

# 第4回産業日本語研究会・シンポジウム 予稿集

平成25年3月1日

於 東京大学 情報学環・福武ホール ラーニングシアター

一般財団法人

高度言語情報融合フォーラム 言語処理学会 日本特許情報機構



## 第4回産業日本語研究会・シンポジウム開催のご案内

平成25年3月

### 産業日本語活動の新しい展開に向けて

近年、経済や企業活動の急速なグローバル化を背景に、日本を代表するグローバル企業のみならず、中小企業も含めて、国内外を問わない事業展開が不可欠となっています。

また、産業活動の現場に目を向けると、異なる分野の多くの専門家が共同作業する大規模プロジェクト、そして、組織を跨ぐ統合・共通化プロジェクトが利便性の向上や業務効率化を目的として、あらゆる場面で見受けられるようになっていきます。

このような時代背景に基づき、テクニカルライティング、機械翻訳などによる、正確かつ円滑な情報発信力と知的生産性の向上が、これまで以上に強く求められています。

産業日本語研究会は、情報発信力の強化や知的生産性の向上を目的とし、各種活動を行っています。「産業日本語」とは、情報を正確に伝達できるような情報発信力の強化、そして、コンピュータ処理されやすいような知的生産性の向上に資する産業や科学技術の記述に用いられるべき、日本語の枠組みです。

「産業日本語研究会・シンポジウム」は、これまで3回にわたり開催し、産業分野・科学技術分野における情報発信力や知的生産性の向上に貢献するとともに、わが国産業界全体の国際競争力の強化に資するような日本語の枠組みのあり方について総合的な議論を行いました。

4回目となる本シンポジウムでは、産業日本語活動の報告を行うとともに、文書を客観的に分かり易くするための他分野での取り組み、例えば、テクニカルコミュニケーションやシステム・ソフトウェアの開発文書ライティング等の各種取り組みのご紹介も交えながら、執筆・利用・翻訳など、それぞれの立場で日本語文書に関わりを持つ方々が一堂に会し、前述の時代背景をもとに、産業日本語活動の新しい展開のきっかけとなる場にしたいと考えております。また、前回に引き続き、関連技術デモ展示に触れていただき、産業日本語を実感していただくことも予定しております。

本シンポジウムは、以下の日時、内容にて行ないます。参加費は無料です。ぜひ多数の方にご参加いただきたく、ご案内申し上げます。よろしくお願いいたします。

産業日本語研究会世話人会

顧問：長尾 眞（京都大学名誉教授）

代表：井佐原 均（豊橋技術科学大学）

辻井 潤一（マイクロソフトリサーチアジア研究所）

橋田 浩一（産業技術総合研究所）

隅田 英一郎（情報通信研究機構）

横井 俊夫（日本特許情報機構 特許情報研究所）

潮田 明（富士通研究所）

松田 成正（日本特許情報機構）

## 第4回産業日本語研究会・シンポジウム

- 主催：高度言語情報融合フォーラム (ALAGIN)、言語処理学会、日本特許情報機構 (Japio)
- 後援：総務省、経済産業省、特許庁、国立国語研究所、情報通信研究機構、  
工業所有権情報・研修館、情報処理学会、人工知能学会、アジア太平洋機械翻訳協会
- 日時：2013年3月1日(金) 13:00-17:30
- 場所：東京大学 情報学環・福武ホール ラーニングシアター  
(東京大学 本郷キャンパス)  
<http://fukutake.iii.u-tokyo.ac.jp/access/index.html>
- 目的：～産業日本語活動の新しい展開に向けて～
- 参加費：無料 (事前登録制)

### ■プログラム：

- (1) 開会挨拶 13:00-13:10  
長尾 眞  
京都大学名誉教授
- 第一部 13:10-16:15- (2) 招待講演 米国におけるテクニカル・ライティングの設計と実施における諸問題 13:10-13:40  
三田村 照子  
カーネギーメロン大学教授
- (3) 講演1 特許版・産業日本語の取り組み紹介 13:40-14:40  
ー特許ライティングマニュアルー  
松田 成正  
一般財団法人日本特許情報機構 調査研究部長  
／特許版・産業日本語委員会委員  
  
ー図式クレームに基づく請求項文ライティング (構造化レベル) ー  
横井 俊夫  
一般財団法人日本特許情報機構 特許情報研究所顧問  
／特許版・産業日本語委員会委員  
  
ー図式クレームに基づく請求項文ライティング (オントロジー化レベル) ー  
橋田 浩一  
独立行政法人産業技術総合研究所知能システム研究部門上席研究員  
／特許版・産業日本語委員会委員  
  
ー総合的な特許ライティング支援環境ー  
谷川 英和  
IRD国際特許事務所 所長 弁理士  
／特許版・産業日本語委員会委員

(休憩 15分)

- (4) 講演 2 産業文書に関する言語処理技術の開発と期待 14:55-15:15  
井佐原 均  
国立大学法人豊橋技術科学大学 情報メディア基盤センター教授
- (5) 講演 3 『日本語スタイルガイド』の活用と TC 技術検定 15:15-15:35  
高橋 尚子  
一般財団法人テクニカルコミュニケーター協会専務理事  
／國學院大學経済学部教授
- (6) 講演 4 システム開発文書の品質とはなにか 15:35-15:55  
藤田 悠  
システム開発文書品質研究会 (ASDoQ) 事務局長  
／長野工業高等専門学校助教

(7) 全体討論 (質疑応答) 15:55-16:15

第二部 16:15-17:30

(8) 産業日本語関連デモ展示の紹介 (各 5 ~ 10 分) 16:15-17:00

- 「『特許明細書半自動生成システム』 Patent Generator」  
有限会社アイ・アール・ディー
- 「可読性診断技術 —産業日本語ライティング支援への活用—」  
東芝ソリューション株式会社 IT 技術研究所
- 「Patent Structure Viewer」  
株式会社インテック
- 「特許意味検索のプロトタイプシステム」  
株式会社富士通研究開発研究所 メディア処理システム研究所
- 「データベースアプローチに基づく特許機械翻訳システム」  
有限会社サイバープロ
- 「多角的な文書比較『歌詠』とカスタマイズ容易な日本語精査『鶯』」  
株式会社クレストック／アイビー・システム株式会社
- 「翻訳支援環境の研究開発」  
独立行政法人情報通信研究機構  
ユニバーサルコミュニケーション研究所 多言語翻訳研究室

(9) 産業日本語関連デモ展示セッション (於 デモ会場) 17:00-17:30

第三部 懇親会・意見交換セッション (於懇親会&デモ会場) 17:30-19:00

※第二部産業日本語関連デモ展示セッションは、福武ホールスタジオ 2, 3 において、出展各社の説明員によるデモンストレーションを行います。

※第三部懇親会は、福武ホールスタジオ 1 (スタジオ 2, 3 と一体) にて、懇親会&意見交換セッションを行ないます。ささやかですが、飲み物など準備しておりますので、こちらも是非ご参加の上、参加の方々でご交流ください。

# プログラム詳細

(1) 開会挨拶 13:00-13:10

長尾 眞  
京都大学名誉教授

第一部 13:10-16:15

(2) 招待講演 米国におけるテクニカル・ライティングの設計と実施における諸問題

三田村 照子 13:10-13:40  
カーネギーメロン大学教授

米国においてテクニカル・ライティングのマニュアルを設計をするにあたり、簡素化された英語、または「制限言語」というのはどういうものかを紹介し、その実施における諸問題について、言語学からの観点とテクニカル・ライターから見た問題点について考察する。また「制限言語」の使用にあたり、機械翻訳、英語教育などの目的や支援システムの設計について、経験に基づいた問題点や導入しやすい条件などについても言及する。

(3) 講演 1 特許版・産業日本語の取り組み紹介 13:40-14:40

ー特許ライティングマニュアルー

松田 成正  
一般財団法人日本特許情報機構 調査研究部長  
／特許版・産業日本語委員会委員

人に理解しやすく、機械翻訳など言語処理技術を活用するコンピュータにも処理しやすい日本語研究のうち、特に特許情報へ応用することを念頭にしたもの「特許版・産業日本語」委員会で研究・分析してきた。その研究・分析の1つの成果である、特許明細書作成実務に密着した「特許ライティングマニュアル」について紹介する。

ー図式クレームに基づく請求項文ライティング（構造化レベル）ー

横井 俊夫  
一般財団法人日本特許情報機構 特許情報研究所顧問  
／特許版・産業日本語委員会委員

図式クレームというグラフィカルな表現ツールを用いた請求項文（クレーム文）のライティング方式を紹介する。請求項（特許請求の範囲）は、特許文書の中核に位置付けられ、特許文章の多くの課題が請求項文に由来する。請求項文を精度高く、効率良く作成・翻訳・検索できるようにすることは、特許文章全体の作成・利用工程を大きく改善することに資する。請求項文ライティングのために、請求項文表現の高水準表現メディアとして、図式クレーム（新たなパテントマップの一種）を提案する。構造化レベルの図式クレームは、構造化言語（図的表現要素を加味された自然言語）を請求項文表現に適用したものである。構造化言語は、論理的な情報を的確に表現するためにデザインされた（自然）言語であり、産業文書・技術文書に広く活用されることを目指す。図式クレームは、請求項文を含む特許文章に対する効果的なライティングマニュアル作成やライティング支援システム開発に向け大きな役割を担うことになる。

－図式クレームに基づく請求項文ライティング（オントロジー化レベル）－

橋田 浩一

独立行政法人産業技術総合研究所知能システム研究部門上席研究員  
／特許版・産業日本語委員会委員

図式クレームに形式的な基盤を与えることによって自動処理の範囲を明確化するため、オントロジーに基づいて図式クレームを定式化する方法と、ユーザインタフェースに関連する課題について解説する。また、そのような定式化に基づく自動処理の可能性について述べ、それに必要な研究開発について論ずる。

－総合的な特許ライティング支援環境－

谷川 英和

IRD国際特許事務所 所長 弁理士  
／特許版・産業日本語委員会委員

特許ライティングを総合的に支援する特許ライティング支援環境、およびその評価について紹介する。特許ライティング支援環境は、以下の3つのシステムから構成される。

1) 特許明細書を半自動生成する「Patent Generator」

本システムにおいて、明細書から再利用され得る文章を自動抽出し、特許部品データベースを構築する。そして、本システムは、特許部品データベースから必要な文章を自動取得することにより、明細書の約40%を自動生成できる。

2) 特許請求の範囲の読解を支援する「Patent Structure Viewer」

本システムは、特許請求の範囲を解析し、請求項の構造を図的に示す特許請求項の細粒度解析機能、請求項の特性も示す注釈付きクレームツリー出力機能等を有し、特許請求の範囲の可読性を大幅に向上する。

3) 特許明細書の可読性を診断する「Knowledge Meister 文書診断」

本システムは、特許請求の範囲、明細書の表現の可読性を診断するシステムであり、人が理解しやすい文章、および高品質な機械翻訳を行うための文章の作成に寄与する。

(休憩 15分)

(4) 講演2 産業文書に関する言語処理技術の開発と期待

14:55-15:15

井佐原 均

国立大学法人豊橋技術科学大学 情報メディア基盤センター教授

自然言語処理の研究と、企業での文書処理のニーズとは十分に情報交流が行われているとは言いがたい。このような観点から、産業文書に対する自然言語処理技術の可能性を実例に沿って概説する。具体的には前編集・対訳辞書・後編集など、翻訳プロセスの中での支援を中心に、産業日本語研究会や関連の活動を通しての企業との関わりの中で始めた企業の実文書の処理に対する支援技術について述べる。

(5) 講演3 『日本語スタイルガイド』の活用とTC技術検定 15:15-15:35

高橋 尚子

一般財団法人テクニカルコミュニケーター協会専務理事

／國學院大學経済学部教授

テクニカルコミュニケーター協会では、2009年に「日本語スタイルガイド」という書籍を発行した。本書は、読者対象を大学生から若手の社会人に想定し、文章作成時に机に置いて、表記や用字・用語などのガイドラインとして活用できるよう企画した。また、「テクニカルコミュニケーション(TC)技術検定試験3級テクニカルライティング」のガイドブックとしても位置付けている。本書の中心部分の紹介、活用方法、TC技術検定の概要、テクニカルコミュニケーションの重要性などを解説する。

(6) 講演4 システム開発文書の品質とはなにか 15:35-15:55

藤田 悠

システム開発文書品質研究会(ASDoQ)事務局長

／長野工業高等専門学校助教

情報システムや組込みシステムを開発する過程で、システムの要求者や開発者は要求仕様書や設計書などの開発文書を作成します。その文書を介して開発を進めるので、限られた開発期間と費用に応じた品質を確保するために十分な開発文書である必要があります。それでは、開発文書の品質はシステム開発にどのような影響をあたえるのでしょうか。どのような文書がシステム開発に十分であるといえるのでしょうか。どうすれば、開発にふさわしい文書を作成することができるのでしょうか。本発表では、これらの疑問の解決に向けて、ASDoQが取り組んでいる活動を紹介します。

(7) 全体討論(質疑応答) 15:55-16:15

第二部 16:15-17:30

(8) 産業日本語関連デモ展示の紹介(各5~10分) 16:15-17:00

『特許明細書半自動生成システム』PatentGenerator

有限会社アイ・アール・ディー

請求項から特許明細書の約40%を半自動生成する特許明細書半自動生成システム Patent Generator をデモ展示します。

「可読性診断技術 —産業日本語ライティング支援への活用—」

東芝ソリューション株式会社 IT技術研究所

理解しにくい文の特徴を考慮した診断規則に基づいて、文の可読性を診断します。診断結果に基づいて文を修正することにより、読みやすい文章を作成することができます。また、機械翻訳の前処理に利用すると、誤訳しやすい表現を指摘するので、簡単に修正して翻訳精度を向上することができます。

### 「Patent Structure Viewer」

#### 株式会社インテック

特許書類における「特許請求の範囲」の部分の解析を行い、その読解を支援します。

- 複数の特許請求項の引用関係を視覚的に分かりやすく表示します。
- 特許請求項の構造を視覚的に表示します。

特許請求項における特徴的単語を強調表示します。

### 「特許意味検索のプロトタイプシステム」

#### 株式会社富士通研究△研究所 メディア処理システム研究所

文の意味構造を活用した、自然文入力による特許ランキング検索のプロトタイプシステムを開発しました。構文解析および意味解析を経て得られた、語と語の間の意味的繋がりを検索の類似度評価に取り入れることにより、従来のキーワードベースの検索システムに比べ、よりユーザの意図に沿った表現を多く含む特許を検索することが可能になります。

### 「データベースアプローチに基づく特許機械翻訳システム」

#### 有限会社サイバープロ

従来の機械翻訳アプローチである、ルールベース、統計ベース、事例ベースに対し、データベースアプローチは、完全対訳節と完全名詞句からなるデータベースの構築と、その応用としての機械翻訳という2段階アプローチをとっている点であり、節の構造を分解しないで、対訳節を対応させている点で特徴がある。これにより、従来自然でわかりやすい翻訳のためには、対訳文の構造を代えなければいけない文の翻訳品質を格段に向上させることができる。また、翻訳に対し、なぜこのような訳になっているかを示すアカウントビリティ機能がついている点で特徴がある。

### 「多角的な文書比較『歌詠』とカスタマイズ容易な日本語精査『鶯』」

#### 株式会社クレストック／アイビー・システム株式会社

歌詠は、任意の文書を多角的に比較し、差分を対比表形式で表示するツールです。段落や文ごとの類似度を指定して目的に合った比較ができます。鶯は、オープンソース形態素解析エンジン MeCab の結果から、文章が正しいか、業界ルールに従っているか判断してコメントするツールです。精査ルールの一部は、カスタマイズが可能です。

### 「翻訳支援環境の研究開発」

#### 独立行政法人情報通信研究機構

#### ユニバーサルコミュニケーション研究所 多言語翻訳研究室

情報通信研究機構（NICT）は、我が国の経済の成長と発展、豊かで安心・安全な社会の実現の原動力である情報通信技術（ICT）分野の研究開発と事業振興業務を進めております。多言語翻訳研究室では、産業や文化の発展に資する多言語翻訳技術の研究開発を進めています。本展示では、翻訳支援環境についてご紹介します。

(9) 産業日本語関連デモ展示セッション（於 デモ会場） 17:00-17:30

第三部 懇親会・意見交換セッション（於懇親会&デモ会場） 17:30-19:00



# 米国におけるテクニカル・ライティング の設計と実施における諸問題

三田村照子

カーネギーメロン大学教授

Language Technologies Institute

Carnegie Mellon University



Carnegie Mellon  
School of Computer Science

第4回産業日本語研究会・シンポジウム

1

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

## Outline

- Introduction
- History of CL & Applications
- Designing a Controlled Vocabulary and Grammar
- Deployment Issues for CL
- Evaluating the Use of Controlled Language



Carnegie Mellon  
School of Computer Science

第4回産業日本語研究会・シンポジウム

2

Copyright © 2013, Carnegie Mellon. All Rights Reserved.

## What is Controlled (or Simplified) Language (CL)?

- A form of language usage restricted by grammar and vocabulary rules
- No single “controlled language” for English
- Controlled language can be used:
  - solely as a guideline for authoring
  - with a checking tool to verify conformance
  - in conjunction with machine translation



## Types of Controlled Language

- **Human-oriented CL:** to improve text comprehension by humans (for technical writers and translators)
- **Machine-oriented CL:** to improve “text comprehension” by computers (for CL checkers or MT systems)



## Examples of Writing Rules

- *Do not use sentences with more than 20 words*
- *Do not use passive voice*
- *Do not make noun clusters of more than 4 nouns*
- *Write only one instruction per sentence*
- *Make your instructions as specific as possible*
- *Use a bulleted layout for long lists*
- *Present new and complex information slowly and carefully*

Q: Which rules can be checked automatically?



## History of CL & Applications

### Roots of CL

- C.K. Ogden's "Basic English" (1930's)
  - 850 basic words
  - an "international language", foundation for learning standard English
  - never widely used



## Roots of CL [2]

- Caterpillar Fundamental English (CFE) - 1970's
  - Non-technical vocabulary and grammar
  - First version had only 850 terms
  - For non-native English speakers
  - Abandoned after ~10 years:
    - insufficient for complex writing
    - CFE difficult to train and enforce



## Examples

**Non CFE:** “*Enlarge* the hole.”

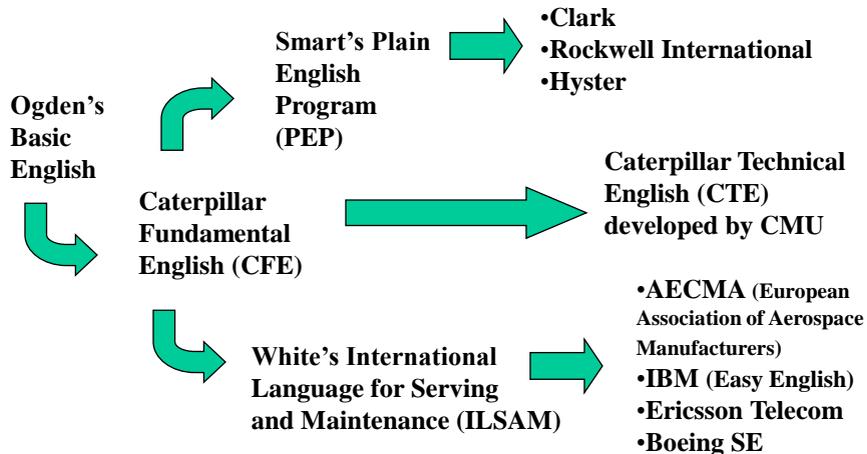
**CFE:** “Use a drill to make the hole larger.”

**Non CFE:** “The brake components must be *matched* during installation.”

**CFE:** “The brake parts with same numbers on the lower ends of the brake shoes must be installed together.”



## History of CLs



## CL Developments

- CL for Technical Documentation
  - Caterpillar Technical English (CTE) by KANTOO (CMU)
  - AECMA's Simplified English (SE)
  - Boeing Simplified English Checker (BSEC)
  - GM's Controlled Automotive Service Language (CASL)
  - Easy English (IBM)
- Simple English Wikipedia
  - [http://simple.wikipedia.org/wiki/Main\\_Page](http://simple.wikipedia.org/wiki/Main_Page)
  - Written in Basic English (Ogden)
  - For learners of English



## CL Checking

- Aids an author in determining whether a text conforms to a particular CL
  - Verify all words & phrases are approved
  - Verify all writing rules are obeyed
  - May offer help to the author when words or sentences not in the CL are found



## CL for Machine Translation

- Technical Translation
  - Large segment of translation market
  - Documentation for complex products (e.g., consumer electronics, computer hardware, heavy machinery, automobiles, etc.)
  - Involves large, specialized vocabulary
  - Writing style may be complicated
- Controlled language reduces ambiguity and complexity while increasing source text quality



# Designing a Controlled Vocabulary and Grammar for Machine Translation (CTE development by CMU)



## Designing Controlled Vocabulary

- Restrict vocabulary size and meaning
- Most useful way to limit ambiguity of input sentences
- Key to improve the accuracy of translation



## Encoding the Meanings of Vocabulary Items

- Limit Meaning per Word/Part of Speech Pair
  - Helps to reduce the amount of ambiguity
- Encode Meanings Using Synonyms
  - Finding separate, synonymous terms
  - Encode them in the lexicon
  - Synonymous terms are marked in the lexicon
  - Used in support of on-line vocabulary checking



## Encode Truly Ambiguous Terms

- When a term must carry more than one meaning in the domain
- Encode in separate lexical entries
- Resulting output structure will be ambiguous
- Lexical disambiguation by machine or by author



## Designing a Controlled Grammar

- What is CL used for?
  - Authoring without CL checker?
  - Authoring with CL checker?
  - Translating with MT?
  - Translating without MT?
- What types of constraints are needed?
- Design focus: to reduce ambiguity



## Problematic Structures

(from CTE Specification by CMU)

- Use of participial forms  
(such as *-ing* and *-ed*)
  - Used in a subordinate clause without a subject  
“When starting the engine...”
  - Reduced relative clauses  
“the pumps mounted to the pump drive”



## Problematic Structures [2]

- Verb Particles “turn on” → “start”
- Coordination of Verb Phrases  
“extend and retract the cylinders”
- Conjoined Prepositional Phrases  
“pieces of glass and metal”
- Quantifiers and Partitives  
“repeat these steps until none are left”



## Problematic Structures [3]

- Coordinate Conjunction of S  
(conjuncts must be the same type)
- Adjoined Elliptical Modifiers  
“if necessary”, “if possible”, “as shown”, etc.
- Punctuation - rules for consistency
  - use of comma, colon, semi-colon
  - quotation marks
  - parentheses



## Problematic Structures [4]

- Relative Clauses - should be introduced by relative pronouns
- Subject gap relative clause  
“The service man can determine the parts which are at fault”
- Object gap relative clause  
“The parts which the service man orders”



## Deployment Issues for CL

- CL cannot be too strict
- Author usability and productivity are important for deployment
- Expressiveness -- Balance vocabulary size vs. complex grammatical expressions
- Productivity of authoring vs. Post-editing



## Deployment Issues for CL (2)

- Controlled Target Language Definition for MT
  - Translated documents at the same stylistic quality level as the source documents
  - Set appropriate expectations about translation quality
  - Controlled language specification for TL
  - Produces more useful aligned corpora for MT/TM



## Deployment Issues for CL (3)

- Controlled Language Maintenance
  - Need to update the terminology and grammar
  - Requires a well-defined process that includes the customer / user:
    - Problem reporting
    - Initial screening of the problems
    - Process monitoring and quality control
    - Support rapid terminology and grammar updates for source and target languages



## Success Criteria for CL Deployment

- Highly-Trained Authors
- Use of Controlled Language Checker
- Technical Domain
- Translation for Dissemination



