語）を的確に同定できる日本語である。具体的には、以下のような要件を挙げることができる。すなわち、

- ① 文章の構成や段の構成は、要点部分の同定がし易い形式にする。
- ② 文構成は、特徴的な語や句によって、要点の度合いを明確に判定できる形式にする。
- ③ 語や句の構成は、特徴点を明解に表現する形式にする。
- ④ 論理の展開を明解に把握できるように、適切なメタ表現を用いる。