## Evaluating the Use of Controlled Language



## Benefits of CL

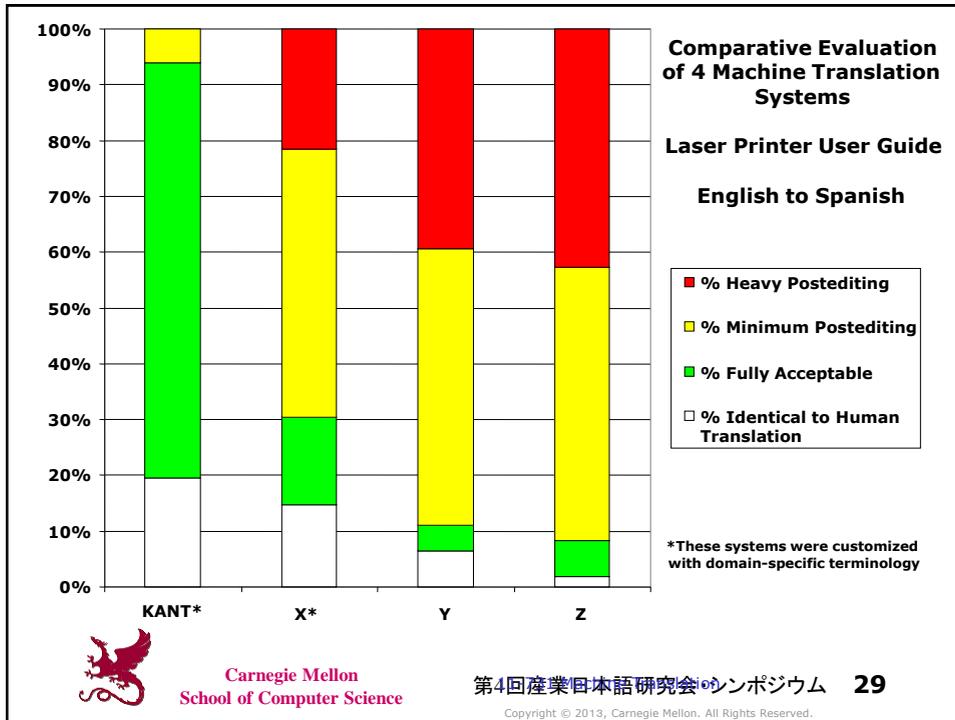
- Improved consistency of writing
- Increased re-use of documents
- Improved authoring quality
  - value of writing guidelines, term management
  - value of standardized authoring
  - improved quality / consistency of training



## Benefits of CL

- Useful for reducing ambiguity
- Ambiguity Test:
  - Average # of syntactic analyses per sentence dropped from 27.0 to 1.04
  - 95.6% have a single meaning representation
  - Lexical constraints achieve the largest reduction in ambiguity
- Improve the quality of translation output





## CL Challenges

- Writing may become more time-consuming
- An additional verification step is required
- Developing a CL may be costly
- For writers and translators, style is more satisfying than productivity, consistency, simplicity, ...
- For end users, simplicity and clarity are a top priority
- CL use must be evaluated carefully

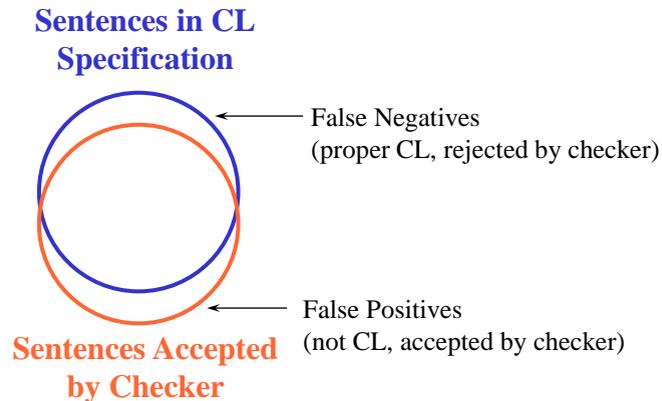


## CL in the Real World

- Software performance  
(shouldn't impact on author productivity)
- Author commitment  
(writing well vs. “getting it to pass”)
- Organizational commitment  
(publishing deadlines vs. CL compliance)



## Specification vs. Coverage



## CL is Justified When ...

- Benefits a large document volume
- Documents are hierarchical, reusable
- Checking well-integrated with document production system
- Controlled source reduces cost of translation to multiple target languages



## Questions?



## 特許版・産業日本語の取り組み紹介

### ー 特許ライティングマニュアル ー

平成25年3月1日  
(財)日本特許情報機構 調査研究部長  
松田 成正 (Shigemasa MATSUDA)



## 「特許版・産業日本語」とは

### 1. 「産業日本語」…造語

定義(コンセプト)は…「人、コンピュータ両方に明晰な日本語」

産業・技術情報を人に理解しやすく、かつ、  
コンピュータにも処理しやすく表現するための日本語



#### 産業日本語プロジェクトの目的

- (1) 翻訳における品質と効率の向上
- (2) 正確かつ円滑な情報発信力の強化
- (3) 知的生産性の向上

### 2. 「特許版・産業日本語」

「産業日本語」のうち、特許明細書への応用に関するもの

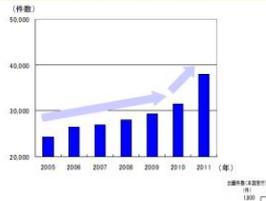
## 「特許版・産業日本語」の背景 ～例えば、海外への出願等～

### 我が国企業の海外への出願推移



- 企業の海外展開に伴い、我が国企業は、特許・意匠・商標ともに、海外への出願を増加させている。
- 日本国特許庁においても、我が国企業による海外出願増加を踏まえた対応が必要。

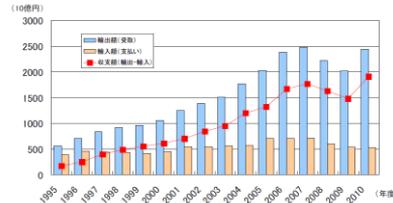
### 日本からの国際特許 (PCT) 出願件数の推移



### 我が国の技術貿易収支の推移



- 我が国企業の技術的優位性等を背景として、我が国の技術貿易収支黒字は拡大。年間2兆円規模に。
- 我が国経済・産業が海外からの収益を維持・拡大していく観点から、グローバル市場においてさらに知財権を活用していくことが重要ではないか。



産業構造審議会 第18回知的財産政策部会 (平成24年6月25日) 資料より

3

## 「特許版・産業日本語」の検討体制

### 検討体制

- 知財 (特許) の専門家 (企業の知財部、弁理士)
- 機械翻訳・特許翻訳の専門家
- 自然言語処理・情報工学の専門家
- 特許情報の専門家



委員会を組織し、専門家が知恵を出し合って検討 (平成19年～)

4

## 「特許版・産業日本語」の取り組み

平成24年度

①特許ライティングマニュアル作成

「-特許ライティングマニュアル-」

②特許オントロジー、図式表現の検討成果を整理

「-図式クレームに基づく請求項文ライティング(構造化レベル)-」

「-図式クレームに基づく請求項文ライティング(オントロジー化レベル)-」

③特許ライティング支援システムを活用した支援機能をまとめる

「-総合的な特許ライティング支援環境-」

5

特許版・産業日本語

## 「特許ライティングマニュアル」とは

(役割・目的)

- 人による特許明細書作成実務を**ガイド**
- コンピュータによる特許ライティング支援機能を実現する**基礎データ**

「文書(document)」  
特有の構造・構成

構造の改善の観点  
明細書の項目(詳細な説明など)  
に応じた記載内容、推奨表現

「文章(text)」

言語(日本語、英語、中国語など)が  
まとまりのある情報を表現する単位

明晰性の改善の観点  
例、長文の分割、係り受けの改善

用語、用字の統一  
-表記統一の観点-

↓  
「言い換え規則」

6

「特許ライティングマニュアル」作成

## 特許「文章」の明晰性の改善の観点 ~「言い換え規則」~

---

### 「言い換え規則」の抽出方法

「検討体制」メンバー(知財、特許翻訳、機械翻訳、情報工学等の専門家)による作業



文書変換のプロセス  
→日本語から日本語への翻訳

内容の同一性を保つ変換  
人がより理解しやすい「言い換え規則」の抽出

翻訳(機械翻訳含む)の適用  
→日本語から英語への翻訳

「翻訳用原稿」を訳した結果  
訳質が改善する「言い換え規則」の抽出

7

## 「言い換え規則」の例

---

8

(再掲)

## 「特許版・産業日本語」の取り組み

平成24年度

①特許ライティングマニュアル作成

「－特許ライティングマニュアル－」

②特許オントロジー、図式表現の検討成果を整理

「－図式クレームに基づく請求項文ライティング(構造化レベル)－」

「－図式クレームに基づく請求項文ライティング(オントロジー化レベル)－」

③特許ライティング支援システムを活用した支援機能をまとめる

「－総合的な特許ライティング支援環境－」

9

特許版・産業日本語の取り組み紹介  
－ 特許ライティングマニュアル －

ご清聴ありがとうございました。



(財)日本特許情報機構  
松田 成正

E-mail : shigemasa\_matsuda@japio.or.jp  
URL : <http://www.japio.or.jp/>

一般財団法人  
**日本特許情報機構**  
Japio Japan Patent Information Organization

10

# 図式クレームに基づく請求項文ライティング - 構造化レベル -

横井俊夫

一般財団法人日本特許情報機構特許情報研究所

## 目次：

1. 請求項文ライティングとは
  2. 構造化図式クレーム
    - 2.1 産業日本語をデザインする
    - 2.2 構造化日本語
      - 2.2.1 構造化をデザインする
      - 2.2.2 構造化文章のテキスト処理
    - 2.3 構造化日本語に基づく図式クレーム - 新たなパテントマップ
      - 2.3.1 構造化図式クレームの表現仕様
      - 2.3.2 図式クレームの基本パターン
  3. 図式クレームに基づく請求項文ライティング
    - 3.1 ライティングのモデル
    - 3.2 ライティングのプロセス
- 参考文献
- 

## 1. 請求項文ライティングとは

特許請求の範囲ライティングが、特許ライティングの骨格となり、請求項文ライティングが、特許請求の範囲ライティングの土台となる。そして、以下の理由から、特許ライティングの中で請求項文ライティングに焦点を絞る。

### (1) 特許文章の多くの課題が請求項文に由来

請求項文は、二文制約（容認性を損なわないように中央埋め込みを 2 文以内とするという規則）に反する長文である。特許文書の各処に請求項文由来の文章が用いられる。特許文書の文章上の多くの課題が請求項文に由来する（図 1）。

### (2) 文章ライティングルール作りを、実データ分析ではなく、モデル化に基づいて

文章ライティングのルール作りのために、23 年度半ばまでは、実特許文章の分析作業を中心にすすめた。この分析作業そのものは、非常に有意義であり、特許文章の実体を具体的に把握することができた。しかしながら、特許文書の著作権問題から、当初予定していたマニュアル制作に分析結果を直接利用することが不可であることが判明した。マニュアルにおけるルール説明には、実例に基づく具体事例が不可欠である。そこで、実データ分析に基づく文章ライティングのルール作りという方針を転換し、モデル化に基づく文章ライティングの枠組み作りに重点を移すことにした。モデル化の対象として請求項文ライティングに焦点を絞り、モデル化の手立て

として図式クレームを採用する。図式クレームに基づく請求項文ライティングは、マニュアル化や支援システム化の効果が大きく、アピール力が非常に高い。

請求項文ライティングの内容を定めるために、請求項文ライティングと請求項ライティングとは異なるということを説明しておく。すなわち、

#### **請求項ライティング：**

請求項や特許請求の範囲は、発明という技術思想そのものを体現する。それらをどのように設計し、制作するかは、知財専門家の英知を傾けた作業となる。請求項ライティングのマニュアル化や支援システム化には、十分な準備と注意深いアプローチが必須である。すなわち、安易な取り組みは避けるべきである。

#### **請求項文ライティング：**

文章としての請求項文に注視する。したがって、次のようなファクターに関する検討は、とりあえず、含まない。すなわち、権利を強力で安定的なものにするために特許請求の範囲全体をどう設計するのか、拒絶への対応をどう考慮しておくのか、多重従属制限や申請料等の関する各国の細則への対応をどうするのか、等々はとりあえず除外する。なお、請求項文は、文章論的には、むしろシンプルである。誰が読んでも、一定の確定的な内容となる文章である。すなわち、レトリックやメタファーは用いられない、書き手に依存する態度情報はない、否定等の論理的曖昧さを生ずる可能性のある表現は避けられている、等々である。

請求項文ライティングが目的とするところは、**請求項文ライティングのマニュアル化や支援システム化によって、知財専門家の請求項ライティングを間接的に支援する仕組みを実現することである。**

請求項文ライティングをモデル化する手立てとして、図式クレームを導入する。以下に、図式クレーム導入の必要性と図式クレームに求められる要件を整理する。

#### (1) 導入の必要性

- ① 知財専門家同士のコミュニケーションツール、知財専門家個人の思索ツールとなる効果的な手段が求められている。
- ② 請求項文の内容を忠実に表現し、簡明に把握できる高水準言語となる表現手段が必要である。
- ③ テキスト形式請求項文への変換、他言語への翻訳、請求項間の類似度計算、請求項からの情報抽出、等々へ効果的に対応できる手段が必要である。
- ④ 請求項のパターン化、オントロジー化によって再利用のための効果的な枠組みを提供できる手段が必要である。

#### (2) 求められる要件

- ① 知財専門家がもつ共通的・日常的な言語直感・知識直感に沿うものであること

- ② 知財専門家にとっては、あくまでもツールであり、その自由な発想を妨げないものであること
- ③ 過去の蓄積にも適用できると同時に、将来的発展を妨げないものであること
- ③ 請求項文ライティングに求められるすべての機能に対応できるものであること

次に、請求項文ライティングの大前提となる特許ライティングに対する考え方を説明しておく。

発明の着想から特許文書一次原稿の作成にいたるプロセスは、非常に多様である。本稿では、特許ライティングのモデルプロセス（図 2）に示したように、技術文書（技術論文や技術報告書）をまとめ終えた段階をライティングの出発点とする。したがって、如何に新発明を発想するか、いかなる発想支援機能を利用するか等の多くの特許文書ライティングの指南書が触れている事項は含まない。そのような事項に係わる段階は、いかなる技術をどのように研究開発するかを試行錯誤する過程の一部であると考えられる。

技術文書にまとめられた技術成果から発明となる部分を切り出し、技術成果を発明という技術思想に創作し直し、この技術思想を特許文書として具現化する。このプロセスが、本稿の特許ライティングが対象とするライティングプロセスである。

技術成果、技術の学術的な成果がすべて発明の対象になるわけではない。少なくとも、合理的な発見は発明とはならない。物の構成、あるいは、方法の手順として、目的とする機能が実現可能、あるいは、実施可能なものでなければならない。当然のことながら、新規性は大前提である。

技術文書の内容で発明の対象と想定された部分が、発明提案書という〇次原稿にまとめられる。発明提案書の構成は、課題、解決手段、効果、概略図面等からなる。図 2 のモデルプロセスでは、発明者と知財専門家との協調作業によって、〇次原稿から一次原稿が作成されるとなっているが、ここでは、この過程を以下のように詳細化する。

#### 発明の提案

→発明の概念仕様：発明提案書（〇次原稿）の作成

↓↑

#### 発明の詳細化と先行技術調査

→発明の詳細仕様：従来技術の問題点と特許請求の範囲の詳細化

↓↑

#### 発明の具現化

→発明の実装仕様：明細書、図面、要約書の具体化

↓↑

#### 発明の実現

→表わす日本語による特許文書：一次原稿の作成

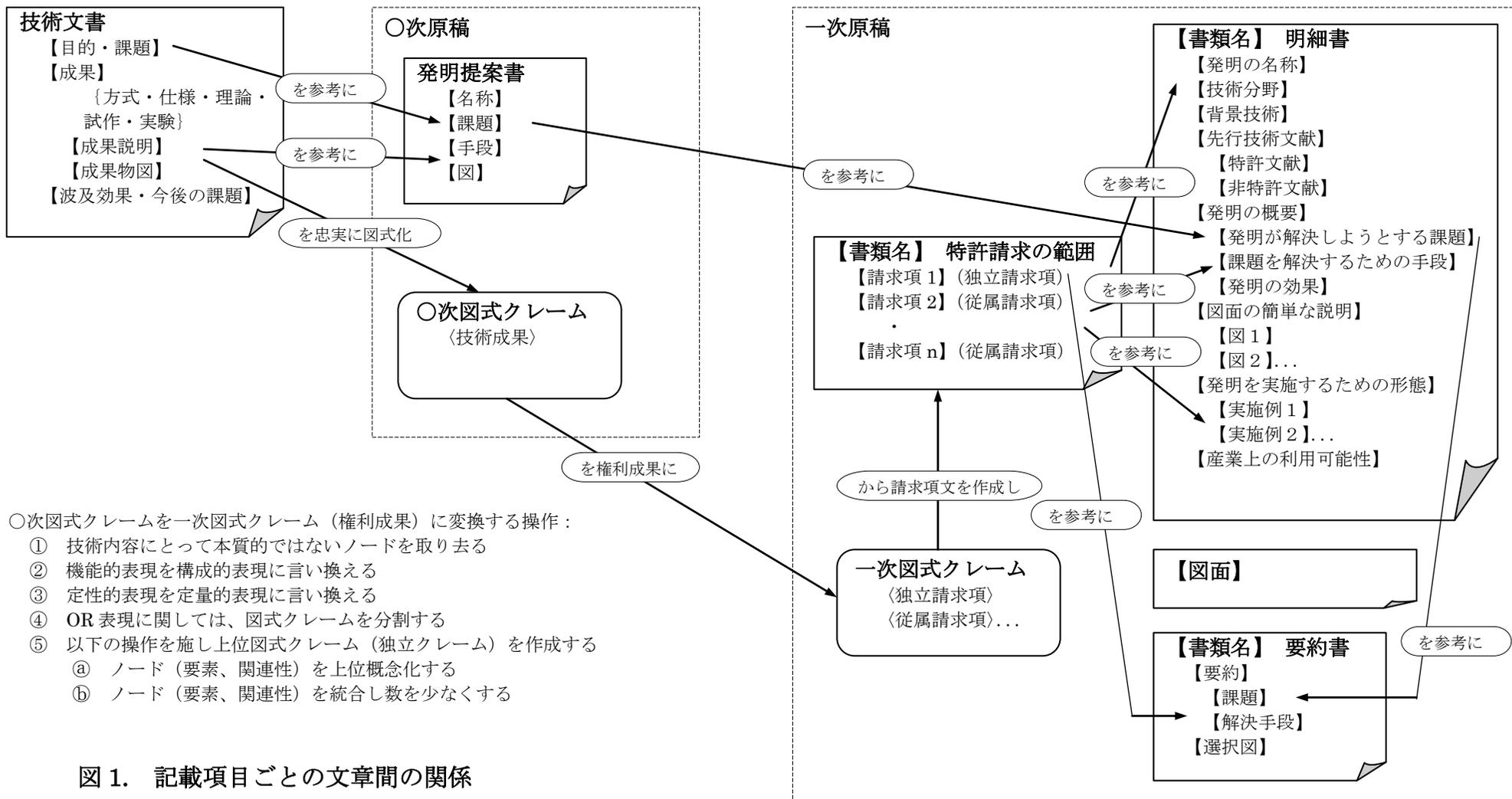


図 1. 記載項目ごとの文章間の関係

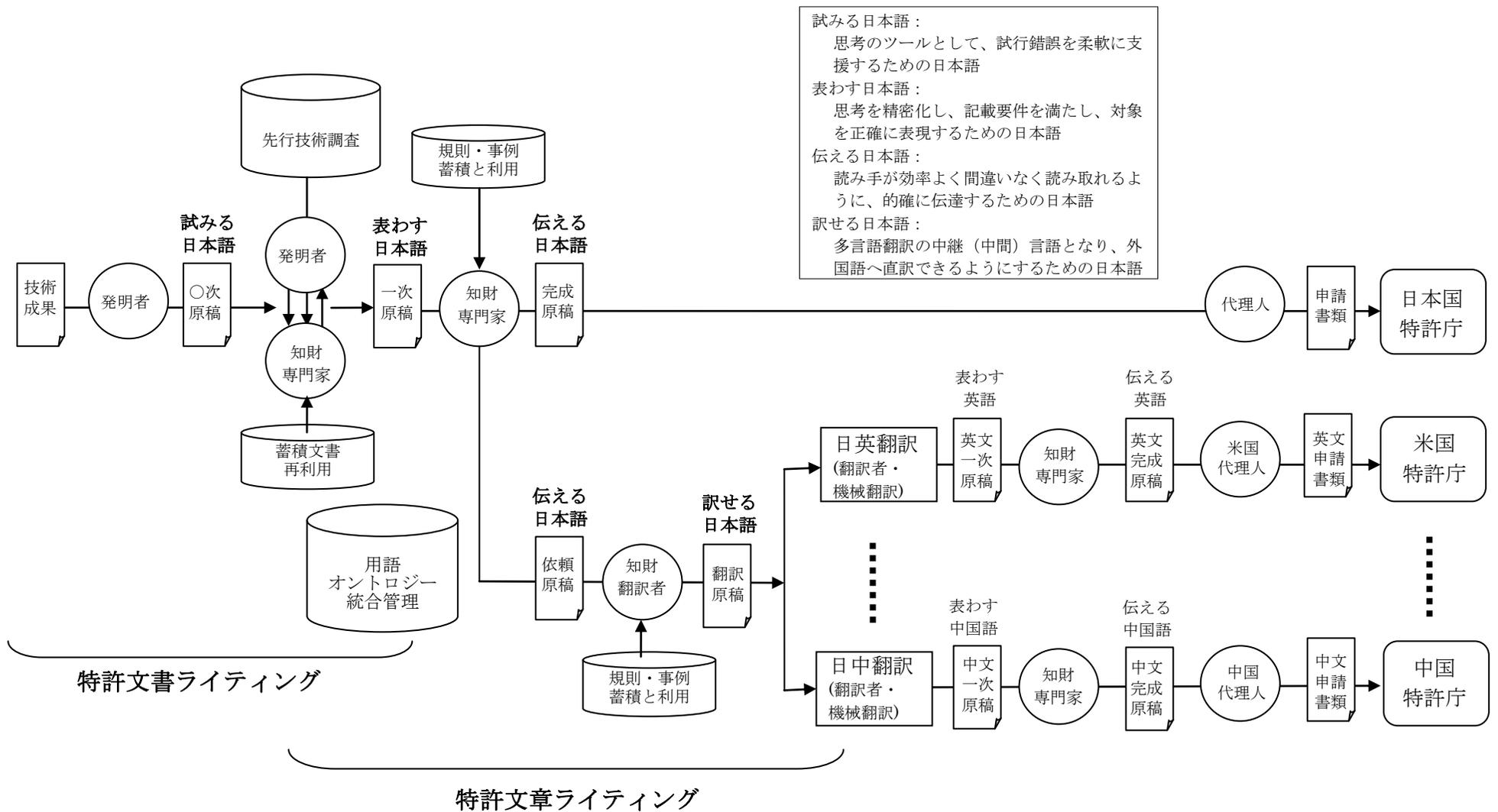


図 2. 特許ライティングのモデルプロセス

## 2. 構造化図式クレーム

### 2.1 産業日本語をデザインする

言語への新たなアプローチとして、合成的なアプローチ、すなわち、言語を工学的にデザインするというアプローチを提案する。そして、このアプローチに基づいて、構造化日本語、さらには、諸外国語に普遍化した構造化言語という言語系を提案する[1]。構造化言語によって表現（記述）された構造化テキスト（構造化文書・構造化文章）によって、文書処理（作成、検索、翻訳、要約、情報抽出等々）が格段に効率化され高機能化される。

本稿では、日本語と和文テキストを中心に、すなわち、構造化日本語と構造化和文テキストを中心に説明する。構造化日本語は、表現対象となる情報が本来保有する構造により密着できるようにするためにグラフィカルな表現要素が組み込まれた日本語である。この構造化日本語に対し、通常日本語を線状化日本語と呼ぶ。

Mind Map[2]やConceptual Mapping等々、様々なグラフィカルな表現形式が提案され、利用されている。構造化日本語は、これらの既存の図式表現形式と大きく異なる考え方に基づいてデザインされた言語である。構造化日本語は、あくまでも、日本語である。素直に、日本語として、読み書くことが出来る。

構造化和文テキストを読み下すことによって、線状化和文テキストが得られる。構造化和文テキストは、読み方の方略にしたがって、多数の読み下し方が可能である。すなわち、ひとつの構造化和文テキストに多数の、時には、無数の線状化和文テキストが対応付けられる。この事実が、構造化和文テキストが表現対象の情報の構造に如何に密着した表現形式であるかを示している。

構造化日本語、構造化和文テキストは、論理的な明晰さや厳密さが強く求められる情報の表現に適している。仕様書、契約書、特許文書、レファレンス文書、設計・運用マニュアル等々の産業技術文書、さらには、法令・司法関連文書などへ効果的に適用できる。また、従来、散文的に語るしかできなかったテクニカルライティングが、体系立った工学的技術として議論できるようになる。すなわち、構造化日本語は、デザインされた産業日本語である。

つぎに、言語をデザインするとはである。言語をデザインするとは、人間に知覚される言語のアフォーダンス[3]を高め、高度な利用価値をもたらし、効果的で効率的なインタラクションを実現するような言語を設計することである。人間と言語とのインタラクションには、2つの側面がある。ひとつは、言語そのものとのインタラクションであり、もうひとつは、言語を媒介とした表現対象とのインタラクションである。

まず、言語そのものとのインタラクションにおけるアフォーダンスである。ユーザーの母語、ユーザーのその分野における日常的な母語の用法、これらから外れれば外れるほど、そのユーザーにとってのアフォーダンスは低下する。したがって、日本人ユーザーにとつ

ては、日本語であること、そして、産業日本語としては、その産業技術分野の日常化された用法に則ること、これらがアフォーダンスを高める。

このアフォーダンスの観点に立てば、エスペラントをはじめとする国際共通語 (Common Language) が何故普及しないのかが明らかとなる。共通化のために、誰にとってもアフォーダンスの低い言語が無理やり設計されるからである。その点、Globish[4]のように、すでに世界で広く使われている言語 (英語) をベースにする国際共通語は、健全である。

次に、表現対象とのインタラクションにおけるアフォーダンスである。表現対象となる情報を出来るだけ素直になぞることが出来れば、アフォーダンスは高まる。ただし、この「なぞる」というところに課題がある。情報は、言語から独立して存在するのではなく、言語によって表現されて、初めて、情報となる。したがって、アフォーダンスを高めるには、人間による情報の理解過程・伝達過程に沿うことが肝要となる。

このアフォーダンスの観点に立てば、制限英語や制限日本語などの制限言語 (Controlled Language) における「制限する」ということと「デザインする」ということが本質的に異なることが明らかとなる。機械処理の都合のために制限したりすることや簡易化したりすることは、多くの場合、人間のためのアフォーダンスを高めることと真逆の行為であるからである。

## 2.2 構造化日本語

デザインされた産業日本語として、構造化日本語を提案する。構造化日本語は、あくまでも日本語である。日本語として書き、日本語として読むことが出来る。この、日本語であるということが重要である。

表現対象となる情報は、三次元空間に時間軸を加味した四次元空間を構成する。ことに、物理的な具象物に係わる情報はそうである。そして、抽象物に係わる情報は、具象物へのメタファーを用いて構成される。一方、言語は一次元である。その誕生の過程からも、また、幼児の言語習得の過程からも、言語の主軸となるのは、音声言語である。発声器官や聴覚器官の線状性 (一次元性) が、音声言語に線状構造を課すことになる。文字言語も、音声言語に準じて、基本は線状構造となる。

言語による情報の表現や言語表現からの情報の読み取りには、四次元の情報と一次元の言語表現との間のマッピング操作が求められることになる。もちろん、言語には、マッピングのための色々な仕組が、形態的にも構文的にも用意されてはいるが、十分に手当てされているというわけではない。手当ての不十分さや手当ての不適切な使用が、言語表現に曖昧さや非明晰さを生む要因となる。

構造化日本語では、表現対象とのインタラクションにおけるアフォーダンスを高めるために、(自然) 日本語の線状性という制約を緩める。そして、日本語としての基本的な仕組を

保持することによって、言語そのものとのインタラクションにおけるアフォーダンスを維持し、あるいは、より高める。このようにデザインされた日本語が構造化日本語である。構造化日本語は、情報を論理的に明晰に表現するのに適していることから、デザインされた産業日本語となる。

### 2.2.1 構造化をデザインする

構造化日本語をデザインするためには、まず、構造化そのものをデザインしなければならない。情報は四次元空間を構成するとした。しかし、構造化が、この四次元に直接対応するわけではない。構造化が対応すべきなのは、四次元情報が言語表現されたときに持つ構造である。情報の構造は、言語表現されることによって定まる。言語によって、情報をどのように理解していくのか、情報をどのように伝えるのか、すなわち、言語による情報の理解過程・伝達過程が創り出す構造がデザインすべき構造化である。

そこで、言語による情報の理解過程・伝達過程が創り出す構造化とはである。一例を挙げよう。事柄としての情報を<A 社が B 装置を開発した>であるとしよう。この情報の言語表現として、以下のようなものを対応付けることができる。

- ①「A 社が B 装置を開発した。」
- ②「A 社は、B 装置を開発した。」
- ③「B 装置は、A 社が開発した。」
- ④「A 社が開発したのは、B 装置である。」
- ⑤「B 装置を開発したのは、A 社である。」

この 5 つの文は、事柄としての情報としては、同じものを表現している。事柄としての情報の構造としては、以下の事象構造[5]がすべての文に対応付けられる。

[事象 時間性=過去：

事象基→<開発する>

動作主→<A 社>

動作対象→<B 装置>]

それでは、5 つの文の違いは、何を反映しているのであろうか。これらの違いが反映するのは、人が事象構造をどのように理解しようとしているのか、どのように伝えようとしているのか、そこにおける違いである。人間による情報の理解や伝達の過程をまとめて情報の処理過程と呼ぶとすると、反映するのは、この情報の処理過程の違いである。そこで、この情報の処理過程の構造を処理構造と呼び、その構造化に取り組むことになる。

情報の処理構造は、言語学における談話構造に関する議論の中で、適切とは言い難い用語ではあるが、情報構造と呼ばれているものにおおよそ対応する。情報構造では、文の構造を 2 つの成分からなる構造に見立てる[6]。旧情報成分と新情報成分、あるいは、既知情報

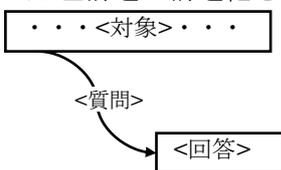
成分と未知情報成分の2つである。

言語学においては、この情報構造をもう一步進めて、文連鎖にまたがる理解や伝達を対象にした議論の展開がある[7]。この議論では、3つの成分からなる構造が見立てられる。そう見做すことができる。質問対象成分と質問成分と回答成分、あるいは、予測対象成分と予測成分と結果成分の3つである。構造化日本語では、この3つの成分からなる構造を前提にして、構造化をデザインする。基本的な構造化として、<対象><質問><回答>という三つ組み構造を採用する。

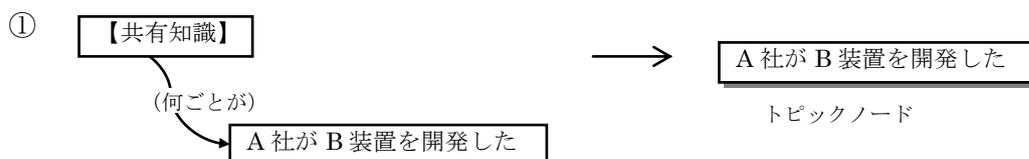
5つの文の情報の処理構造は、その違いを反映し以下のようなになる。対象成分は、それぞれの文の前に位置する文の回答成分である。そして、回答成分は、後に位置する文の対象成分となる。

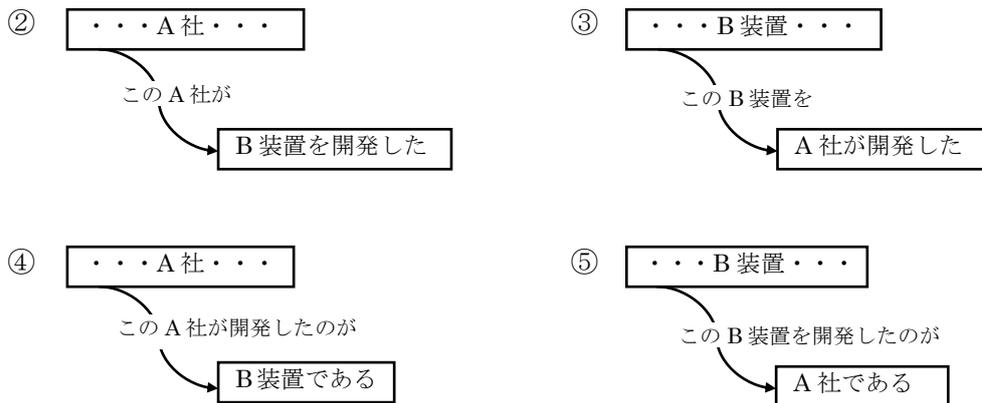
- ①[処理：対象→<【共有知識】>  
質問→<(何ごとが)>  
回答→<A社がB装置を開発した>]
- ②[処理：対象→<・・・A社・・・>  
質問→<A社が(したのが)>  
回答→<B装置を開発した>]・・・
- ③[処理：対象→<・・・B装置・・・>  
質問→<B装置を(したのが)>  
回答→<A社が開発した>]・・・
- ④[処理：対象→<・・・A社が開発した・・・>  
質問→<A社が開発したのが>  
回答→<B装置である>]・・・
- ⑤[処理：対象→<・・・B装置を開発した・・・>  
質問→<B装置を開発したのが>  
回答→<A社である>]・・・

処理構造の構造化をグラフィカルな表現形式で表わすと、以下となる。



5つの文に対するグラフィカルな表現は、以下となる。





## 2.2.2 構造化文章のテキスト処理

構造化日本語による構造化和文テキストに対するテキスト処理、すなわち、作成、検索、翻訳、要約について概観する。構造化の情報を活用することによって、通常の線状化テキスト処理に比べ、高機能で高精度なテキスト処理が可能となる。ただし、ここでは、作成と翻訳のみを取り上げる。

### (1) 作成

構造化テキストライティングは、以下の3つのライティングステップが交錯しあいながら展開するプロセスとなる。

#### ① 俯瞰と接近

テキスト全体の全容を把握するために俯瞰し、テキスト各部分の詳細を詰めるために接近する。俯瞰と接近がダイナミックに繰り返される。

#### ② 書き加え

不全感を感じるノードに対して質問アークと回答ノードを書き加える。感じ取られた不全感を質問として表現し、不全感を満たすように回答を書き足す。

#### ③ 書き換え

質問アークを付け替えることによって、部分構造化テキスト全体の移動操作がなされる。その際に求められる<質問>や<回答>の調整内容も明解に把握される。

いずれ、このライティングプロセスを支援するグラフィカルなエディターが用意されることになる。ワードプロセッサではなく、構造プロセッサである

この構造プロセッサには、構造化テキストと線状化テキスト（通常のテキスト）との間の相互変換の機能も用意される。広く流通する文書では、現状のままの形式が継続することになる。したがって、構造化テキストを線状化する機能が必要となる。一方、膨大に蓄積されたテキストを再利用するために、線状化テキストを構造化する機能が必要となる。

構造化テキストの線状化の操作は、二次元グラフをたどりながら構造化日本語テキストを読み下す操作となる。ひとつの構造化テキストに多数の、場合によっては、非常に多数の読み方、すなわち、線状化テキストが対応付けられる。そこで、基本的な読み方として、いくつかの推奨方略が用意されることになる。

## (2) 翻訳

構造化日本語が表現する構造は、情報が本来持つべき構造である。したがって、基本的には、この構造は、言語に依存しない、あるいは、そう仮定してもそれほどの不都合は生じない。ただし、<質問>と<回答>という構造は、言語に依存せずに保存されるとしても、<質問>部分と<回答>部分の言語表現の方式は、言語に依存することになる。日本語のような主題（話題）優勢言語と英語のような主語優勢言語[8]では、異なってくる。すなわち、

主題優勢言語：<質問>=<主題成分> <回答>=<題述成分>

主語優勢言語：<質問>=<主語成分> <回答>=<述語成分>

この<質問>部分と<回答>部分の表現方式を変換する仕組みを整備すれば、ほぼ現状の機械翻訳システムを用いて、構造化言語間に高精度な機械翻訳を実現することができる。なお、<質問>と<回答>の基本的な表現方式のパターンは、20~30程度に整理されると思われる。例えばとして、3.1節で用いた5つの例を英訳すると以下となる。

① A-company developed B-apparatus

② . . . A-company . . .

This A-company

developed B-apparatus

③ . . . B-apparatus . . .

This B-apparatus

was developed by A-company

④ . . . A-company . . .

This A-company developed

B-apparatus

⑤ . . . B-apparatus . . .

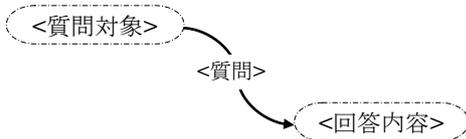
This B-apparatus was developed by

A-company

## 2.3 構造化日本語に基づく図式クレーム - 新たなパテントマップ

### 2.3.1 構造化図式クレームの表現仕様

構造化図式クレームでは、関係アークを以下のように特化する。



想定される質問は、以下のように整理される。ただし、以下の<質問>仕様は、当面の図式クレームに必要となるものに限定する。

<質問>:=

- <対象に係わる事象を問う質問> |
- <事象に係わる対象を問う質問> /\*分裂構文形式の質問\*/ |
- <事象に関係する事象を問う質問>

<対象に係わる事象を問う質問>:=

- <対象が単一の場合の質問> |
- <対象が確定複数の場合の質問> |
- <対象が任意複数の場合の質問>

<対象が単一の場合の質問>:=

- これ<格辞> /\*全体に対して問う\*/
- /\*格辞とは、格助詞や格助詞相当語句をさす\*/ |
- この<部分><格辞> /\*全体の<部分>に対して問う\*/ |
- その<属性><格辞> /\*全体の<属性>に対して問う\*/ |
- その<構成要素><格辞> /\*全体の<構成要素>に対して問う\*/ |
- この<部分>の<属性><格辞> /\*<部分>の<属性>に対して問う\*/ |
- この<部分>の<構成要素><格辞> /\*<部分>の<構成要素>に対して問う\*/

<対象が確定複数の場合の質問>:=

- これら<格辞> /\*全部に対して問う\*/ |
- この第 i<格辞> /\*i 番目に対して問う\*/ |

<対象が任意複数の場合の質問>:=

- これら<格辞> /\*全部に対して問う\*/ |
- この各々<格辞> /\*各々に対して問う\*/ |

<事象に係わる対象を問う質問>:=

<問われる対象><問われる対象要素と回答要素を除く事象>のが

<問われる対象>:=<対象に係わる事象を問う質問>

<事象に関係する事象を問う質問>:=

<順接原因となる事象を問う質問> |

<順接結果となる事象を問う質問> |

<逆接原因となる事象を問う質問> |

<逆接結果となる事象を問う質問> |

<仮定条件前件となる事象を問う質問> |

<仮定条件後件となる事象を問う質問> |

<例示となる事象を問う質問> |

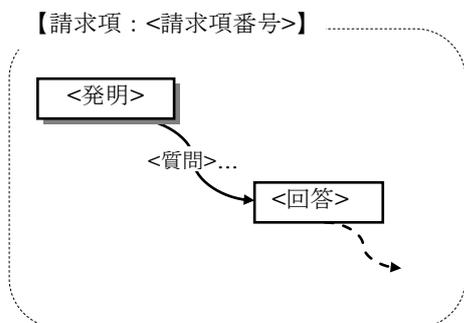
<汎化となる事象を問う質問> |

<AND となる事象を問う質問> |

<OR となる事象を問う質問>

### 2.3.2 図式クレームの基本パターン

図式クレームの構成は、構造化日本語における構造化文章パターンとして以下のように表現される。この構成を基にして、図式クレームの基本パターンを定める。基本パターンは、請求項の発明内容を類別するための基本的要件にしたがって定められる。



独立請求項に対応する図式クレームの基本パターンを下表のように定める。

独立請求項	発明の対象		取り上げ方		視点の定め方	
	物	方法	発明のみ	従来と対比	要件列挙	関連列記
基本パターン 1	○		○		○	
基本パターン 2	○		○			○
基本パターン 3	○			○	○	
基本パターン 4	○			○		○
基本パターン 5		○	○		○	
基本パターン 6		○	○			○
基本パターン 7		○		○	○	
基本パターン 8		○		○		○

従属請求項に対応する図式クレームの基本パターンを下表のように定める。内的付加と外的付加を同時に含む場合には、2つの基本パターンを融合させたパターンによって対応する。

従属請求項	発明のタイプ		付加の方法	
	物	方法	内的	外的
基本パターン 9	○		○	
基本パターン 10	○			○
基本パターン 11		○	○	
基本パターン 12		○		○

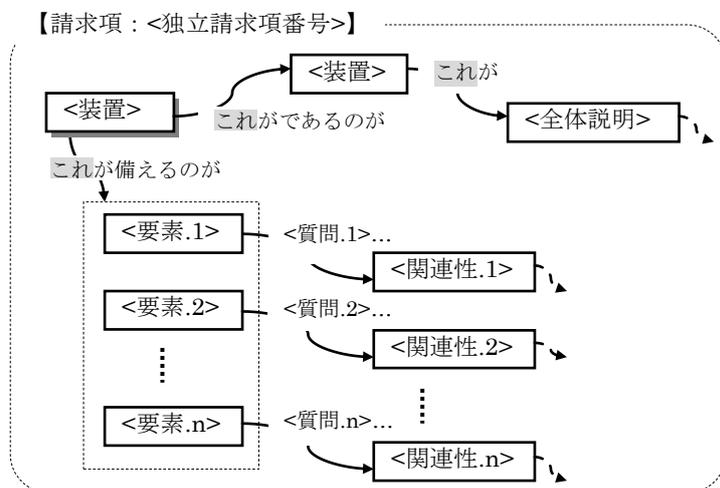
さらに、従属請求項に対する基本パターンとして、下表の2つを加える。<方法>を用いての<物>、<物>を用いての<方法>に対応する基本パターンである。

従属請求項	発明のタイプ		用いる手段	
	物	方法	方法	物
基本パターン 13	○		○	
基本パターン 14		○		○

以下に、以下に、基本パターン 1、基本パターン 7、基本パターン 9、基本パターン 10 を例示する。

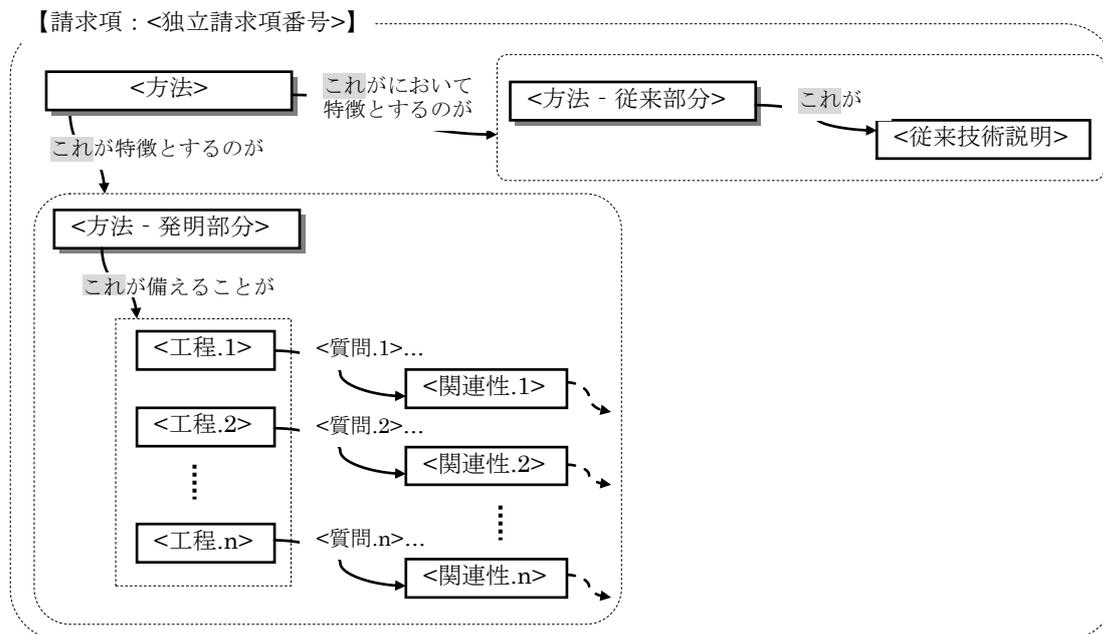
### [基本パターン 1]

物に関する発明で、発明技術のみを取り上げ、構成要素を列挙する独立請求項



### [基本パターン 7]

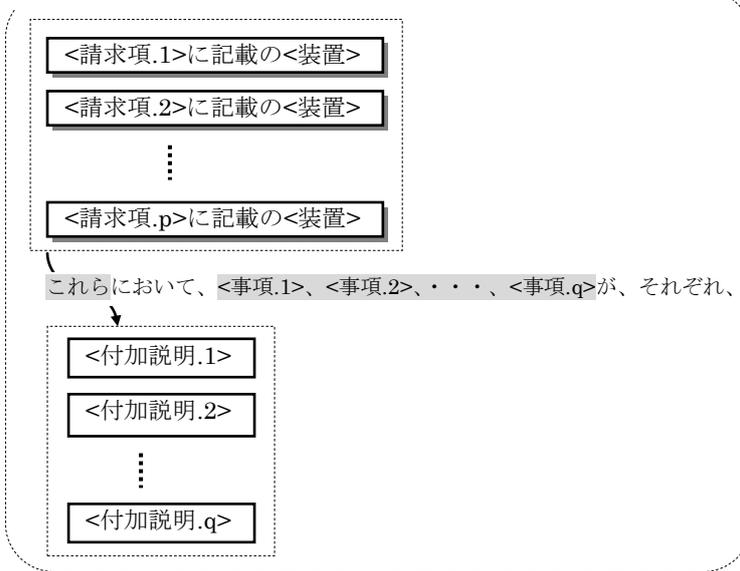
方法に関する発明で、従来技術と対比させ、要件を列挙する独立請求項



### [基本パターン 9]

物に関する発明で、内的付加による従属請求項

【請求項：<従属請求項番号>】

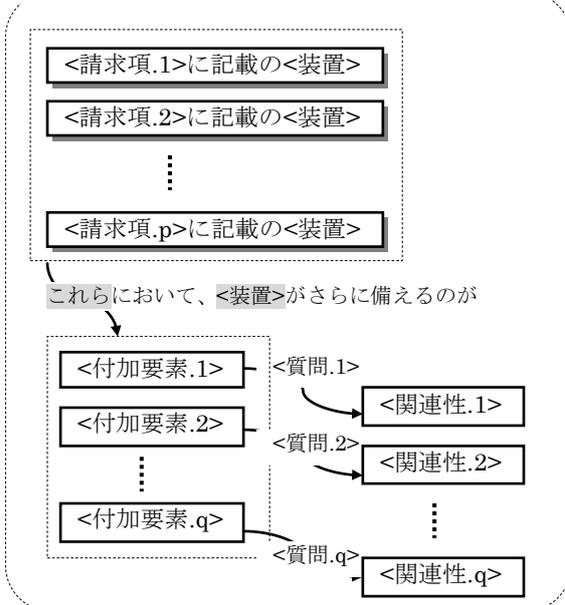


### [基本パターン 10]

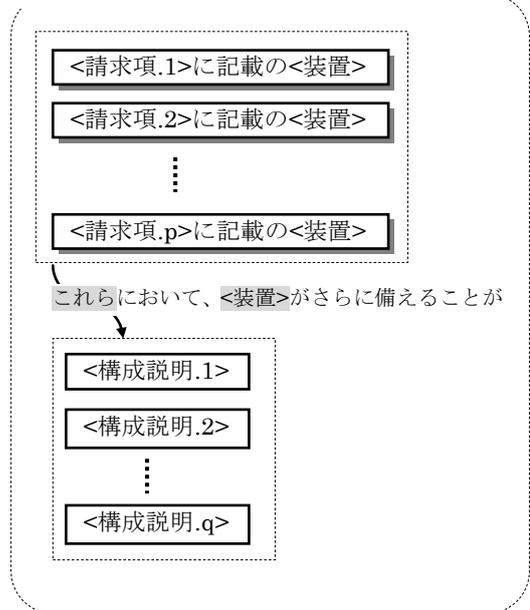
物に関する発明で、外的付加による従属請求項

(左：要件列挙、右：関連列記)

【請求項：<従属請求項番号>】



【請求項：<従属請求項番号>】



### 3. 図式クレームに基づく請求項文ライティング

#### 3.1 ライティングのモデル

請求項文ライティングの図式クレームに基づくライティングモデルを図3に示す。

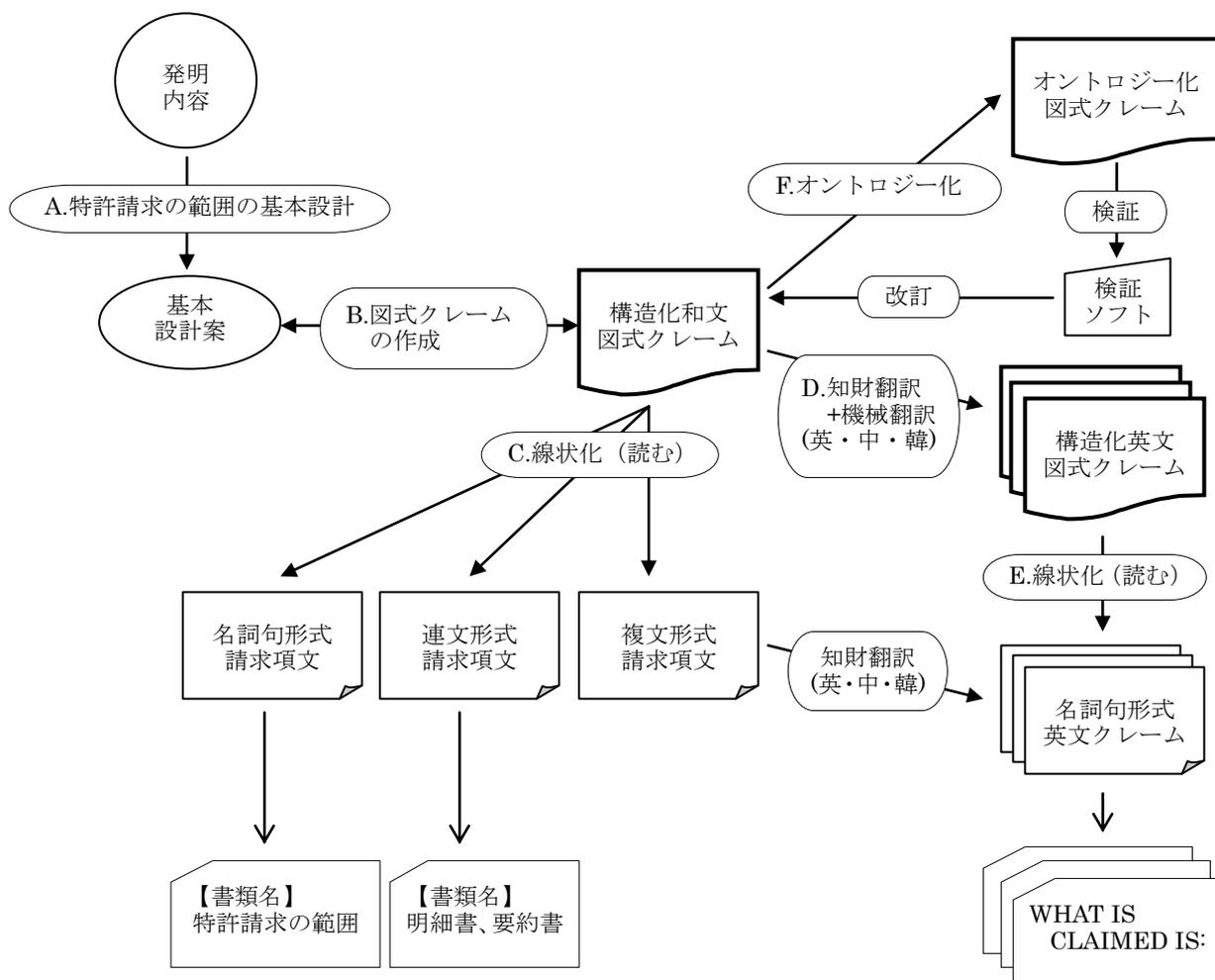


図3. 図式クレームに基づく請求項文ライティングのライティングモデル

## 3.2 ライティングのプロセス

請求項文ライティングのステップは、図式クレームを用いる以下のようなプロセスである。本稿では、例題となる請求項文に関して、著作者の最終的な許諾が得られていないためにプロセスの具体的な説明は割愛する。

- A. 特許請求の範囲の基本設計
- B. 構造化図式クレームの作成
  - B.1 基本パターンの選択
  - B.2 実現パターンへの書き換え
  - B.3 図式クレームへの書き換え
- C. 構造化図式クレームの線状化
  - C.1 名詞句形式テキストへと読む
  - C.2 複文形式テキストへと読む
  - C.3 連文形式テキストへと読む
- D. 構造化図式クレームの翻訳（英訳）
  - D.1 翻訳原稿の作成
  - D.2 英文図式クレームへの翻訳
- E. 英文図式クレームの線状化
  - E.1 名詞句形式テキストへと読む
- F. 構造化図式クレームのオントロジー化

## 参考文献

- [1]横井俊夫：言葉をデザインする - デザインされた産業日本語 - 、Japio YEAR BOOK 2012（2012年11月）
- [2]トニー・ブザン、バリー・ブザン著、神田昌典訳：ザ・マインドマップ、ダイヤモンド社（2005年11月）
- [3]佐々木正人：アフォーダンスー新しい認知の理論、岩波書店（1994年）
- [4]ジャン=ポール・ネリエール、ディピッド・ホン著、一般財団法人グローバル人材開発訳：世界のグロービッシュ、東洋経済新報社（2011年3月）
- [5]町田 健：言語構造基礎論 - 文の意味と構造、勁草書房（2011年10月）
- [6]堀川智也：日本語の「主題」、ひつじ書房（2012年）
- [7]石黒 圭：日本語の文章理解過程における予測の型と機能、ひつじ書房（2008年）
- [8]吉村公宏：英語世界の表現スタイルー「捉え方」の視点から、青灯社（2011年5月）

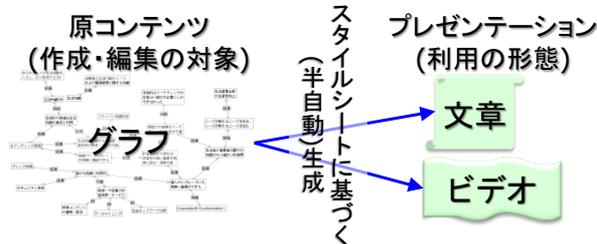
# 図式クレームに基づく 請求項文ライティング — オントロジー化レベル —

2013-03-01

橋田 浩一@産総研

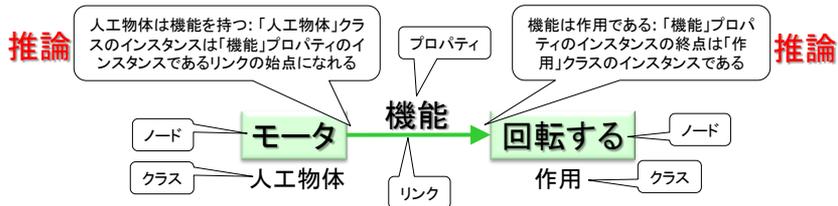
## ビジョン

- 原コンテンツ(作成・編集の対象)はオントロジーに基づくグラフ表現
- コンテンツの種類や目的に応じた仕方で作成・編集
  - 請求項を具体化して実施例を作成
  - 実施例を抽象化して請求項を作成
  - 審査のコメントに応じて改訂
- 原コンテンツから多様なプレゼンテーションを生成
  - 文章やビデオや朗読等のスタイルシートを適用することにより、グラフから各種のプレゼンテーションを生成する。
  - 特許請求項等は完全自動生成



# オントロジー

- 有限個のクラスとプロパティとそれらの間の制約
- クラス: 1項述語(集合)
  - RDFグラフのノードでインスタンスを表現
  - 上位オントロジーでは数十個: 人工物体、性質、作用、形状、...
- プロパティ: 2項関係
  - RDFグラフのリンクでインスタンスを表現
  - 上位オントロジーでは数十個: 派生、構成要素、属性、位置、対象、材料、機能、目的、状況、部分構造、内部構造、...



3

# グラフ構造の効能

風が吹くと土ぼこりが立つ。その土ぼこりが目に入る。それで盲人が増える。その盲人が三味線を買う。ゆえにその三味線を作るため猫が殺される。だから鼠が増える。増えた鼠が桶をいじるので、桶の需要が増える。それで桶屋が儲かる。

論理的構造がすぐわかる ↓ 加工が容易

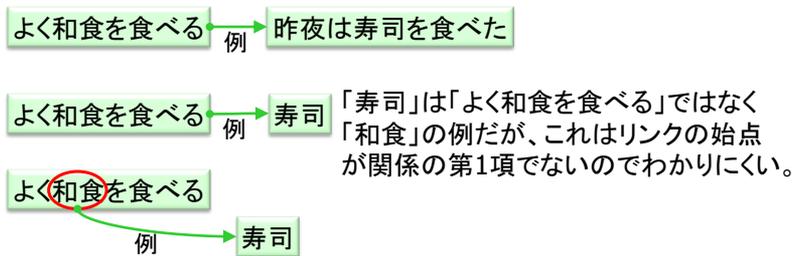


要約

4

## ノード内のテキストを読まなくても論理的構造がすぐわかるための要件

- リンクの意味が瞬時にわかる。
  - リンクのタイプ(プロパティ)の個数が100程度以下であり、その名前から意味が即座にわかる。
  - リンクの端点が関係の項である。



5

## プロパティの名前Pとその意味

- 関係名詞: 始点のPが終点
  - 健の妻が奈緒美
  - 容器を使う
- 他動詞: 始点が終点をP
  - 健が奈緒美を愛する
- 前置詞・後置詞: 始点P終点
  - dinner after work
  - 健が帰る
- その他
  - 宏が健の友人

6

# オントロジーに基づく図式クレーム

- 請求項の意味的明確化(簡略化ではない)
  - 請求項の通常のテキスト表現と意味的に等価
  - テキスト表現より曖昧性が低く、理解と作成が容易
- 作り方
  - プロパティとして定義されている関係は原則としてリンクで表現する。
  - ノードの共有で表わしにくい共参照はテキストに埋め込んだ**インデクス**の共有で表現する。

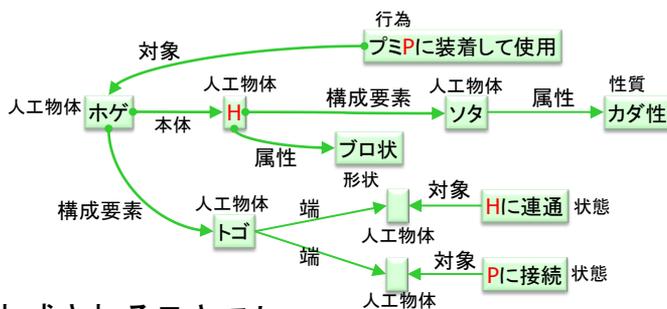
制御装置C

CとPの間にRを設置

このノードをグラフで表現してインデクスの埋め込みをなくすこともできるが、表現が複雑になる

7

## 図式クレームの例



自動生成されるテキスト:

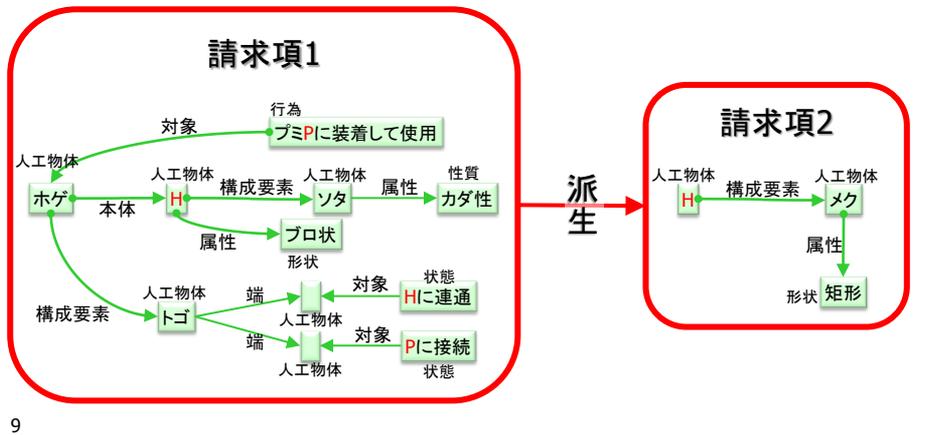
- 次のようなホゲ。
- プミPに装着して使用する。
- ホゲ本体Hはプロ状であり、カダ性を有するソタから構成される。
- 一端がHに連通し他端がPに接続するトゴを備える。

プロパティに対応

8

## 請求項の間の派生関係

- 請求項への部分的変更(置換と追加)によって従属請求項が派生する。



9

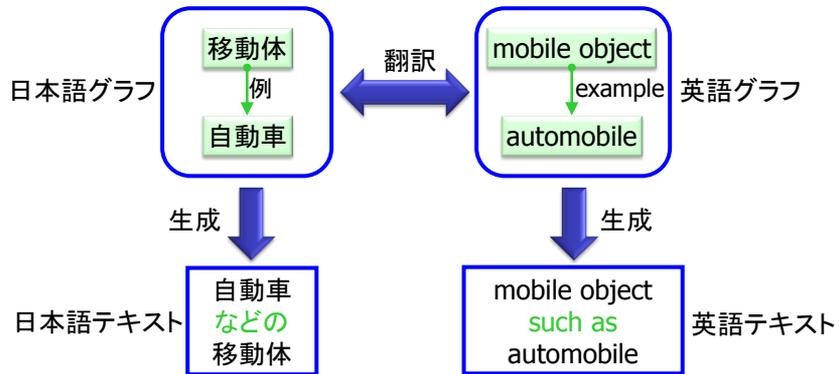
## 特許文書の作成と活用の促進

- 理解・作成を容易に
  - 発明者と弁理士との協働による作成・編集
  - 審査官の指摘に対応する修正
  - 技術を使いたい人が発明の内容を簡単に理解
- 活用を容易に
  - 相互運用性
    - グラフ表現のスキーマ(オントロジー)の標準化
  - 自動推論による意味的妥当性の検証
    - 例: 人工物体の構成要素は物体である
  - 検索、翻訳、読解支援、...

10

## 例: 翻訳

- グラフの翻訳 = 各ノードのテキストの翻訳
- グラフから生成した文章を翻訳するのではない。



# 総合的な特許ライティング支援環境

IRD国際特許事務所 所長・弁理士 博士(情報学)  
特許版・産業日本語委員会委員  
谷川 英和  
htanigawa@ird-pat.com

## 目次

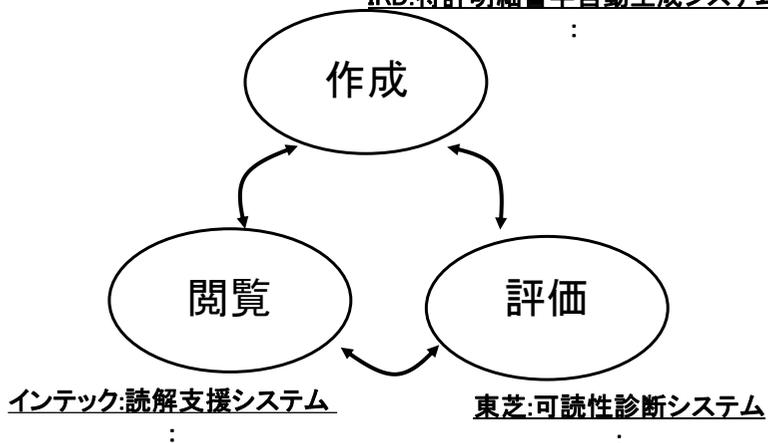
- 1.特許ツールMAP
- 2.特許ライティング支援環境
- 3.特許明細書半自動生成システム(PatentGenerator)
- 4.読解支援システム(Patent Structure Viewer)
- 5.可読性診断システム(KnowledgeMeister文書診断)
6. 評価結果と今後

# 1.特許ツールMAP

フェーズ 種類	戦略立案 (企画)	発明構築	特許調査	特許明細書 作成	出願処理	中間処理	登録手続	権利維持	権利無効化	権利行使
業務 推進系	特許検索 ツール	発想支援 ツール	特許検索 ツール	明細書 作成支援 ツール	出願支援 ツール				特許検索 ツール	
	特許マップ 作成支援 ツール	発明支援 ツール	特許マップ 作成支援 ツール	翻訳 ツール					文献検索 ツール	
管理系				案件管理 ツール	特許管理ツール		特許管理ツール		包袋管理ツール	
分析 評価系	特許分析 ツール			特許明細書 分析ツール						特許分析 ツール
										特許価値 評価ツール

# 2.特許ライティング支援環境

IRD:特許明細書半自動生成システム



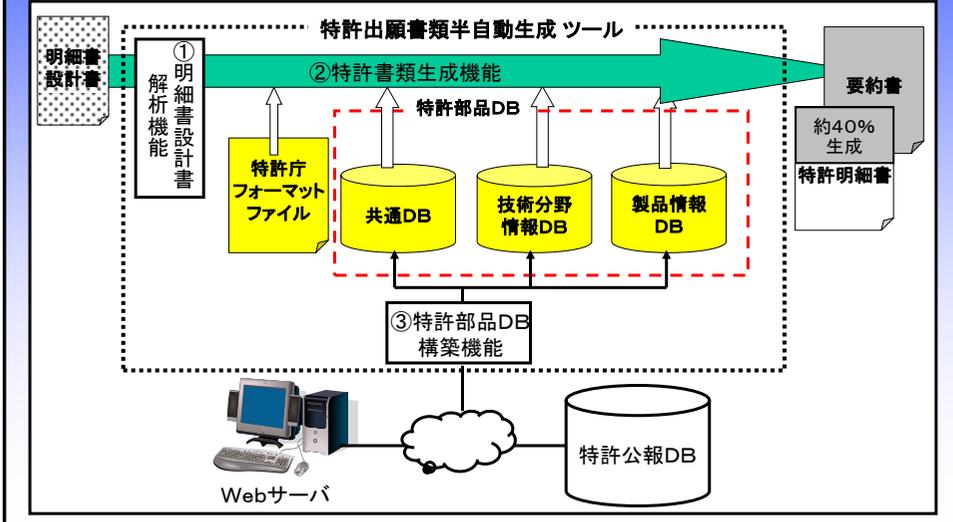
## 特許ライティング支援システムの評価委員

岩永 勇二	平田国際特許事務所 副所長 弁理士
熊野 明	東芝ソリューション株式会社 プラットフォームソリューション事業部 ソフトウェア開発部 参事
新森 昭宏	株式会社インテック 先端技術研究所 研究開発部 部長
谷川 英和	IRD国際特許事務所 所長 弁理士
松本 征二	知崇国際特許事務所 所長 弁理士
的場 成夫	有限会社 夢屋 代表取締役 弁理士
横山 淳一	富士通株式会社 知的財産権本部 特許統括部 シニアマネージャー 弁理士

## 3.特許明細書半自動生成システム (PatentGenerator)

# (1)特許明細書半自動生成システム(PatentGenerator)

弁理士のノウハウ、発明を構成する部品の文章をDB化し、利用する  
 → 特許書類作成効率1.8倍、特許書類品質1.4倍に！



# (2)入力ファイル(明細書設計書)の例

①発明の概要

④全体課題・効果

⑤課題・効果

②実施形態番号

③用語説明

全体課題＞出願書類を効率的に作成できなかった。  
 <全体効果>出願書類を効率的に作成できる。  
 【請求項1】(1)(発明情報、説明情報に基づいて、出願文書情報を作成する)。  
 <課題>出願書類を効率的に作成できなかった。  
 <効果>出願書類を効率的に作成できる。  
 発明は、発明を特定する情報である1以上の発明特定情報を有する情報であり、発明を特定する情報である1以上の発明特定情報を有する発明特定情報に記述される発明情報記憶部と、前記発明特定情報に関する説明を示す情報である説明情報が記述される説明情報記憶部と、前記発明情報記憶部に記憶している発明情報、及び前記説明情報記憶部に記憶している説明情報に基づいて、特許出願の文書に関する情報であり、発明情報に関する1以上の実施の形態を含む情報である出願文書情報を作成する作成部と、前記作成部が作成した出願文書情報を入力する出力部と、を備えた出願文書情報作成装置。  
 【請求項2】(1)(1ー出願文書情報＝説明情報を用いて発明特定情報を説明する文書)。  
 <課題>出願書類を効率的に作成できなかった。  
 <効果>出願書類を効率的に作成できる。  
 前記作成部は、発明情報に発明特定情報に関する説明を示す情報を含む出願文書情報を作成する。請求項1記載の出願文書情報作成装置。  
 【請求項3】(1)(1、2ー発明情報において2以上の発明がグループ化されており、出願文書情報では、一の発明のグループが一の実施の形態に含まれる)。  
 <課題>複数の発明の情報を合成して、出願書類を効率的に作成できなかった。  
 <効果>複数の発明の情報を合成して、出願書類を効率的に作成できる。  
 発明情報は、一の発明に関する情報であり、1以上の発明特定情報を有する情報である単一以上の有する情報であり、当該2以上の単独発明情報が1以上のグループに記述されており、  
 発明情報において一のグループに含まれる発明が同一の実施の形態に含まれるように出願文書情報を作成する。請求項1または請求項2記載の出願文書情報作成装置。

コメント：クレーム、実施形態設計書、クレームの構成要素を単に並べた情報など(発明ごとになっていなくてもよい)。  
 コメント：一時記憶可能。  
 コメント：発明特定情報の一部の説明、全部の説明、関連する説明。  
 コメント：発明でなくともよい。例えば、フローがなくてもよい。クレームが含まれても含まれてもよい。  
 コメント：特許出願できる状態であってもよく、その手前のたき台であってもよい。  
 主に実施の形態の文書。結果や手順、詳細などは、あってもなくてもよい。  
 コメント：表示、書様、送信、印刷など。  
 コメント：実施の形態の番号が示されてもよく、単に、グループに分かれているだけでもよい。

©IRD国際特許事務所 谷川英和 8

### (3)特許庁フォーマットファイル(ひな形情報)

#### 特許明細書の書式

- ・<〇〇>は、システム変数であり、入力ファイルから抽出して、代入される
- ・<<〇〇>>は、ユーザ定義変数であり、共通DBから検索し、代入される

【書類名】明細書

【発明の名称】<すべての発明の名称>

【技術分野】

<技術分野>

【背景技術】

<背景技術>

【先行技術文献】

【特許文献】

<<特許文献例>>

【非特許文献】

<<非特許文献>>

【発明の概要】

【発明が解決しようとする課題】

しかしながら、従来の<発明の名称>においては、<全体課題>という課題があった。

【課題を解決するための手段】

本第#の発明の<発明の名称>は、<請求項>である。

かかる構成により、<効果>。

【発明の効果】

本発明による<発明の名称>によれば、<全体効果>が可能である。

©IRD国際特許事務所 谷川英和

### (4)製品情報

#### 製品(構成要素)の部品に関する情報

- ⇒発明を構成する**構成要素の説明文**
- ⇒**権利範囲を拡大するための説明文**
- ⇒**特許要件を満たすための説明文**

名称	説明
デコーダ	<構成要素>は、デコードされるべき各ソース単語シーケンスについて与えられる補間重みのベクトルに従って、全てのサブシステムからの全てのモデルを線形補間可能である。こうするために、検索に先立って、デコーダはまず、各サブシステムからの句(フレーズ)テーブルをマージしなければならない。全ての句テーブルの句の全てが、デコードの間に用いられる。1つのサブシステムのテーブルで発現するが他のサブシステムのテーブルでは発現しない句も用いられるが、トレーニング中にこの句を獲得しなかったサブシステムによるサポートはない(ゼロ確率)。探索処理は、典型的な多段句ベースデコーダにおけるのと同様に行われる。
翻訳仮説推定手段	<構成要素>は、デコーダからのnベスト出力をとり、ソース及びターゲット分類器に従ったクラスが一致するリストの中で最も高い翻訳仮説を選択する。
統計的機械翻訳装置	トレーニングモジュールによってトレーニングされたモデルと、トレーニングモジュールによって推定された重みWとに基づいて、統計的に翻訳を行う。
格納	<構成要素>は、不揮発性の記録媒体が好適であるが、揮発性の記録媒体でも実現可能である。
<他>	<構成要素>は、通常、MPUやメモリ等から実現され得る。<構成要素>の処理手順は、通常、ソフトウェアで実現され、当該ソフトウェアはROM等の記録媒体に記録されている©IRD国際特許事務所 谷川英和路)で実現しても良い。

## (5)技術分野情報

### 技術分野に関する情報

- ⇒技術分野に固有の**専門用語の説明文**
- ⇒**権利強化を図るための定型文**
- ⇒**文献としての役割を果たすための定型文**

名称	説明
ターゲット言語	ターゲット言語は、通常、アラビア語 (a r)、デンマーク語 (d a)、ドイツ語 (d e)、英語 (e n)、スペイン語 (e s)、フランス語 (f r)、インドネシア語 (マレー語) (i d)、イタリア語 (i t)、日本語 (j a)、韓国語 (k o)、マレーシア語 (マレー語) (m s)、オランダ語 (n l)、ポルトガル語 (p t)、ロシア語 (r u)、タイ語 (t h)、ベトナム語 (v i)、中国語 (z h) である。
受け付け	受け付けとは、タッチパネルや、キーボードなどの入力デバイスから入力された情報の受け付け、有線もしくは無線の通信回線を介して送信された情報の受信、光ディスクや磁気ディスク、半導体メモリなどの記録媒体から読み出された情報の受け付けなどを含む概念である。
出力	出力とは、ディスプレイへの表示、プロジェクターを用いた投影、プリンタでの印字、音出力、外部の装置への送信、記録媒体への蓄積、他の処理装置や他のプログラムなどへの処理結果の引渡しなどを含む概念である。なお、送信や蓄積、処理結果の引渡しについては、出力対象が最終的にユーザに提示されるものとする。

## (6)共通情報

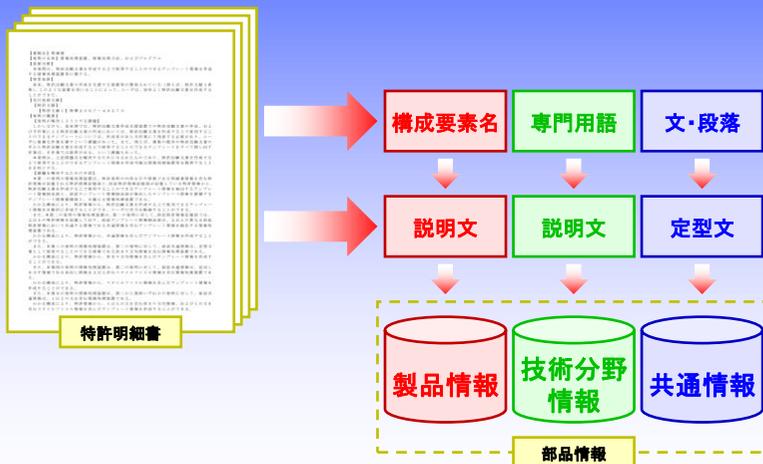
### 特許明細書における**定型文**

- ⇒**任意の技術分野において共通する文章**
- ⇒**技術分野を問わず共通する文章**
- ⇒**特許明細書における定型文**

名称	説明
システム構成1	図1はこの実施の形態のSMTシステム30の全体構造を示す。図1を参照して、SMTシステム30は、クラス依存SMTモデル、ソース文を分類するために用いられる分類器モデル、及びSMTデコーダ内で用いられる句テーブルをトレーニングするためのトレーニングモジュール44を含む。トレーニングセット42はトレーニングデータとして用いられる。トレーニングモジュール44はさらに、一般SMTモデルに割当てられる重みW1を推定する。重みは、開発セット40に基づいて推定される。パイリンガルコーパスはクラスに分けられ、さらに、各クラスについてトレーニングセットと開発セットとに細分される。
システム構成2	図2は図1のSMTトレーニングモジュール74とクラス依存SMTモデル112の3つの集合とを示す詳細なブロック図である。
実施の形態の効果	以上、本実施の形態によれば、～できる。
発明を実施するための形態フッタ	今回開示された実施の形態は単に例示であって、本発明が上記した実施の形態のみに制限されるわけではない。本発明の範囲は、発明の詳細な説明の記載を参酌した上で、特許請求の範囲の各請求項によって示され、そこに記載された文言と均等の意味および範囲内でのすべての変更を含む。

## (7)特許部品DB作成システム

特許明細書、特許公報などの特許文書から、  
特許部品DBを作成する(部品情報(ひな形情報を除く)を抽出する)。



©IRD国際特許事務所 谷川英和  
より効率的な部品情報の抽出・蓄積が可能に!!

13

## 4.読解支援システム ( Patent Structure Viewer Viewer)

## (1) 研究開発の動機

### • 「特許請求の範囲」の低可読性の問題

#### – 特許請求項の読みにくさ

- 文長が長く、構造が複雑
  - 主張する権利範囲を1文で記述
  - 読みやすさよりも、正確性・厳密性が優先
  - 専門家コミュニティが形成され、一定の記述スタイルが確立

#### – 請求項の引用関係の複雑さ

- 多数項引用、多重多数項引用
- 外的付加と内的付加

## (2) Patent Structure Viewer について

以下の機能を提供するソフトウェアです

### 1. 特許請求項の細粒度解析

- 構成要素列挙形式
  - 「～と、～と、～とを備えた〇〇」
- 順次列挙形式(書き流し形式)
  - 「～し、～し、～する〇〇」
- 変形構成要素列挙形式
  - 「～と、～と、～とを備え、～は～し、～は～する〇〇」
- ジェブソン形式
  - 「～において、～を特徴とする〇〇」

### 2. 注釈付きクレームツリーの生成

- クレームツリーの自動生成と「電気回路図」的表現での表示
- 内的付加・外的付加の判定
  - 外的付加=「現状の構成要素の組み合わせに対して別の構成要素を付け加えるもの」
    - 「請求項〇に記載の〇〇であって、～〇〇を更に備えた〇〇。」
  - 内的付加=「上位概念に内包される下位概念を案出するものであり、それには限定、選択が含まれる」
    - 「請求項〇に記載の〇〇であって、前記〇〇は〇〇である〇〇。」

### 3. 特許請求項における特徴的単語の強調表示(オプション)

- 他の特許書類にあまり出現しない単語を検出し、カラー表示



## (5) 実行例③

### 特許請求項の解析結果のCSV出力

- 粗い解析結果と細粒度解析結果の2種類のCSVを自動出力

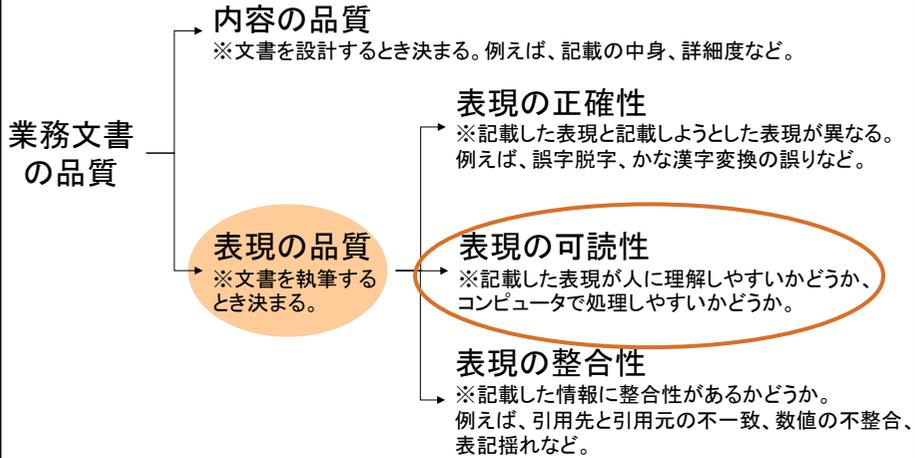
請求項	解析結果
1	請求項1
2	ソース文のクランパースを表現する確率のベクトルを決定するための手段を含み、前記ベクトルの要素はソース文が予め
3	前記ターゲット言語の可能な単語シーケンスの確率に従って、前記ソース文の前記ターゲット言語における最も尤も高い
4	統計的機械翻訳装置。
5	請求項2
6	前記確率ベクトルは一般クラスと
7	複数個の特定クラスと
8	を含み、
9	前記複数個の特定クラスは、前記一般クラスを分割したものである。
10	請求項1に記載の統計的機械翻訳装置。
11	請求項1
12	前記一般クラスに対応する前記ベクトルの1要素は、0から1の範囲で定数である。
13	請求項1又は請求項2に記載の統計的機械翻訳装置。
14	請求項1
15	前記ベクトルの要素を正規化して、前記要素の和が1となるようにするための正規化手段
16	を含み、
17	請求項1
18	請求項1
19	請求項1
20	前記確率のベクトルを決定するための手段は、最大エントロピーモデルに基づいて統計的トレーニングされ、
21	前記確率のベクトルを決定するための手段は、最大エントロピーモデルに基づいて統計的トレーニングされ、
22	前記クラスのそれぞれにエンバディメントを割り当てる。
23	請求項1
24	請求項1
25	請求項1
26	請求項1
27	請求項1

粗い解析結果

細粒度解析結果

## 5.可読性診断システム(KnowledgeMeister文書診断)

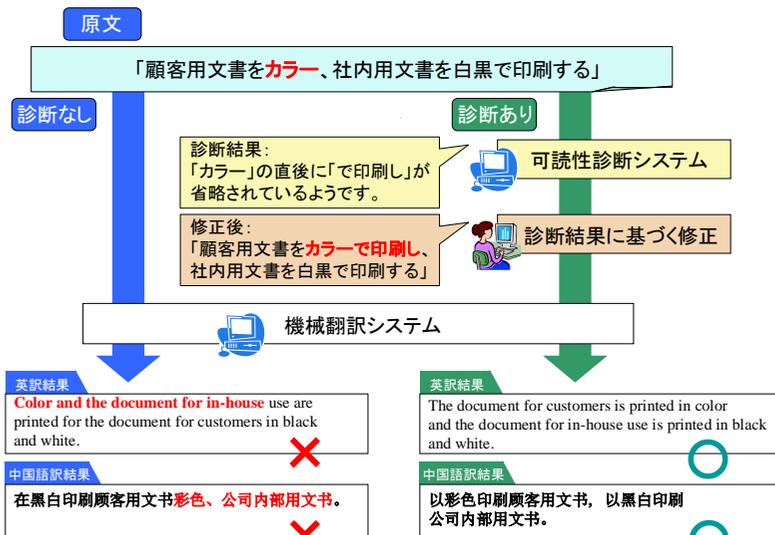
## (1)業務文書の品質



©IRD国際特許事務所 谷川英和

21

## (2)機械翻訳への活用例



©IRD国際特許事務所 谷川英和

Copyright 2012 Toshiba Solutions Corporation

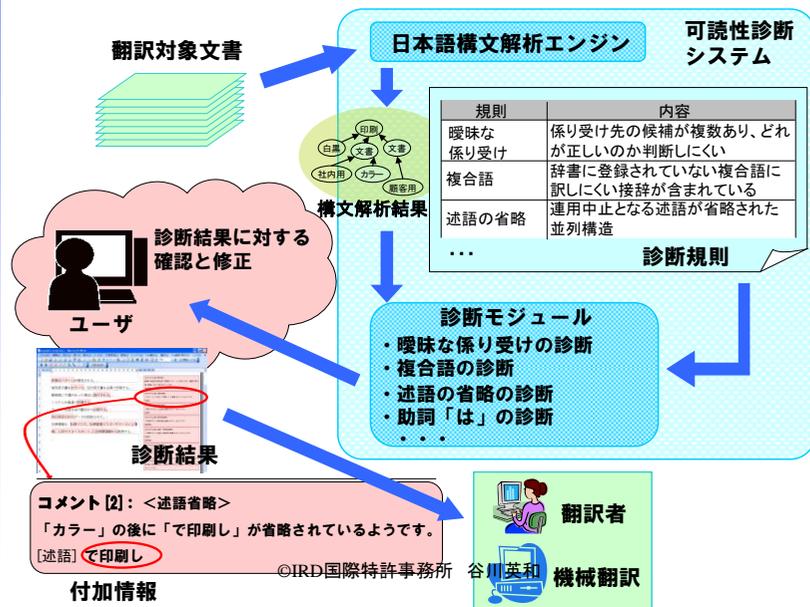
### (3)可読性診断技術の活用シーン

- **機械翻訳の前処理**
  - 曖昧性のある表現が指摘される
  - 原文を修正し、翻訳の精度を向上
- **翻訳コストの低減**
  - 分かりやすい翻訳原文が作成できる
  - 翻訳者からの問合せ、翻訳者の誤解による誤訳が減少
- **人が理解しやすい文章の作成**
  - 仕様書などの技術文書で、正しい情報伝達を支援
  - 理解度が増し、業務効率が向上

©IRD国際特許事務所 谷川英和

23

### (4)可読性診断技術の構成



## (5)技術的特徴

- 翻訳しにくい文の特徴を考慮した診断規則に基づいて、文章の可読性を診断
- 日本語構文解析の結果を使用し、文の構造から誤訳しやすい箇所や曖昧な箇所を検出
- 翻訳の精度を向上するための付加情報や選択肢を生成

コメント [2]: <述語省略>

「カラー」の後に「で印刷し」が省略されているようです。

[述語] **で印刷し**

付加情報

昨日保存されたデータが削除された

コメント [3]: <曖昧な係り受け>

「昨日」 - 「保存」

「昨日」 - 「削除」

選択肢

©IRD国際特許事務所 谷川英和

23

## (6)可読性診断結果の出力例

変換元パターンが複製される。

顧客用文書をカラー、社内用文書を白黒で印刷する。

昨日保存されたデータが削除された。

勤務表に不備があった場合に発行される。

システムが高速で計算する。

データの処理方法の観点から分類する。

社員情報を、社員マスタ、社員管理マスタ・サマータイム情報、人材マスタ・スポット、入社時間情報から取得する。

コメント [1]: <複合語>  
詳細に本登録の複合語「変換元パターン」があります。翻訳に注意の語群「元」が含まれています。  
[登録語]変換元パターン

コメント [2]: <述語省略>  
「カラー」の後に「で印刷し」が省略されているようです。[目的語]  
[述語]で印刷し

コメント [3]: <述語省略>  
「印刷する」の述語が省略されているようです。  
[述語]

コメント [4]: <構文要件>  
0. 「昨日」・「保存」 (1)  
1. 「昨日」・「削除」 (1)  
[選択]

コメント [5]: <述語省略>  
「発行する」の述語が省略されているようです。  
[述語]

コメント [6]: <目的語省略>  
「計算する」の目的語が省略されているようです。  
[目的語]

コメント [7]: <述語・目的語省略>  
「分類する」の述語と目的語が省略されているようです。  
[述語]  
[目的語]

コメント [8]: <述語省略>  
「取得する」の述語が省略されているようです。  
[述語]

©IRD国際特許事務所 谷川英和

## (その他)産業日本語との関連

### (1)特許明細書半自動生成システム(PatentGenerator)

- ・産業日本語ライティングルールに従った特許部品DBを準備すれば、ルールに従った明細書の出力が可能 (明細書の標準化が可能)

### (2)読解支援システム(Patent Structure Viewer)

- ・産業日本語ライティングルールに従った場合、解析精度を一段と向上できる可能性がある
- ・今後の開発により、産業日本語ライティングルールに従った構造を有するか否かのチェックが可能となる可能性がある

### (3)可読性診断技術(KnowledgeMeister文書診断)

- ・可読性診断の判定基準と産業日本語ライティングルールとの共通性
- ・ライティングの支援システムや支援環境としての活用

## 6.評価結果と今後

### 1.特許ライティング支援システムを構成する各ツールの評価

- 概ね良好(使用すれば効果がある)

### 2.各ツールの改善案のまとめ、改善

- PatentGenerator: 化学分野等において、関連特許を自動検索し、動的に特許部品DBを作成する等の機能が必要
- KnowledgeMeister: 特許書類に特化した用語等のチェックが必要

### 3.特許ライティング支援環境のあるべき姿の提案

- さらに「教育支援環境」が必要

ご清聴ありがとうございました

# 産業文書に関する言語処理技術の開発と期待

井佐原 均

国立大学法人豊橋技術科学大学  
情報メディア基盤センター教授

**講演概要：** 自然言語処理の研究と、企業での文書処理のニーズとは十分に情報交流が行われているとは言い難い。このような観点から、産業文書に対する自然言語処理技術の可能性を実例に沿って概説する。具体的には前編集・対訳辞書・後編集など、翻訳プロセスの中での支援を中心に、産業日本語研究会や関連の活動を通しての企業との関わりの中で始めた企業の実文書の処理に対する支援技術について述べる。

## 1. 翻訳プロセスと機械翻訳

機械翻訳を用いた翻訳過程で精度を向上するポイントとしては、前編集、機械翻訳、後編集の3つが考えられる。前編集においては、日本語を人間にとっても、機械にとっても分かりやすく記述するための規格を導入し、入力文を制約することにより、機械翻訳結果の品質向上が可能となる。機械翻訳に関しては、翻訳エンジンそのものの精度向上はもちろん大切であるが、既存のシステムの活用においては、対訳辞書の整備が翻訳精度の向上に直結する。情報発信型の翻訳が、情報受信型の翻訳と異なる点の一つは入力をコントロールできることである。入力文を規格化し、対象分野の対訳辞書を整備することにより、機械翻訳出力の精度が向上し、速報性が重視される文書や、内部での利用のための文書の翻訳には十分な精度となることが期待される。出版物やウェブ上で一定期間提示される文書などの場合には、さらに翻訳の精度を高めるための後編集作業が必要となる（図1）。

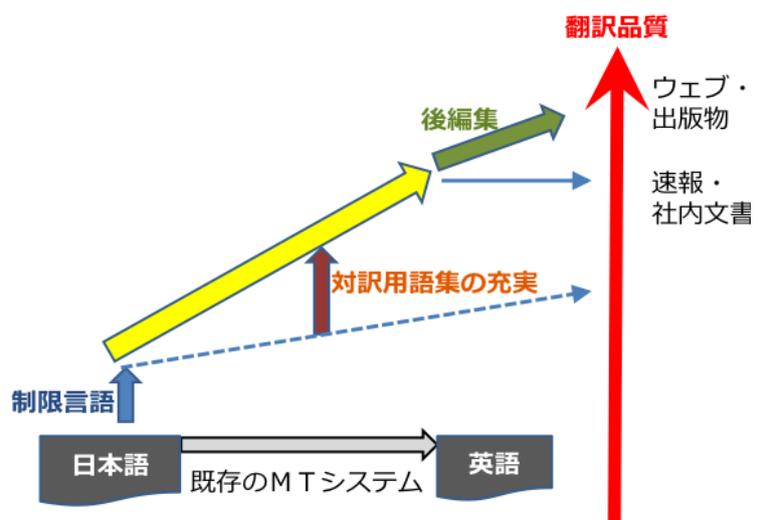


図1 翻訳プロセスの精度向上

## 2. 個別ニーズへの適合

機械翻訳をはじめとする自然言語処理技術を産業文書に適用する場合には、各企業や分野に対してシステムを適合させる必要がある。機械翻訳システムそのものを適合させる場合、用例翻訳や統計翻訳といったコーパスベースの翻訳システムであれば、学習用の対訳文書を準備することによって、機械翻訳システムの精度を向上することが可能となる。我々が平成18年度から5年間にわたり、科学技術振興調整費重点課題解決型研究等の推進「日中・中日言語処理技術の開発研究」で実施した日中機械翻訳システムの開発は、科学技術文献を対象に、このような考えに沿ったものである。個々の企業で大量のデータを集積することには困難が伴うが、企業文書のデータの共有によって大量データを利用可能とする試みとしては、T A U S (Translation Automation User Society: <http://www.translationautomation.com/>) の活動がある。

本稿では機械翻訳のアルゴリズムを改良するのではなく、既存の機械翻訳システムの利用を前提に、入力文書をコントロールし、対訳専門用語辞書を構築することによって、翻訳精度を向上し、必要な場合にはクラウドソーシング後編集によって求める翻訳精度を実現する枠組みについて述べる。このような取り組みによって機械翻訳をはじめとする自然言語処理技術の活用の幅が広がることを期待している。入力文の制約においては、規格化日本語という考え方により、企業や自治体文書を翻訳しやすい文書に変更することを試みている。対訳専門用語辞書の構築においては、マニュアル文書を対象に、そこに出現する「意味のあるひとまとまりの語句」を半自動で取り出すことを目指しているクラウドソーシング後編集においては、大学紹介の英語ホームページを対象に多言語での後編集実験を行い、クラウドソーシング後編集の有効性を検討している。

## 3. 規格化日本語・産業日本語

企業内文書や自治体のホームページを対象に、どのような規則によって原文を制約すれば、機械翻訳出力が向上するかを検討している。

企業内文書について言えば、自動車関連企業の文書を精査し、平易な基準（規格）とテンプレートをその企業の社員に提示した。基準とテンプレートを用いて、実際に社内文書を記述してもらい、それにより機械翻訳結果がどのように変わるかを調査している。この実験に先立ち、平易な基準に沿って作成されたマニュアルとその機械翻訳出力が、日本語および英語として適切な読みやすいものになっているかどうかを被験者実験で確かめた。評価はこの企業の社員を被験者に行い、「既存マニュアルの文」と「書き換えた文」を日本人社員20名に提示し、書き換え後の文が日本語のマニュアルとして、より適切な（読みやすい）文になっていることを確認した。また、「既存マニュアルの文の機械翻訳出力」と「書き換えた文の機械翻訳出力」を外国人社員8名に提示し、書き換えた文の機械翻訳出力がより良い英文になっていることを確認した。

現在、ISO TC37 SC4 において、ISO/AWI 24620-1: Language resource management

- Simplified natural language -- Part 1: Basic concepts and general principles の検討が行われており、ここでの成果が反映される予定である。

#### 4. 対訳用語辞書の整備

実際の産業文書を高精度で翻訳するためには、その文書に出現する用語の辞書を整備することが必要である。このような用語は分野や企業に特化したものであり、かつ新規の語彙が常に作成されるという特徴がある。このため人手による作成は速度と経費の点で困難であり、文書から（半）自動的に用語を取り出す技術が必要となる。翻訳を対象とする場合、取り出すべき対象は単語だけでは不十分であり、頻出する言い回しなど、意味のあるひとまとまりの句を取り出し、その全体に対して対訳（となる語句）を与える必要がある。

ここでは、英語文書が与えられた時に、翻訳に必要な対訳用語集を半自動で作成する手法を実際の文書の処理を例に紹介する。実験では自動車メーカーの整備マニュアル（英日対訳）を対象とした。原文は英語であり、これを日本の翻訳会社が日本語に翻訳した。翻訳に際して、翻訳会社は翻訳用の日英対訳用語集を作成している。本実験で用いたのは、マニュアルの英語原文（10 万文）と日英対訳用語集（100 エントリー）である。課題は英語原文から英語の用語（100 エントリー）を実用的な作業量で抽出するタスクとなる。

我々のシステムは、文書中の用語の接続情報（統計情報）を用いることにより、意味のあるひとまとまりの語句の抽出を可能としている。本研究の目標は、全自動の辞書構築ではなく翻訳前の英語テキストから自動的に抽出した用語句候補を人手で眺めて、分野特有の用語集を取り出すことである。このためには人手で調査できる程度の数の、また正解を十分に含んだ、用語句候補を取り出す必要がある。

処理の手順は以下のようなになる（図 2）。

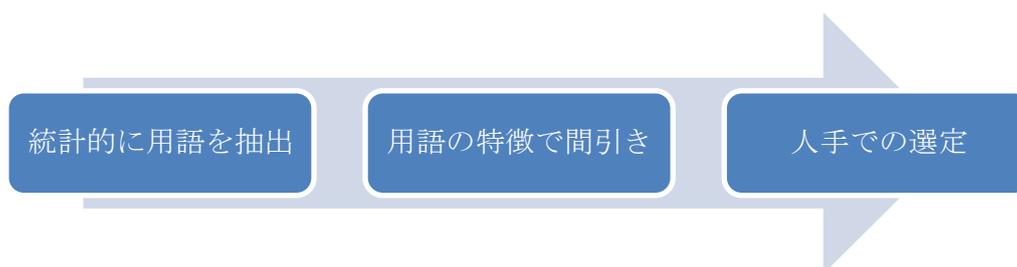


図 2 半自動用語抽出の手順

用語自動抽出システムにより、103,828 文からなる英語版マニュアルから 65,712 個の用語候補を抽出した。抽出した用語の中に抽出したい 100 語句の内の 83 語句を確認した。この 83 語句を部分形態素列として含む語句は 1,767 語句が抽出されていた。再現率で見れば、83%ということになるが、適合率は極めて低く、抽出した文字列は膨大

な量であり、ノイズを多く含んでいる。この6万強の語句を人間が見て、100個弱の語句を選び出すのは現実的ではない。このため、正解データの取りこぼしを増やさず、かつ正解でない語句を排除するような手法が必要となる。

このため、取り出した語句の品詞情報・字種情報や接続情報を利用しての間引きを行った。この結果、マニュアル原文から取り出した専門用語候補65,712個を4,947個に絞ることができた。この中で正解語句を含む文字列は175文あり、正解語句としては間引き前に83個含まれていたものが、72個維持されている。5000程度の候補から正解を目視で選ぶことは十分実用的であると思われる。間引きに用いた条件は単体では候補数を2万程度にしか減らすことができなかった。複数の条件の組み合わせにより、今回の結果を引き出している。

用語候補を5000個から、さらに絞り込むことは可能と思われるが、今回の実験では次のような問題もある。今回は正解データが人手で作成した用語集であり、機械処理では制約しがたいものもあった。また、5000個を目視する中で新しい用語を確認することもあろう。なお、我々のシステムは本来、単語や複合語だけではなく、より長い語句も取り出し可能であり、入力支援などにも有効な文字列を取り出すことを目的としており、一般的な用語（名詞・複合名詞）抽出システムとは異なる。ウェブ上で公開されている用語抽出システムでも同じデータを用いた実験を行ったが、同じ72個の用語を得るためには少なくとも上位8,763個を見なくてはならなかった。

## 5. 集合知（クラウドソーシング）後編集

機械翻訳システムの性能は徐々に向上しており、翻訳支援や情報獲得支援には利用可能な状況になっている。しかし、未だに精度は完璧ではなく、前処理や後編集などが必要とされている。しかし、絶えず更新される情報をプロの翻訳者に依頼し、後編集するには膨大なコストが必要になる為、誰もが利用できるわけではない。

本節では機械翻訳システムの出力を、母語話者ではあるが翻訳のプロではない人が複数人で後編集することを提案する。この集合知により、プロの翻訳者の後編集によって得られるような理解可能な翻訳文を低コストで得ることが可能であることを示すことが目的である。

各文を複数名で後編集する場合、二人目以降は、原文と、機械翻訳システムによる翻訳出力と、それまでの後編集結果を参考にして、更に良い文を作ることができる。他の人の後編集結果を参考にする事ができるので、翻訳技術の乏しい人でも参加する事ができる。また、修正に自信がある文だけを後編集することができる。さらに、後編集の対象となる文章に対して、ある程度の知識を持っている人が参加することでプロが行う後編集より適切な文が得られると考えられる（図3）。

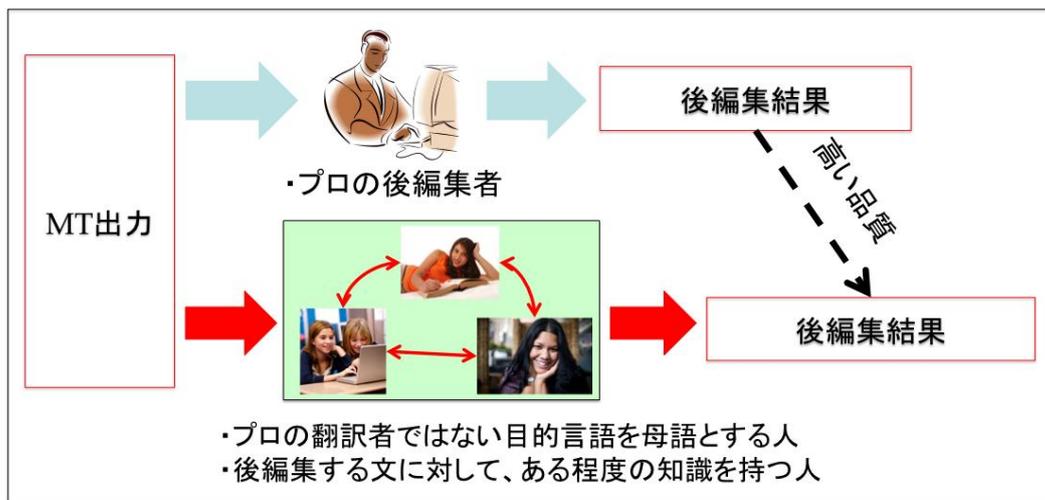


図3 プロによる後編集と集合知（クラウドソーシング）後編集

実験は豊橋技術科学大学の英語版ホームページに機械翻訳システムを設置し、多言語に翻訳することで行った。この翻訳結果に対し、大学内の留学生が自分の母語に訳された結果を後編集した。後編集を行った言語は、アラビア語・インドネシア語・スペイン語・ドイツ語・フランス語・ベトナム語・ポルトガル語・中国語・韓国語の9か国語である。また、詳細な統制実験として、日本人学生4名による日本語の後編集実験も行った。後編集終了後に、プロの翻訳者によって、どの後編集が適切かを判定した。適切な後編集結果が無い場合は、プロの翻訳者が後編集を行った。

実験の結果（表1）と得られた知見を以下に示す。ここでは後編集量の多かった4か国語（中国語・インドネシア語・スペイン語・ベトナム語）について説明する。

表1 後編集量と選択結果

	中国語	インドネシア語	スペイン語	ベトナム語
センテンス数(文)	1,638	1,287	1,849	1,344
有効センテンス数(文)	1,627	1,286	1,848	1,343
LCPE が選択された数(文)	1,088	875	894	783
LCPE 以外の CPE が選択された数(文)	171	37	329	116
MT が選択された数(文)	250	127	291	134
プロの翻訳者によって修正された数(文)	118	247	334	310
LCPE が選択された(%)	67%	68%	48%	58%
LCPE 以外の CPE が選択された(%)	11%	3%	18%	9%
MT が選択された(%)	15%	10%	16%	10%
プロの翻訳者によって修正された(%)	7%	19%	18%	23%

LCPE: Last Crowd Post Edit（最後に行われた後編集の結果）

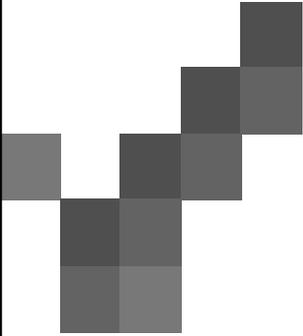
たとえば、中国語についてみれば、クラウドソーシング後編集を行わない場合は、MT 出力を正しいとした 15%以外の文を修正しなくてはならないが、クラウドソーシングの結果、67%が修正不要となり、プロの翻訳者による後編集は 18%で済む。（もし、他の後編集結果も参照するならば、修正は 7%で済む。）また、18%の修正部分についても、MT 出力を修正する場合の TER が 67.822 であるのに対し、後編集結果を修正する場合の TER が 35.829 であり（表 2）、修正の作業量ははるかに少なくなる。

表 2 TER (Translation Err Rate) による評価

言語	MT 出力とプロによる 修正文間	LCPE とプロによる 修正文間
インドネシア語	38.160	20.238
スペイン語	35.634	25.046
ベトナム語	47.287	18.620
中国語	67.822	35.829

## 6. おわりに

産業日本語の一分野として、機械翻訳を対象に実社会のデータを取り扱うための適合（チューニング）技術について述べた。技術は実際のユーザの下で使われてこそ意味を持つものである。本稿と講演により、ご興味をお持ちになられた方は是非お問い合わせください。



# 『日本語スタイルガイド』 の活用とTC技術検定

産業日本語研究会・シンポジウム2013(2013.3.1)

高橋 尚子  
一般財団法人テクニカルコミュニケーター協会専務理事  
國學院大學経済学部教授

JTCA 1



## 本日の流れ

- テクニカルコミュニケーター協会とは
- 『日本語スタイルガイド』
  - 出版までの経緯
  - 目次内容
  - ライティング部分の紹介
- 『日本語スタイルガイド』の活用事例
  - TC技術検定3級[TW]ガイドブック
  - 「日本語スタイルガイド」出前セミナー
  - 高等教育におけるTC専門課程制度
- 最後に・・・新しい動き

JTCA 2

## (1) 一般財団法人テクニカルコミュニケーター協会とは

- 略称: TC協会
  - 英文名称: Japan Technical Communicators Association (略JTCA)
  - 創設: 1992年1月・・・昨年20周年を迎える
  - 製品やサービスの使用説明を扱う専門家の団体
  - 目指すもの
    - 「使用説明」の品質向上によって、誰もが安全かつ簡単に最新の技術を利用することができ、仕事や生活の質を高めることができる社会の実現
    - TC技術の発展と普及を促し、その応用範囲の拡大
    - TC技術に携わる人々の間で情報共有と情報交換
    - デザイン、ユーザビリティ設計、システム設計、国際マーケティングなど関連分野との交流
  - 会員数: 法人会員約100社、個人会員約200名
- URL → <http://www.jtca.org/index.html>

JTCA

3

## TC協会の沿革

1989.6	第1回TCシンポジウム「いま、日本のマニュアルを考える」開催。以降、毎年開催
1992.1	<b>任意団体「テクニカルコミュニケーター協会」として設立</b>
1992.6	「マニュアル評価ガイドライン」発行。以降、毎年さまざまなテーマで調査報告書発行
1997.3	「テクニカルコミュニケーション 技術検定<分野:テクニカルライティング>3級ガイドブック」発行★
1998.2	TC 技術検定試験<分野:テクニカルライティング3級>を実施(2003年以降の呼称は初級)。以降、2009年まで毎年1回実施
1999.12	TCシンポジウムを大阪で開催。以降、毎年関西地区でも開催
2000.3	「テクニカルコミュニケーション 技術検定<分野:テクニカルライティング>2級ガイドブック」発行★
2000.4	日本マニュアルコンテスト単独開催開始。以降、毎年実施
2001.2	TC 技術検定試験<分野:テクニカルライティング2級>を実施(2003年以降の呼称は上級)。以降、2009年まで毎年1回実施
2009.1	<b>一般財団法人テクニカルコミュニケーター協会として活動</b>
2009.7	『日本語スタイルガイド』(初版)発行★
2010.2	新しい枠組みでのTC技術検定試験3級テクニカルライティング[TW]を実施。以降、毎年2回7月・2月実施

## TC協会の活動内容

- 公益事業
  - TCシンポジウム
    - ・・・協会発足の前から毎年実施、今年25周年を迎える
  - 日本マニュアルコンテスト
  - 標準規格の策定
  - 学術研究および産学協同プロジェクトの推進
  - 海外市場向け製品の表示情報に関する調査など
- 収益事業
  - 人材育成(セミナー、高等教育など)
  - TC技術検定の実施(国内、国際とも)
  - 出版(関連書籍、会誌など)
  - 受託事業(マニュアル評価など)

## 「テクニカルコミュニケーション」とは

- 製品やサービスに関して、利用者が求める情報を、正確にわかりやすく表現すると共に、効果的に伝達し、提供すること。略して「TC」という。この情報の提供を実現するためのさまざまな技術を「TC技術」という。
- TC技術の特徴
  - 印刷媒体から画面に表示される電子媒体まで、さまざまな媒体で提供する
  - 文章、図解、イラスト、写真、音声、動画など、さまざまな要素を適切に配置し、構成する
  - 読み手の属性・状況・要求を的確に想定し、製品やサービスとのよい関係をつくる
  - 適用分野は、使用説明(取扱説明書、カタログなど)、技術文書、業務マニュアル、ビジネス文書、Web掲載情報、論文まで多種多様である(ただし、法令文や広告文などは適用外)

## 「テクニカルライティング」とは

- 技術的な情報を、正確かつわかりやすく効率的に伝えるための文章作成技術
  - その誕生・・・
    - 工業化が進展し、新しい科学技術を応用した製品を安全に間違いなく取り扱えるよう、使用方法や操作手順などを伝える必要  
例えば...航空機のメンテナンスマニュアルなど
  - その発展・・・
    - 1970年以降、業務でのコンピュータ利用が急増し、情報システムという見えない仕組みを理解し、使えるようになるために説明が必要に
    - 1980年代後半以降、専門知識を持たない個人のPC利用が広まり、複雑な機能を理解し、操作できるように、適切な説明技術が重要に
  - その後・・・
    - 各企業や任意団体での技術の追求、わかりやすいマニュアル作成への取り組みが始まる
- ※ 専門的で難解な科学技術論文をわかりやすく書く「サイエンスライティング」とは一線を画す

## (2)『日本語スタイルガイド』

- 2009年7月 初版発行  
2011年4月 第2版発行(改訂常用漢字に対応)
- 価格: 2,520円(税込)
- 総発行部数: 4300部(2012年末現在)
- 執筆者:  
TC技術検定部会委員を含めTC協会関係者(雨宮 拓・勝田豊彦・黒田 聡・高橋慈子・高橋尚子・徳田直樹・富永敦子・永山嘉昭)
- 本書の概要:
  - 長年使用説明の制作に携わった経験を土台として、わかりやすく効率的に情報を伝えるための日本語作文技術の要点をまとめたもの
  - テクニカルライティングの知識や経験に基づき、より広範な日本語の文章表現技術に応用できるよう、基本的な指針を集大成したもの
- 利用方法:
  - 常にデスクに置いて関連ページを参照する
  - TC技術検定3級[TW]受験のガイドブックとする

## 発行までの変遷

1992.6	「マニュアル評価ガイドライン」調査報告
1994.5	「人材育成のための教育カリキュラム」調査報告
1995.5	「TC資格認定制度に関する調査研究報告書」発行
1996.3	「TC技術検定制度に関する調査報告書」発行
1997.3	「テクニカルコミュニケーション 技術検定<分野:テクニカルライティング> 3級ガイドブック」発行①
2000.3	「テクニカルコミュニケーション 技術検定<分野:テクニカルライティング> 2級ガイドブック」発行②
2003.1	「テクニカルコミュニケーション技術検定ーテクニカルライティング試験ガイドブック」発行③
2004.1	「マニュアルライティングーテクニカルコミュニケーション技術検定・テクニカルライティング分野受験対策テキスト」発行
2009.7	『日本語スタイルガイド』(初版)発行④
2011.4	「日本語スタイルガイド」(第2版)発行

JTCA

9

## ①「テクニカルコミュニケーション 技術検定 <分野:テクニカルライティング>3級ガイドブック」

- 1997年3月 初版発行
- 執筆者: TC技術検定専門委員会から選ばれた委員
- 本書の概要
  - 1994年および1995年の調査報告書に基づき、TC技術試験の技術要素を解説した3級の受験用ガイドブックとして発行
  - 章構成は、マニュアルの制作工程に沿っている。「ライティング技法」の2章は、最もページ数を割いて記述

### <目次>

- 1章 企画・構成
- 2章 **ライティング技法**
  - 2.1 日本語文法
  - 2.2 用字・用語
  - 2.3 文章表現の基本
  - 2.4 文法と文章表現
  - 2.5 読みやすい文章表現
  - 2.6 文章の組み立て方
- 3章 ビジュアル表現
- 4章 制作に関する技術
- 5章 付帯する技術
- 6章 評価・管理
- 付録 検定する技術の区分

JTCA

10

## ②「テクニカルコミュニケーション 技術検定 ＜分野：テクニカルライティング＞2級ガイドブック」

- 2000年3月 初版発行
- 執筆者：TC技術検定専門委員会から選ばれた委員11名
- 本書の概要
  - 3級ガイドブックと1997年に発行した調査報告書に基づき、2級としての内容を深め、加筆し、TC技術試験の技術要素を解説した2級の受験用ガイドブックとして発行
  - 章構成などは3級と同じ

### ＜目次＞

- 1章 企画・構成
- 2章 **ライティング技法**
  - 2.1 日本語文法
  - 2.2 用字・用語
  - 2.3 文章表現の基本
  - 2.4 文法と文章表現
  - 2.5 読みやすい文章表現
  - 2.6 文章の組み立て方
- 3章 ビジュアル表現
- 4章 制作に関する技術
- 5章 付帯する技術
- 6章 評価・管理
- 付録 検定する技術の区分

## ③「テクニカルコミュニケーション技術検定－ テクニカルライティング試験ガイドブック」

- 2003年1月 初版発行  
2005年10月 初版2刷
- 価格：3,800円(税込)
- 執筆者：TC技術検定専門委員会を含めTC協会関係者
- 本書の概要
  - 2級ガイドブックをベースに、新たな制作工程(下線部)や媒体を追加し、2級・3級どちらでも使用できる受験ガイドブックとして発行  
(\*\*の一部を3級の技術要素から除外)

### ＜目次＞

- はじめに TC技術検定試験概要
- 第1章 企画・構成・**表現設計**
- 第2章 **ライティング技術**
  - 2.1 日本語文法
  - 2.2 用字・用語
  - 2.3 文章表現の基本
  - 2.4 文法と文章表現
  - 2.5 読みやすい文章表現
  - 2.6 文章の組み立て方
- 第3章 ビジュアル表現
- 第4章 制作に関する技術\*\*
- 第5章 **電子マニュアル\*\***
- 第6章 付帯する技術
- 第7章 評価・管理\*\*
- 付録 TC技術検定試験例題

## ④「日本語スタイルガイド」目次

### 第1編 文書作成における説明技術

- 第1章 さまざまな実用文
- 第2章 日本語スタイルガイドとテクニカルライティング技術
- 第3章 文書として仕上げるテクニカルコミュニケーション技術

### 第2編 日本語スタイルガイド

- 第1章 文法、用字・用語、表記を理解する
- 第2章 読みやすく書く
- 第3章 誤解されないように書く

### 第3編 テクニカルライティング技術の要点

- 第1章 情報理解に基づきコンテキストを組み立てる
- 第2章 読み手のことを考えたライティング
- 第3章 文章の完成度を高め安定させる

ライティング技術

### 第4編 テクニカルライティング技術の基礎

- 第1章 表現設計の基本を知る
- 第2章 構造化に配慮して文書を設計する
- 第3章 ツールを活用して作業効率を高める
- 第4章 コンプライアンスに配慮する

### 付録

漢字とひらがなの使い分け、外来語(カタカナ)表記ガイドライン、改定常用漢字表  
TC技術検定試験の概要、TC技術検定3級試験例題

## ライティングに関わる部分の紹介(第2編)

### 第2編 日本語スタイルガイド

#### 第1章 文法、用字・用語、表記を理解する

- 1.1 日本語文法
- 1.2 用字・用語
- 1.3 表記

#### 第2章 読みやすく書く

- 2.1 文体と時制に留意する
- 2.2 一文一義で書く
- 2.3 文末の書き方を使い分ける
- 2.4 1文の長さに気をつける
- 2.5 主語と述語を正しく使う
- 2.6 修飾語を正しく使う
- 2.7 助詞を正しく使う
- 2.8 句読点を正しく打つ

#### 第3章 誤解されないように書く

- 3.1 否定表現に気をつける
- 3.2 可能表現に気をつける
- 3.3 比喩に気をつける
- 3.4 強調に気をつける
- 3.5 範囲、起点の明確な文を書く
- 3.6 数値を使って表現する
- 3.7 注意が必要な表現

※ 1つの文について、文法に従い、正確に伝えるための書き方、使い分けを解説

## ライティング部分の詳細(例)

- 助詞の使い方を具体的な例で示し、解説を行っている

(例)2.7.3 「に」と「へ」を区別して使う

表 2.21

例	解説
× カーソルを次の項目へ移動します。 ○ カーソルを次の項目に移動します。	上の文も間違いではないが、「項目」という到達点については「に」を使うように統一するとよい

(例)2.7.4 「より」と「から」を区別して使う

表 2.22

例	解説
× 4月より9月まで ○ 4月から9月まで	「4月」が基点なので、「より」を使わずに「から」を使う。
○ 4月より9月のほうが、問い合わせ件数が多い。	「4月」と「9月」を比較しているので、「より」を使う。

## ライティングに関わる部分の紹介(第3編)

- 2編「日本語スタイルガイド」に沿って、書くことだけでは実現できない技術をまとめたもの
    - 情報を理解し、整理する
    - 伝える目的を明確にし、相手(読み手)の知識や意識に配慮する
    - コンテキストを組み立て、全体の構想を練る
    - 文章を推敲し、書き直して品質を高め、完成度を高める
- ※一つ一つの文が正確で読みやすくても、伝わらないことがある。読む相手を考えて書くという態度を解説

### 第3編 テクニカルライティング技術の要点

#### 第1章 情報理解に基づきコンテキストを組み立てる

- 1.1 対象の理解と情報の整理
- 1.2 伝える「目的」と伝える「相手(読み手)」を明確にする

#### 第2章 読み手のことを考えたライティング

- 2.1 読み手の「知識」に配慮する
- 2.2 読み手の「意識」に配慮する
- 2.3 読み手の「心情」に配慮する
- 2.4 「認知や行動の仕方」に配慮する

#### 第3章 文書の完成度を高め安定させる

- 3.1 文章の推敲とリライト
- 3.2 情報の更新に伴うリライト

## テクニカルライティング技術の「土台」

- TC協会での調査報告書の「土台」となったもの
  - 「わかりやすいマニュアルを作る 文章・用字用語ハンドブック」  
テクニカルコミュニケーション研究会編  
(日経BP出版センター、1995年初版)
  - 「説得できる文章・表現の200の鉄則」  
永山嘉昭編著(日経BP出版センター、1992年初版)
  - 「ザ・テクニカルライティング」高橋昭男著(共立出版、1993年初版)
- 周辺技術として根拠としてきたもの
  - 「誰のためのデザイン? 認知科学者のデザイン言論」  
D.A.ノーマン著、野島久雄訳(新曜社、1990年)
  - 「一目でわかる表現の心理技法」  
海保博之著(共立出版、1992年)
- さらにさかのぼり、すべての原典は・・・
  - 「理科系の作文技術」木下是雄著(中央公論社、1981年初版)

## (3)『日本語スタイルガイド』の活用事例

- ① TC技術検定3級[TW]試験のガイドブック
- ② 「出前セミナー」のテキスト
- ③ 高等教育におけるTC専門課程のテキスト

## ① TC技術検定3級[TW]試験での活用

- TC技術検定試験とは、説明技術を高め、専門家を育成し、一定レベルに到達したかどうかを評価する制度
- 目的
  - テクニカルコミュニケーション技術の到達度を、全国規模の統一的な基準で検定し、技術の向上や人材の育成に活用する。
  - テクニカルコミュニケーション技術やテクニカルコミュニケーションの職域についての社会的認知度を高め、優秀な人材の確保や発掘に役立てる。
  - テクニカルコミュニケーション技術の開発、人材育成に必要な教育システム、評価システム、および教材の開発を促進させる。
- 検定の種類
  - 2級使用説明制作実務[MP]試験
  - 2級 使用説明制作ディレクション[DR]試験
  - 3級 テクニカルライティング[TW]試験
- 実施時期
  - 年2回(2月:2級[MP]・3級[TW]、7月:2級[DR]・3級[TW])

## TC技術検定実施の変遷

1998.2	「テクニカルコミュニケーション技術検定<分野:テクニカルライティング>3級試験」開始、2003年2月まで毎年実施
2001.2	「テクニカルコミュニケーション技術検定<分野:テクニカルライティング>2級試験」を追加、2003年2月まで毎年実施
2004.2	実施中の2試験の枠組みを見直し、「テクニカルライティング分野初級」試験、「テクニカルライティング分野上級」試験と改称、2009年2月まで毎年実施 「マニュアル制作ディレクション分野」試験を追加、毎年実施
2007.2	「マニュアル制作ディレクション分野」試験の技術内容を見直し、2009年2月まで実施
2010.2	『日本語スタイルガイド』に基づき試験内容と方式を見直し、「TC技術検定3級テクニカルライティング[TW]試験」開始、2010年から毎年2月・7月実施中
2010.7	「TC技術検定2級使用説明制作ディレクション[DR]試験」と改称、2012年から毎年7月に実施中
2010.10	「テクニカルライティング分野上級」試験を土台とし、印刷媒体を対象にライティングおよび制作ツールの活用、企画・構成や表現設計、ビジュアル表現など制作実務に根ざしたものに改訂し、「TC技術検定2級使用説明制作実務[MP]試験」開始、2011年から毎年2月に実施中

## 3級[TW]試験の検定内容

### ■ 技術水準

- 日本語の表記、文法、用字・用語を含む日本語スタイルガイドやライティング技術
- ライティング技術を深めるために必要なテクニカルライティング技術の要点(情報理解と文脈の組み立て、推敲とリライト、読み手を考えたライティング技術)、
- テクニカルコミュニケーション技術の基礎(表現設計、文書の構造化、ライティングツール、コンプライアンスなど)

### ■ 到達目標

- ビジネスや生活で情報を伝えるとき、論文や業務文書をはじめ、説明書や技術文書を書くとき、読みやすく、わかりやすく、正確な文章を書くことができる。

### ■ 受験対象

- 日本語の文章表現技術を高めたい方
- テクニカルライティングを学習した方であれば、マニュアルや使用説明などの制作実務経験を持たなくても受験可能

## 3級[TW]試験の検定内容(技術要素)

- 本書の編・章・節・項を順に、大区分・中区分・小区分・細目と当てはめて検定内容を設定
- 第2編⇒大区分2「日本語スタイルガイド」の技術要素

### 中区分2：読みやすいライティング

小区分	細目
1 文体と時制	(1) 文体の違いを理解し、本文を「ですます調」で書ける
:	
2 一文一義	(1) 1つの文に1つの主題で書ける (2) 1つの行動を1つの文で書ける

### 中区分3：誤解されないライティング

小区分	細目
1 否定表現	(1) 肯定文で書ける
:	
5 範囲、起点の表現	(1) 「以」の意味を理解し、「以下」「ら」「はじめ」「ほか」は、「～を含む」と表現できる (2) 「未満」「～を超えて」の意味を理解し、「以下」「以上」に置き換えられる

### 3級[TW]試験の検定内容(例題)

#### ■ 区分2.2.2「読みやすいライティング」の例題

【問】一文一義になっていない文は、次のうちどれか。

- ア アジャスターが本体に収納されていない状態で動かすことは禁止されています。
- イ キャリングケースなどに収納する際は、16 ページに記載されている「電源を切る」の手順に従ってください。
- ウ パスワードの入力に3 回続けて失敗すると、それ以降の操作ができなくなります。
- エ 表示言語は15 言語の中から選ぶことができますが、お買い上げ時は英語に設定されています。

#### ■ 区分2.3.5「誤解されないライティング」の例題

【問】参加者の人数が最も明確に表現できているものは、次のうちどれか。

- ア 中村様ら、6 名が出席します。
- イ 中村様をはじめ、6 名が出席します。
- ウ 中村様を含め、6 名が出席します。
- エ 中村様以下、6 名が出席します。

## ② 出前セミナーでの活用

- 『日本語スタイルガイド』の主要な内容に沿った、マニュアルなど使用説明のライティングに焦点を当てたセミナー

#### ■ セミナープログラム概要

1. 制作環境の変化と現在の使用説明
2. テクニカルライティング技術とテクニカルコミュニケーション技術
3. 日本語スタイルガイドに沿ったライティング技術
4. 適切なコンテキストの組み立て
5. 構造化に配慮した文書設計

#### ■ 講義形式・所要時間

- 講義＋実技演習(ワークショップ)・3.5時間

#### ■ 受講対象者

- マニュアルなど、使用説明の現場担当者および管理者  
関連部門でライティング技術に関心のある方

#### ■ 開催条件

- 企業単位・事業所単位、1開催あたり25名まで

### ③ 高等教育におけるTC専門課程での活用

- **テクニカルコミュニケーター専門課程制度**(略称:TC専門課程)・・・大学など高等教育機関でTC協会が定めた「知識と能力のリスト」(右表)にあげる分野ごとの内容を修得できる科目を履修し、必要単位を修得した学生を認定する制度(大学から設置申請が必要)

- \*「TC基礎」のテキストとして使用
- \*\*TC技術検定3級[TW]試験合格で代替可、テキストとして使用
- 2012年度から、筑波大学情報学群知識情報・図書館学類で開始

分野		必要 単位数
TC基礎*		2
制作実習		
情報収集と分析		2
企画・設計		2
情報アーキテクチャー		2
制作管理・ディレクション		2
デザイン・表現設計		2
日本語ライティング**		2
英文ライティング		2
周 辺 分 野	ユーザーインターフェイス	2
	コンプライアンス	
	異文化理解	
合計		18

JTCA

25

### ●最後に・・・新しい動き

- 文部科学省 新指導要領において国語・言語教育の見直し
  - 初等・中等教育すべての教科において、「言語学習を充実」させる
  - 国語において、「題材を読む」中心の教育から、「日常生活に関わることも含め、構成を考えて、的確に話す・聞く・書く態度を目標にする
  - URL → [http://www.mext.go.jp/a\\_menu/shotou/new-cs/youryou/index.htm](http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/index.htm)
  - TC協会でも調査を開始
- ワールドワイドでTC教育・資格検定の開始
  - 国際TC検定試験、およびTC TrainNet(独tcworld主催)
  - TC協会は、アジア地区の包括的認定パートナー
  - 日本では、本年6月から募集開始
- 学術研究の開始
  - 「テクニカルコミュニケーション学術研究会」が2012年3月に発足
  - URL → <http://www.as-tc.org/#top>

JTCA

26

第4回産業日本語研究会・シンポジウム

# システム開発文書の品質 とは何か

システム開発文書品質研究会 (ASDoQ) 事務局長  
長野工業高等専門学校 助教  
藤田 悠



Association of System Documentation Quality

## 目次

1. システム開発と文書の関係
2. 「システム開発文書品質研究会」の紹介
3. ASDoQにおける取り組みの紹介



Association of System Documentation Quality

## システム開発と文書の関係

## システム開発と文書

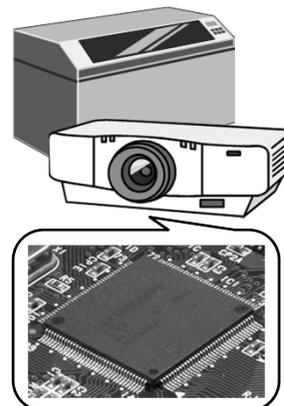
要求定義書・設計書

プログラム

システム

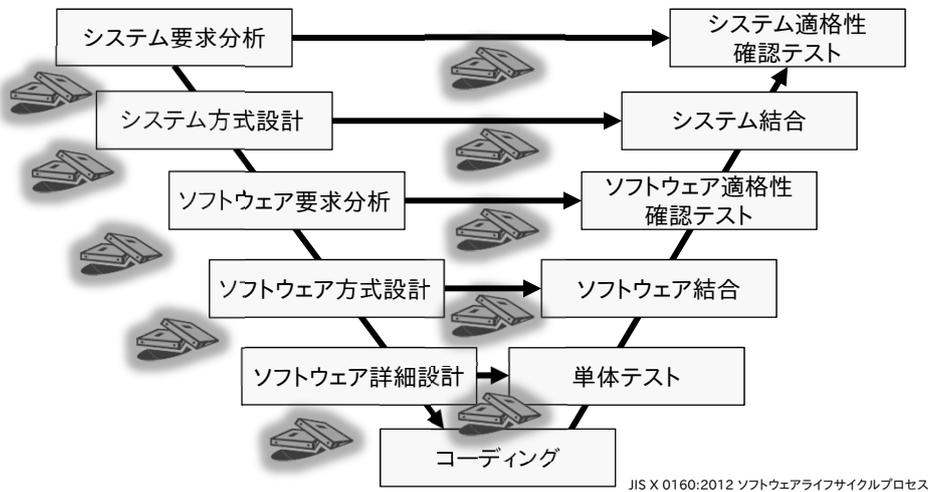
1.機能  
1.1 温度制御  
ランプ温度が120度以上になったとき、ファンを回転させ温度上昇をおさえる。140度以上になったときは、ブザーを発生して警告する。警告から10分間で温度低下がみられない場合、スタンバイ状態に移行する。  
1.2 動作表示  
LCDに動作を表示する。  
動作毎に表示する文字列を以下の表に定義する。


```
int main()
{
  while(1) {
    PORT_A = 0x01;
    PORT_B = 0x02;
    if(REG_A == 1)
      PORT_C = 0x01;
    else
      PORT_C = 0x00;
  }
  return 0;
}
```



システム開発の過程では、プログラム言語や自然言語などで文書を作る

# システム開発で作る文書



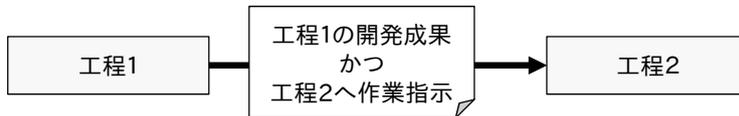
複数の工程をつなぐために、文書を介して、開発を進める

# システム開発における文書の役割

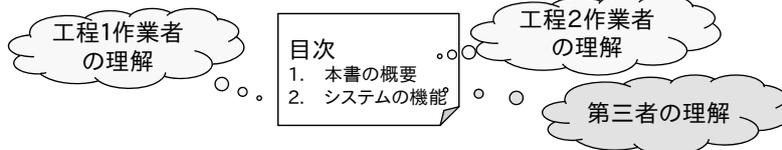
- 各開発工程ですべきことを定義する



- 開発成果を表し、次工程に指示する

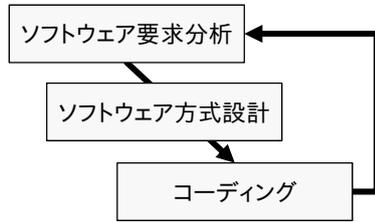


- 文章によって正確に伝達する

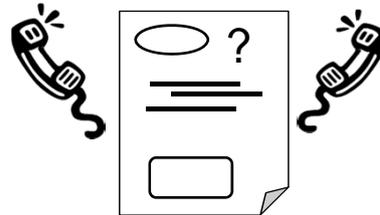


# 文書が開発に与える影響

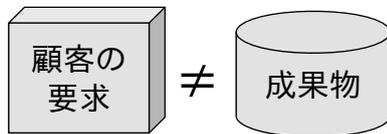
## 開発工程間の手戻り



## 問い合わせコストが増加



## 要求と成果物の不一致



## 保守や修正のコスト増加

```
int main()
{
  while(1) {
    PORT_A = 0x01;
    PORT_B = 0x02;
    if(REG_A == 1)
      PORT_C = 0x01;
    else
      PORT_C = 0x00;
  }
  return 0;
}
```

# 「システム開発文書品質研究会」の紹介

## システム開発文書品質研究会 (ASDoQ)<sup>アスドック</sup>

種別 任意団体

設立 2011年7月11日

会員 個人会員(83名), 法人会員(12社),  
法人登録会員(25名) (2013.2.11現在)

会費 原則無料

定期研究会: 年に3~4回  
技術発表, 作業部会の報告

活動 ASDoQ大会: 年に1度開催  
講演, チュートリアル

ワークショップ: 特定のテーマを合宿で集中討議

作業部会: 現在, 3テーマを設置

## 研究会の目的

1. 文書品質の提案
2. 計測技術の研究
3. 文書品質の普及



(参考) 設立当初に考えていた素朴な疑問

- ・開発文書の品質を明らかにし, それを高めることを目指せば, システム開発は今以上に, 開発生産性と開発品質を高められるのではないか?
- ・ソースコードは, 様々なメトリクスが計測され, その品質が議論されているのに, なぜ, 仕様書や設計書などは, 手がつけられていないのか?

## 研究活動とその活用

- 研究成果(非競争領域)
  - 開発文書の品質の定義
  - 開発文書品質の計測方法
  - 開発文書品質の向上方法
  - 品質の高い開発文書例
- 研究成果の活用(競争領域)
  - 文書品質の計測・向上
  - プロセス品質検証の透明性向上
  - 技術者教育カリキュラムの開発・実施
  - アウトソーシング時の提供文書の品質向上
  - 文書品質計測プログラムの開発・改良
  - 文書品質改善ビジネスの発展

研究会への参加者は、  
「システム開発文書品質研究会  
著作物取扱規則」に合意し、  
成果の自由な利用を保証する

品質の定義を共有し、各活動領域で応用、発展させる

## 会員の活動領域

- 組込みシステム・情報システムの開発
- テクニカル・ライティングの指導
- ソフトウェア・ドキュメンテーションの指導
- 自然言語処理の研究
- 大学, 高等専門学校の教員
- 技術書籍の編集
- コミュニケーションの研究
- 形式手法, XDDPなどの普及・推進

など多種多様

様々な活動領域に発展させ、様々な側面から開発文書の品質を追求する

## ASDoQにおける取り組みの紹介

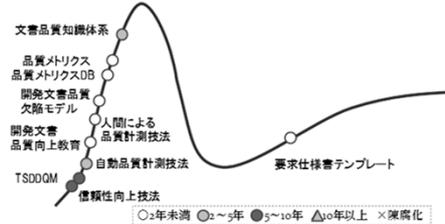
## 各作業部会での活動

- ロードマップ部会 (主査:名古屋大学 山本修一郎)  
開発文書に関連する研究状況を整理, 分析する.  
システム開発文書技術白書を作成.
- 用語定義部会 (主査:イオタクラフト 塩谷敦子)  
「システム開発文書品質」に関連する用語を説明する.  
先行研究での文書品質の性質を整理する.
- 人材育成部会 (主査:名古屋大学 山本雅基)  
文書を作成できる技術者を育成する.  
文書の品質を高めるために必要な教育の検討.

## ロードマップ部会

### 目的

システム開発文書品質に関する技術知識を整理し、文書品質研究のロードマップを作る。



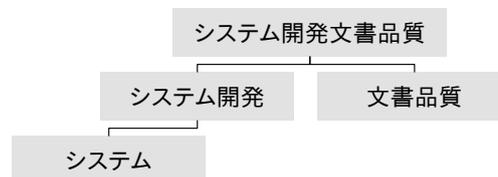
### 活動内容

システム開発文書品質に関する技術知識体系をまとめる。(研究ロードマップ「White Paper」)  
文書品質に対するニーズとシーズのギャップを明確にする。

## 用語定義部会

### 目的

ASDoQにおいて、文書の品質の研究活動を進める上で、共通の概念として統一しておくべき用語を定義する。



### 活動内容

「システム開発文書品質」を構成する用語や、文書の品質を研究するうえで必要になる用語の定義や扱い方などをまとめる「用語集」を作成中。

## 人材育成部会

### 目的

文書を作成する能力や評定する能力をもつ技術者  
およびこれらの能力をもつ技術者を育成する管理者  
のための教材を作る。

ムーブレボ通報システムの  
組込みソフトウェアの  
要求仕様書を作成する

ムーブレボ



### 活動内容

品質が高いと認める開発文書をメンバー自身が  
作成する. 文書作成の過程で得られた経験を共有する.

## 「システム開発文書の品質特性」への取組み

～巨人の肩に立つ～

- 技術文書を作成する技術を学ぶ  
システム開発文書の属性をまとめる
- 研究・教育・実務での事例を学ぶ  
優れた取り組みを積み重ねる

- 将来, 文書品質特性のデータベースとなる
- 文書品質の計測を行う際に, 参考とする
- システム開発文書品質の特性の参考とする

## 共通する文書力と開発のための文書力

- 技術文書に共通する文書品質

- 特許文書
- マニュアル
- システム開発文書

などに共通する文書技術を生かす。



- 「システム開発」のための文書品質

- 配慮しなければならない記述
- より厳密性を求められる内容
- 分野ごとに求められる特徴

などの知識を「形式知化」する。

## 品質が高い文書は人を育てる



Web :<http://asdoq.jp/>

E-mail:secretariat@asdoq.jp (事務局)

ASDoQ

検索



# 有限会社アイ・アール・ディー(IRD 国際特許事務所)

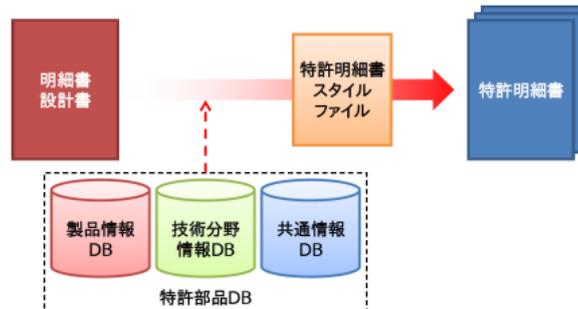
本社住所	〒540-0008 大阪市中央区大手前1丁目7番31号 OMMビル8F私書箱53号	■アピールポイント 特許明細書半自動生成システムは、特許権を取得したい発明の内容を記載した書類(例えば、特許請求の範囲)から、特許明細書等を半自動生成するシステムです。 本システムにより、特許明細書の約40%を自動生成することができます。これにより、特許明細書作成の効率化、特許明細書の高品質化を図ることができます。
URL	http://corp.ird-pat.com	
展示名	特許明細書半自動生成システム PatentGenerator	
お問合せ先	担当 渡辺 TEL 06-6944-4530 FAX 06-6944-4531 E-mail twatanabe@ird-pat.com	

## 【産業日本語との関連】

特許請求の範囲に記載されている請求項を自然言語処理することにより、特許明細書を半自動生成する。

## 【詳細】

明細書設計書の内容に応じて、3種類の特許部品DB(製品情報DB、技術分野情報DB、共通情報DB)から必要な文章を取得し、特許明細書スタイルファイルの内容に従い、特許明細書を生成します。



明細書設計書とは、権利化したい発明の内容を記載した書類です。主に、特許請求の範囲(請求項)を記載します。

部品情報とは、特許明細書を構成する情報(文章)のうち、再利用性の高い情報(定型文)のことです。部品情報は、性質・特性により、製品情報、技術分野情報、共通情報に分類することができます。

製品情報とは、製品の部品を説明する文章(定型文)です。

技術分野情報とは、技術用語(専門用語)を説明する文章(定型文)です。

共通情報とは、特許明細書における定型文です。

# 東芝ソリューション株式会社

本社住所	〒105-6691 東京都港区芝浦 1-1-1(東芝ビルディング)
URL	<a href="http://www.toshiba-sol.co.jp/">http://www.toshiba-sol.co.jp/</a>
展示名	<b>可読性診断技術</b> —産業日本語ライティング支援への活用—
お問合せ先	担当部署 IT 技術研究所 TEL 042-340-6366 FAX 042-340-6013 E-mail kano.toshiyuki@toshiba-sol.co.jp

## ■アピールポイント

理解しにくい文の特徴を考慮した診断規則に基づいて、文の可読性を診断します。

診断結果に基づいて文を修正することにより、読みやすい文章を作成することができます。また、機械翻訳の前処理に利用すると、誤訳しやすい表現を指摘するので、簡単に修正して翻訳精度を向上することができます。

産業日本語ライティング支援ツールとしての活用を検討しています。

## 【産業日本語との関連】

読み手に誤解されやすい表現のチェックを目的として、可読性(理解しやすさ)を判定する技術を開発している。この可読性の判定基準は、産業日本語のライティングルールと共通する点が多く、ライティング支援ツールとして活用できると考えられる。

## 【詳細】

技術文書は、その内容が正確であるだけでなく、内容が読み手に正しく効率的に伝わるのが望ましい。また、機械翻訳等の機械処理においても、書き手の意図が正しく理解され、正確に処理できるのが望ましい。ここでは、記述された文章の理解しやすさを、その文章の可読性とよぶ。可読性の低い文章は、読み手に誤解される危険性が大きく、内容を理解するのに時間がかかる。また、機械処理においても、書き手の意図とは異なる意味に解釈される可能性が高い。可読性診断技術は、文章の理解しやすさを診断し、可読性が低い箇所を検出する技術である。

専門分野の文章の翻訳を翻訳者に依頼する場合、翻訳者が必ずしもその分野に精通しているとは限らない。そこで、翻訳者が内容を正しく理解できるような文章、すなわち可読性の高い文章が求められる。可読性が低い文章は、誤訳を生じさせ、翻訳された文章の修正や再翻訳などの作業増加を招く。

機械翻訳においても、可読性が低い文章では、良い翻訳結果が期待できない。現在の機械翻訳製品は、簡潔でわかりやすい文章に対しては、実用レベルに達した翻訳結果を出力する。しかし、可読性が低い文章を入力すると、正しい翻訳は難しくなり、翻訳の精度が低下する。機械翻訳には人間並みの言語理解能力を期待できないので、翻訳者に依頼するときより一層、可読性の高い文章が求められる。

機械翻訳により誤訳が発生する文の例を次に示す。

[例文 1] 顧客用文書をカラー、社内用文書を白黒で印刷する。

[日英翻訳結果 1] Color and the document for in-house use are printed for the document for customers in black and white.

[日中翻訳結果 1] 在黑白印刷顾客用文书彩色、公司内部用文书。

この例文 1 の場合、「カラー」の直後に述語がないため、機械翻訳は「カラー」の係り先を正しく認識できず、原文の意味とは異なる意味に翻訳してしまう。印刷機能として「カラー」と「白黒」の選択が有ることを知っている人にとっては、その知識で補完することにより、この文を正しく解釈できるが、機械翻訳では「カラー」と「白黒」の関係を捉える前に、「カラー」とその直後の「社内用文書」との並列関係を機械的に捉えてしまい、誤訳を導いている。

このような可読性が低い文を、事前診断で発見できれば、文章作成者に対して修正を促すことにより、人間翻訳や機械翻訳の精度と効率の向上が期待できる。

可読性診断システムでは、構文解析結果を用いて、可読性診断機能群に属する個別の診断機能により、文の可読性を順に診断する。現在取り組んでいる可読性診断機能を次の表に示す。

診断機能名称	説明	※
曖昧な係り受け	係り受け関係の解釈が複数ある箇所を指摘する。	2
複合語	辞書に登録されていない複合語で、直訳では訳しにくい接辞(「可」、「未」、「無」など)を含むものを指摘する。 例: ペット可賃貸マンション、文書管理システム未導入部門	3
述語の省略	述語動詞が省略されている箇所を指摘する。	3
主語の省略	主語が省略されている述語動詞を指摘する。	3
目的語の省略	目的語が省略されている述語動詞を指摘する。	3
主語と述語が離れている	述語動詞から離れている主語を指摘する。	2
目的語と述語が離れている	述語動詞から離れている目的語を指摘する。	2
長い修飾部	修飾部が長い箇所を指摘する。	2
述語の数	述語動詞が多く含まれている文を指摘する。	1
助詞「は」	格の曖昧性がある助詞「は」を指摘する。	2

※産業日本語ライティングルールとの対応 1=短文化、2=構文明確化、3=語彙明確化  
可読性診断システムにより、例文 1 に対して次のような診断結果が得られる。

「カラー」の直後の「で印刷し」が省略された可能性がある。

利用者は、診断結果を参考にして「で印刷し」を補足し、例文 2 のように修正できる。修正結果を機械翻訳すると、正しい意味の訳文が得られる。

[例文 2] 顧客用文書をカラーで印刷し、社内用文書を白黒で印刷する。

[日英翻訳結果 2] The document for customers is printed in color and the document for in-house use is printed in black and white.

[日中翻訳結果 2] 以彩色印刷顾客用文书，以黑白印刷公司内部用文书。

この可読性診断技術を、産業日本語ライティング支援ツールとして活用できないか、評価を行った。適用可能性を検証するために、特許文書の翻訳を想定した評価実験を実施した。その結果、誤訳しやすい文の 7 割を検出することができた。また、「主語の省略」診断を除いた評価では、指摘箇所の 85%で誤訳が発生しており、誤訳の危険性を診断できる見通しが得られた。

展示会場では、可読性診断技術により文章中の可読性が低い箇所を検出するデモを展示する。

# 株式会社インテック

本社住所	〒136-8637 東京都江東区新砂 1-3-3
URL	http://www.intec.co.jp
展示名	<b>Patent Structure Viewer</b>
お問合せ先	担当部署 先端技術研究所 研究開発部 TEL 03-5665-5011 FAX 03-5665-5095 E-mail psv_support@intec.co.jp

## ■アピールポイント

特許書類における「特許請求の範囲」の部分の解析を行い、その読解を支援します。

- 複数の特許請求項の引用関係を視覚的に分かりやすく表示します。
- 特許請求項の構造を視覚的に表示します。
- 特許請求項における特徴的単語を強調表示します。

## 【産業日本語との関連】

特許版・産業日本語で記述されていない既存の特許書類の読解を支援します。

## 【詳細】

Patent Structure Viewer は、特許書類(公開特許公報、特許公報等)における、「特許請求の範囲」の読解を支援するシステムです。「特許請求の範囲」には、「特許請求項」(クレーム)が記述されます。特許請求項とは、発明の権利範囲を記述するためのもので、特許書類中で最も重要な箇所です。特許請求項は通常、複数個記述されますが、その引用関係は複雑なものになることがあります。また、特許請求項は通常、複雑な構造を持った長文で記述されるため、弁理士や知的財産権専門家以外の一般の人にとっては極めて読みにくいものになっています。

Patent Structure Viewer は、特許書類のテキストファイルを入力し、解析を行い、結果をブラウザで表示できる形で出力します。以下の機能を提供します。

- ・注釈付きクレームツリー表示機能
- ・特許請求項の細粒度解析機能
- ・特許請求項における特徴的単語の強調表示機能

## <関連発表>

1. 「特許請求の範囲」読解支援のための言語処理技術の改良と統合化, (2010年12月, 第1回特許情報シンポジウム)
2. 特許請求項の細粒度解析について, (2010年6月, 日本知財学会第8回年次学術研究発表会)
3. 引用形式請求項における内的付加と外的付加の判定とタグ付けについて, (2009年6月, 第7回日本知財学会学術研究発表会)
4. 手がかり句を用いた特許請求項の構造解析, (2004年, 情報処理学会論文誌, pp891-905, Vol.45, No.3)

注釈付クレームツリーと特許請求項の細粒度解析結果は、印刷することができるほか、JPEG形式のファイルで保存することができます。また、特許請求項の解析結果を2種類のCSVファイル(粗い解析結果と細粒度解析結果)として出力することができます。

特開 2008-26300 を用いた表示例を以下に示します。

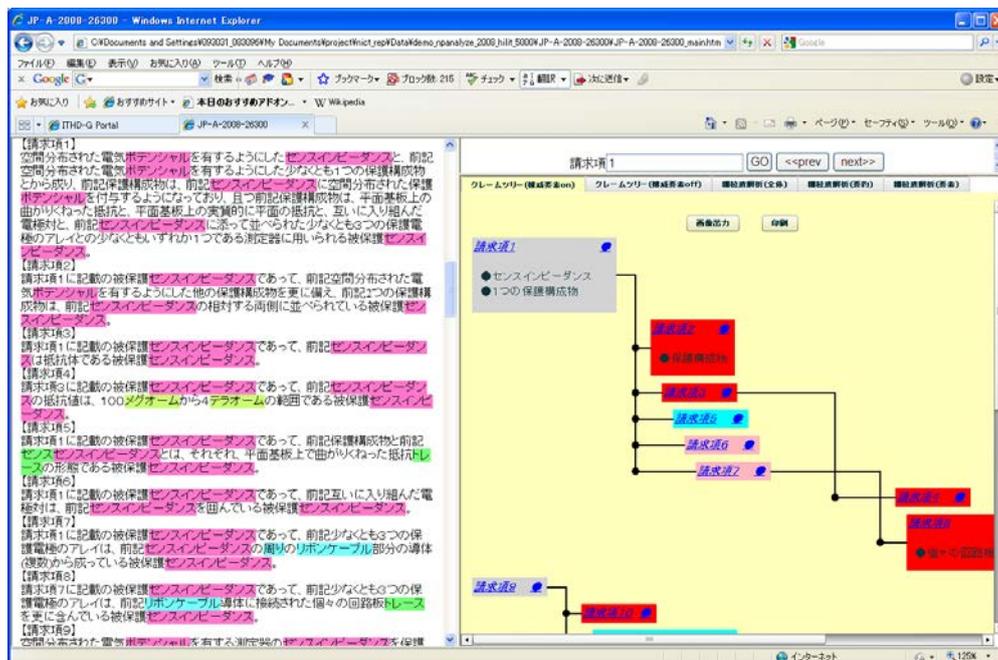


図1. 注釈付クレームツリーの表示(右側)と、特徴的単語の協調表示(左側)

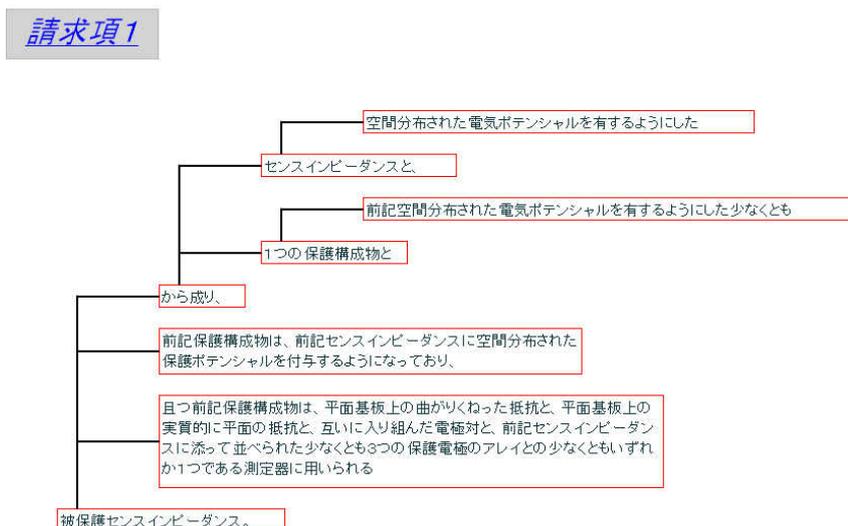


図2. 請求項1の細粒度解析結果

# 株式会社富士通研究所

本社住所	〒211-8588 川崎市中原区上小田中 4-1-1
URL	<a href="http://jp.fujitsu.com/group/labs/">http://jp.fujitsu.com/group/labs/</a>
展示名	<b>特許意味検索の プロトタイプシステム</b>
お問合せ先	担当部署 メディア処理システム研究所 TEL 044-874-2489 FAX 044-754-2923 E-mail <a href="mailto:ushioda@jp.fujitsu.com">ushioda@jp.fujitsu.com</a>

## ■アピールポイント

文の意味構造を活用した、自然文入力による特許ランキング検索のプロトタイプシステムを開発しました。構文解析および意味解析を経て得られた、語と語の間の意味的繋がりを検索の類似度評価に取り入れることにより、従来のキーワードベースの検索システムに比べ、よりユーザの意図に沿った表現を多く含む特許を検索することが可能になります。

### 【産業日本語との関連】

産業日本語の代表例として取り上げられている特許文書について、知的生産性の向上のための日本語意味処理の効果の検証を行っています。検証対象の1つとして、特許検索への適用例を御紹介します。

### 【詳細】

インターネット検索はじめ私達が通常目にする自由入力の検索システムでは、ユーザが指定した1つあるいは複数個のキーワードを入力として、それらのキーワードを含む(あるいは多く含む)文書を検索するタイプが主流です。しかし特許検索のように技術的に込み入った内容の文書を検索する場合、個々のキーワードがお互いどのように意味的に繋がっているかをユーザが指定できた方がユーザの意図に近い文書が検索できると考えられます。たとえば「塩素ガス、劣化、低減」と言う3つのキーワードを検索式(検索クエリー)としてキーワードベースの検索システムに入力した場合、ユーザの意図が「塩素ガスによる劣化を低減させる」ことなのか、「塩素ガスによって劣化が低減する」現象なのか、あるいは「塩素ガスの低減によって劣化する」ことなのか判別することができませんが、自然文で入力できればユーザの意図はより明確になります。そこで、特許検索においてはいわゆる概念検索と呼ばれる検索方式も導入されていて、ユーザは単語ではなくて自然文(文章)を検索クエリーとして入力できるようになっています。しかしながら概念検索システムもその多くが、システム内部では文章から単語や複合語を切り出して、検索対象文書内でのそれらの出現頻度をもとに検索を行っています。その意味で多くの概念検索システムもキーワードベースの検索であると言えます。

富士通研究所では、長年の機械翻訳研究で培った高精度の構文解析および意味解析技術を活用して、クエリー中の語と語の間の意味的繋がりを検索の類似度評価に取り入れることにより、従来のキーワードベースの検索システムに比べ、よりユーザの意図に沿った表現を多く含む特許を検索する技術を研究開発しています。今回の展示ではこの特許意味検索システムのプロトタイプをご紹介します。

### 特許意味検索システムの概要

特許意味検索システムでは文書が表す意味を、文書中に含まれる文の意味構造の総和と定義しています。本研究における意味構造とは格文法に基づいたグラフ構造により表現されます。すなわち、意味構造は、単語の概念を表すノードおよびノード間の関係を表すアークからなる有向グラフにより表されます。図1は、「太郎は花子に本をあげた。」の意味構造を有向グラフにより表した例です。図1において、○で囲まれたものがノードを表し、□で囲まれたものがアークを表します。この例ではノードは、「GIVE」、「HANAKO」、「TARO」、「BOOK」の4つです。アークは「中心」、「目的」、「動作主」、「対象」、「述語」、「過去」の6つです。アークはノード間の関係を表します。1ノードにしか繋がらないアークはそのノードの属性を表します。

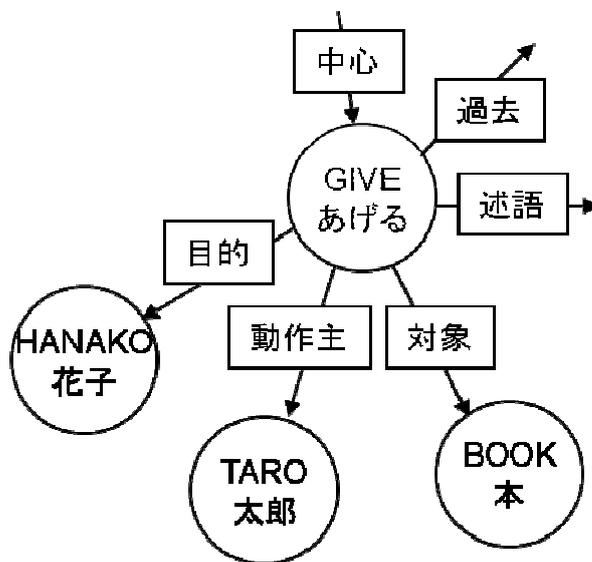


図1. 意味構造

意味を有向グラフとして表現したとき、意味検索はグラフマッチングと考えることができますが、単純なグラフマッチングでは、グラフの複雑性や表現の多様性、また意味解析処理の精度の問題などによりマッチ率に限界があります。そこで、本システムでは、意味構造をアークとそのアークにつながるノード(1つか2つ)の部分グラフに分解し、この部分グラフをベースとする検索アルゴリズムを考案し実装しました[1]。意味構造の抽出には、日英翻訳エンジン ATLAS[2]の翻訳過程で生成される意味解析結果を用いています。ATLAS は中間言語方式の翻訳方式を採用しており、原文を辞書と文法規則に基づき解析して意味構造を計算しています。

[1]大倉清司,潮田明(2012) 意味検索のプロトタイプシステムの構築. 言語処理学会第18回  
年次大会予稿集. 2012.

[2] 富士通.英日・日英翻訳ソフト ATLAS.  
<http://software.fujitsu.com/jp/atlas/>

# 有限会社サイバープロ

本社住所	〒525-0011 滋賀県草津市片岡町292-5
URL	http://www.cyberpro.co.jp
展示名	<b>データベースアプローチに基づく特許機械翻訳システム</b>
お問合せ先	担当部署 営業部 TEL 077-568-1931 FAX 077-568-1931 E-mail cyberoikeda@yahoo.co.jp

## ■アピールポイント

従来の機械翻訳アプローチである、ルールベース、統計ベース、事例ベースに対し、データベースアプローチは、完全対訳節と完全名詞句からなるデータベースの構築と、その応用としての機械翻訳という2段階アプローチをとっている点であり、節の構造を分解しないで、対訳節を対応させている点で特徴がある。これにより、従来自然でわかりやすい翻訳のためには、対訳文の構造を代えなければいけない文の翻訳品質を格段に向上させることができる。また、翻訳に対し、なぜこのような訳になっているかを示すアカウントビリティ機能がついている点で特徴がある。

## 【産業日本語との関連】

制限言語としての産業日本語の確立のためには、語や文法の制約ではその記述能力の保証ができない。そのためには、文中での意味の単位である「節」と、「完全名詞句(節を含まないすべての名詞句)」の内、対象文書記述のためのものを抽出し、制約しなければいけない。更に、機械翻訳の品質を保証するためには、制約した節と名詞句に対する対訳を一意に決めておかなければいけない。この2つの制約を課した文の作成を支援したり、自由に書かれた文を同意の制約文に書き換えたり、制約文を翻訳したりしようとするのが本システムで、産業日本語成立のまさに中心課題を解決するものである。

## 【詳細】

文の節分解は、例えば次のように行う。

S0=このようにして車体を浮上させた場合には、摩擦駆動は行われず、磁気誘導による推進駆動、さらにはこの推進駆動にプロペラによる補助推進駆動を加えた推進駆動となる。	S0= In the case where the chassis is made to float in this manner, frictional drive is not provided, and propelling drive derived from magnetic induction, or auxiliary propelling drive using propellers is added.
S0=@連用形タ接続:た場合には、@受動態-未然形:ず、_、_。(〔S1〕,〔S2〕,〔S3〕) S1=このようにして_を浮上させる(〔N4〕) S2=[_]が摩擦駆動を行う(〔:人〕) S3=_、さらには_を加えた_となる(〔N5〕,〔N6〕,〔N7〕) N4=車体 N5=磁気誘導 N6=推進駆動 N7=プロペラによる補助推進駆動	S0=In the case where @Passive:, @Passive and @Passive:.(〔S1〕,〔S2〕,〔S3〕) S1= the _ is made to float in this manner(〔N4〕) S2=[_] provide frictional drive(〔person 〕) S3=[_] add _, or _ to _(〔N5〕,〔N6〕,〔N7〕) N4= chassis N5= magnetic induction N6= propelling drive N7= auxiliary propelling drive using propellers

この文(関数列)分解の特長は、分解要素(関数)がすべて節(文を含む)か名詞句になっている

ことで、関数内の埋め込みには、活用、態、時制などの変換が伴っていることである。このようにしても、S1 のように使役形日本文が受動態英文の対応することや、S3 のように内部統語構造は異なるものになることや、N7 のように複合名詞句を分解せずに翻訳しないとうまく対応しないものがある。また、これらがルールベースの機械翻訳で誤訳になる1つの原因であった。

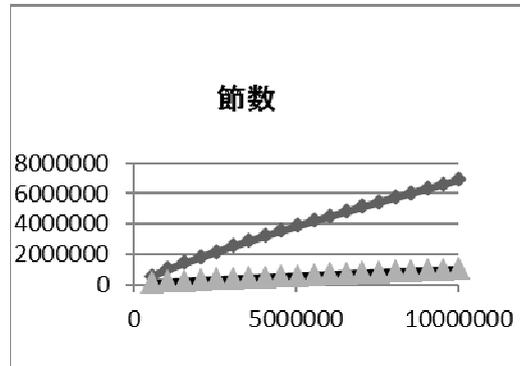
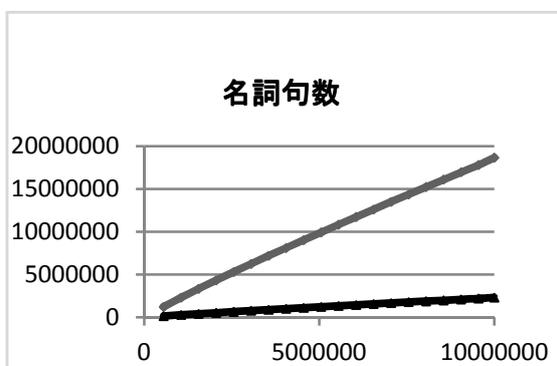
本システムは、これらの問題を解決するため、実在する特許翻訳文(NTCIR-10)から、すべての対訳節、対訳名詞句を自動的に切り出してデータベース化した。そして、その対訳対の内部構造での対応もすべてデータベースに格納した。すべての要素(ここでは、節と名詞句)をすべて具体的な表層レベルで管理するのが、データベースアプローチという所以である。従来のルールベースは、少ない要素(レキシコンと品詞並び規則)ですべての文の表層構造を理解しようとしてうまくいかなかったが、データベースアプローチは、品詞という中間構造物を使わず、具体的な表層レベルでの表現を基本にしている点が重要である。

また、要素分解の段階で、従来共起(collocation)として扱われていた、動詞—目的語、副詞—動詞、形容詞—名詞などの従属関係を、すべて節や名詞句中に閉じ込めてしまった。例えば、上例の S2 のように、[摩擦駆動を行う]—[ provide frictional drive]という対応は、分解して、[[摩擦駆動]—[ frictional drive]]と[を行う]([N])—[ provide ]([N])]の2つの要素には分解できず、一体として扱わなければならない。これも誤訳の一因であった。

データベースアプローチは、このように従来は「例外」扱いされていた語や表現の関係を、表現レベルで網羅的に蓄積することで、例外を出さない言語間の対応(alignment)を実現している。

このアプローチで重要なことは、「節や名詞句の対応をどのようにして網羅的に収集するか」といういわゆる「辞書構築」の問題である。特に、上記の例でも示したように、ある言語で名詞句になっていても、翻訳文では、それを動詞句を使って表現している場合である。この問題を解決するために、名詞句に含意されている節(ここでは、「暗黙節」という。)を見つけ出すことである。これまでの研究で、隠れ節は、「形容詞+名詞」、「名詞+名詞」、「代名詞とその指示対象」、1つの動詞を共有する複数の名詞(主語や目的語)、「文の名詞化」、「接頭語+名詞」、「名詞+接尾語」などに節が隠されていることがわかっており、頻出するこのような名詞句に対しては、実際に名詞句から隠れ節を切り出した。この情報が、節対応(clause alignment)を自動的に見つけるのに重要な役割を果たす。

明示的な節や名詞句を自動的に抽出するために、日本語では Cabocha、英語では Stanford Parser を使い、これで解析した結果を更に、変換して節と名詞句を作成した。特徴的な点は、明示節(黒線)だけではなく、暗黙節(灰線)も切り出している点である。下2図は、1000 万件の英文から抽出した節と名詞句の数である。これから、総名詞句数、節数が推測できる。



これで総名詞句数、節数がわかる。1000 万文格納してもまだ少しずつ増加していく。

# 株式会社クレストック / アイビー・システム株式会社

本社住所	〒431-3105 静岡県浜松市東区笠井新田町 676 番地 株式会社クレストック
URL	<a href="http://www.crestec.co.jp">http://www.crestec.co.jp</a>
展示名	<b>多角的な文書比較『歌詠』と カスタマイズ容易な日本語精査『鶯』</b> 担当部署 アイビー・システム(株) やまと歌担当宛
お問 合せ先	TEL 011-271-7117 FAX 011-271-0550 E-mail <a href="mailto:yamatouta@ivysystem.co.jp">yamatouta@ivysystem.co.jp</a> <a href="http://www.ivysystem.co.jp/yamatoUta/index.html">http://www.ivysystem.co.jp/yamatoUta/index.html</a>

## ■アピールポイント

歌詠は、任意の文書を多角的に比較し、差分を対比表形式で表示するツールです。段落や文ごとの類似度を指定して目的に合った比較ができます。

鶯は、オープンソース形態素解析エンジン MeCab の結果から、文章が正しいか、業界ルールに従っているか判断してコメントするツールです。精査ルールの一部は、カスタマイズが可能です。

## 【産業日本語との関連】

原文と産業日本語文章を文ごとに比較して結果を保存しておく、翻訳に適した文にするための参考情報になります。類似した文の並びから表記のゆれを発見できます。形態素情報を利用して用語の抽出や日本語の精査ができます。

## 【詳細】

### <システムの形態>

1. デスクトップアプリケーションであるスタンドアロン版(Windows XP / Vista / 7 / 8 32/64bit)
2. セキュア・ウェブアプリケーションとしてのカスタマイズ版(Windows Server / Linux)

### <比較機能『歌詠』の特徴>

#### 差分の対比表示

Microsoft Word の比較と異なり、差分結果を分かり易い対比表形式で Web Browser に表示します。外部アプリケーション (Microsoft Word) は不要です。比較結果は、目的に適した表示へ動的に切り替え可能です。ウェブ版では基準文書と2つ以上の文書を同時に比較できます。

#### 移動した要素の検知

比較する文書の間で内容の位置だけが変化していた場合、移動元と移動先を表示します。

#### 差分計算

段落(改行)や文(日本語は句点「。」、英語はピリオド「.」)ごとに区切り、文字または形態素の単位で比較します。差異の大きさは、Levenshtein の編集距離を用いて算出します。

#### 類似度

比較対象の編集距離が小さいほど、差異は小さくなり、類似度は高くなります。したがって一定の類似度閾値以上の場合は、変更された箇所と見なし、下回る場合は、別の内容の要素であると判断します。比較する前に類似度を指定し、比較対象文書に即した所望の結果を取得できます。

#### 比較オプション

文書の先頭から順番に内容を比較する機能、または類似度が最大の要素を探してそれを比較する(文書の構成が異なる文書や行の入れ替えがある表の比較のための)機能を選択できます。

**表記ゆれの検出**

類似度が高い文すべてをリストアップし、文書内にある表記ゆれの可能性を検出します。

**英文比較**

和文と和文以外(欧米文)を区別した比較処理によって、類似度精度を向上しています。

**<日本語精査機能『鶯』の特徴>**

**精査ルール**

以下のように CSV 形式で記述しているため、ルールの削除、変更、追加が可能です。

文章中から用語として取り出す形態素パターンの一部 (\*は任意の値)

○-含める ×-除く	着目する単語				直前の単語				直後の単語			
	単語	品詞	細分1	細分2	単語	品詞	細分1	細分2	単語	品詞	細分1	細分2
○	*	名詞	サ変接続	*	*	接頭詞	名詞接続	*	*	動詞	*	*
×	*	名詞	サ変接続	*	*	*	*	*	*	動詞	*	*
○	*	名詞	サ変接続	*	*	*	*	*	*	名詞	*	*

**用語の抽出**

以上のルールによって、文章中にある見逃し易い用語を抽出し、用語辞書にある登録済の用語と照合します。一般的な用語か、新しい用語か、用語にゆれがないか発見する支援をします。

**文書の内容精査の一例**

- 文章内にある誤り可能性(脱字、誤字、不要な文字)を指摘します。
- 誤りやすい慣用句、用字・用語のチェック、活用形のチェックをします。
- 特許請求の範囲において、「前記」や「該」に続く語句が初出の場合、指摘します。

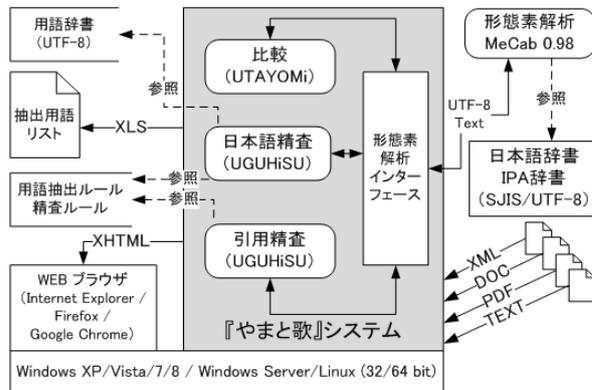
形態素解析の結果から誤り可能性を抽出するルールの一部 (\*は任意の値)

直前の単語				直前の単語の例外				直後の単語				直後の単語の例外			
単語	品詞	細分1	活用型	細分1	活用型	活用	基本形	単語	品詞	細分1	活用型	細分1	活用型	活用	基本形
*	助詞	係助詞	*	*	*	*	*	*	助詞	係助詞	*	*	*	*	*
*	助動詞	*	*	*	*	*	*	*	動詞	*	*	自立	*	*	*

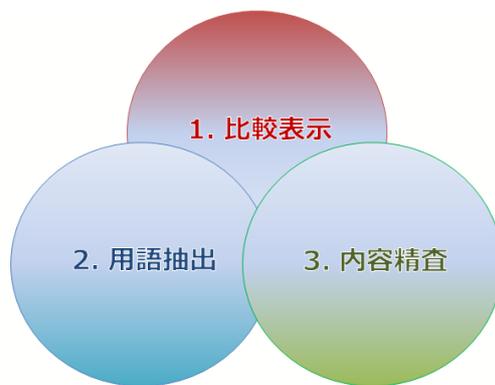
**精査結果の表示**

文書比較時、精査を同時に実行すると比較前後における精査コメントの変化も確認できます。

**<モジュール構成概要>**



**<主な機能>**



# 独立行政法人情報通信研究機構

<b>本社住所</b>	〒184-8795 東京都小金井市貫井北町 4-2-1	<b>■アピールポイント</b> 情報通信研究機構(NICT)は、我が国の経済の成長と発展、豊かで安心・安全な社会の実現の原動力である情報通信技術(ICT)分野の研究開発と事業振興業務を進めております。 多言語翻訳研究室では、産業や文化の発展に資する多言語翻訳技術の研究開発を進めています。 本展示では、翻訳支援環境についてご紹介いたします。
<b>URL</b>	<a href="http://www.nict.go.jp/index.html">http://www.nict.go.jp/index.html</a>	
<b>展示名</b>	<b>翻訳支援環境の研究開発</b>	
<b>お問合せ先</b>	担当部署 ユニバーサルコミュニケーション研究所 多言語翻訳研究室 TEL 0774-98-6300 FAX 0774-98-6300 E-mail <a href="mailto:ltg-info@khn.nict.go.jp">ltg-info@khn.nict.go.jp</a>	

## 【産業日本語との関連】

英語をわかりやすい日本語に翻訳するときに、産業日本語が関係する。

## 【詳細】

NICT多言語翻訳研究室では、「みんなの翻訳」という翻訳支援環境を、東京大学図書館情報学研究室と共同で運営しています。「みんなの翻訳」はボランティア翻訳者の支援を目的としています。

ボランティアの翻訳者は、様々な文書を翻訳しています。ボランティアの翻訳者は、翻訳により、世の中に貢献していると言えます。たとえば、マニュアルの日本語訳は、日本人のユーザーにとっては、大変有難いものですし、ブログの翻訳は、他のメディアが注目しない場所や人々について光を当てると言えます。したがって、ボランティアの翻訳者を支援することは、世の中に貢献することと言えます。

また、日本にいるボランティアの翻訳者は、現在、数千人程度ですが、外国語特に英語を翻訳できる潜在的なボランティア翻訳者の数は、数十万人程度ではないかと思われます。そのため、翻訳をしたい人が、簡単に翻訳ができる環境を提供すれば、現状よりも、もっと多くの人が翻訳をするようになり、より多くの外国の情報を取り込めるとともに、日本の情報を発信することもできるようになると思います。

このような動機から、NICT多言語翻訳研究室では、東京大学図書館情報学研究室と共同で、「みんなの翻訳」を開設しました。

みんなの翻訳の特徴は、(1)東京大学で開発された高機能な翻訳支援エディタ QRedit を誰もが利用できることと、(2)みんなの翻訳で公開されている翻訳には、「一定の条件の下で、二次的著

作物を作成し、それを公開しても良い」というライセンスが付与されているため、適切な使用であれば、翻訳を利用できるということと、(3)三省堂の協力により「グランドコンサイス英和辞典(36万項目収録)」が翻訳支援に利用できることです。

翻訳支援エディタ QRedit の基本設計理念は、以下の4点に集約されます。(1)新たな情報・機能を提供するのではなく、翻訳者が現に行っている作業の手間を省く、(2)システムが決めるのではなく翻訳者が決めるのに必要な情報を提供する、(3)翻訳者の発想を豊かにする情報を表示する、(4)できるだけシンプルにする。これらの方針は、翻訳者へのインタビューおよび現状の翻訳支援技術の水準に基づいて決めました。

翻訳結果を共有するためには、原文と翻訳文の使用許諾について考慮する必要があります。たとえば、当然ですが、原文の著者が翻訳文の公開を許可していない場合には、翻訳文は公開できないので、翻訳結果を共有することはできません。

そのため、みんなの翻訳の利用者には、原文と翻訳文の使用許諾について確認を求めています。また、みんなの翻訳の利用者には、各自が翻訳した文は、二次的利用ができるように許可することを求めています。そのために、システムは、みんなの翻訳の利用者が翻訳文を保存するときに、使用許諾などを確認しています。このようにして、みんなの翻訳では、原著者や翻訳者の著作権を尊重しつつ、翻訳を共有できる仕組みを準備しています。

また、NICTは、みんなの翻訳で得た知見を元にして、翻訳支援エディタだけでなく、機械翻訳や対訳抽出や対訳文アライメントなどを統合した翻訳支援環境も開発しています。

不許複製 禁無断転載

高度言語情報融合フォーラム (ALAGIN)

言語処理学会

一般財団法人日本特許情報機構 (Japio)

発行日 平成25年3月1日

発行者 高度言語融合フォーラム (ALAGIN) 内

産業日本語研究会・シンポジウム事務局

本シンポジウムに関する問合せ先

産業日本語研究会・シンポジウム事務局

email: [info@alagin.jp](mailto:info@alagin.jp)