

第9回産業日本語研究会・シンポジウム 予稿集

平成30年3月5日

於 東京・丸ビルホール（丸の内ビルディング7階）

高度言語情報融合フォーラム



一般財団法人
日本特許情報機構

Japio

第9回産業日本語研究会・シンポジウム

まえがき

平成30年3月

ユニバーサルコミュニケーションに資する産業日本語に向けて

～グローバル化の波を越え、人工知能の活用を促進していくために～

産業日本語研究会では、産業・科学技術情報の発信力強化や知的生産性の向上を通じて、わが国産業界全体の国際競争力の強化に資するような、人間が理解しやすく機械が処理しやすい「産業日本語」のあり方を研究しています。この「産業日本語」は、明瞭な日本語文の作成と高品質な翻訳文の低コスト作成に寄与することなどを目的としています。

情報を効率的・効果的に機械処理することで、言語等の違いを乗り越え、すべての人が容易に情報の伝達・共有を可能とするユニバーサルコミュニケーション技術が求められております。そこで、今回のシンポジウムでは、グローバル化と人工知能という2つの観点から、ユニバーサルコミュニケーションに資する日本語を議論する場を提供したいと考えています。

海外進出や顧客の多国籍化などによる企業活動のグローバル化にともない、国内外への産業・科学技術情報の発信先に応じた情報発信が進められております。前者の視点からは、日本語を母語としない者も含めた様々なバックグラウンドを有する者との円滑なコミュニケーションのためにも、思考及び情報伝達の基盤である日本語を見直す必要があります。

後者の視点からは、ニューラルネットワークをはじめとする人工知能技術をいかに活用するかが重要となります。多言語での情報発信や文章データ処理の高度化など、様々な場面での活用が想定されますが、人工知能は万能ではなく、出力される情報は入力データの質に影響を受けます。そのため、入力データとなる日本語のあり方を見直す必要があります。

このような背景のもと、グローバル化が一層進み、人工知能技術が実用化される時代において、求められる日本語スキルや産業・科学技術文章データ活用を考える上で示唆に富む最新の取り組みや研究についてご紹介していただきます。

本シンポジウムが、産業日本語のさらなる普及につながり、わが国産業に大いに貢献できる機会になると期待しております。産業界、学术界などからの、多くの皆さまのご参加をお待ちしております。

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)
委員：辻井 潤一	(産業技術総合研究所)
橋田 浩一	(東京大学)
隅田 英一郎	(情報通信研究機構)
柏野 和佳子	(国立国語研究所)
潮田 明	(産業技術総合研究所)
横井 巨人	(日本特許情報機構)

- 主催：高度言語情報融合フォーラム（ALAGIN）、日本特許情報機構（Japio）
- 後援：総務省、文部科学省、経済産業省、特許庁、千代田区、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、
言語処理学会、日本経済団体連合会、日本知的財産協会、
アジア太平洋機械翻訳協会、大学技術移転協議会、フジサンケイビジネスアイ
- 日時：2018年3月5日（月） 13：00-18：00
- 場所：東京・丸ビルホール（丸の内ビルディング7階）
<http://www.marunouchi-hc.jp/hc-marubiru/access.html>
- テーマ：ユニバーサルコミュニケーションに資する産業日本語に向けて
～グローバル化の波を越え、人工知能の活用を促進していくために～
- 参加費：無料（事前登録制）※シンポジウム後の懇談会・意見交換会は要会費

■プログラム：

【オープニング】 13:00－13:10

(1) 開会挨拶：

長尾 眞 産業日本語研究会 世話人会 顧問／
公益財団法人国際高等研究所 所長／
京都大学 名誉教授

【第一部】グローバル時代の使える日本語文書スキル 13:10－14:40

- (2) 基調講演：「大学におけるライティング教育―《考える》学生を育てる―」 ----- 1
佐渡島 紗織 早稲田大学 教授
- (3) 基調講演：「日本に住む外国人の80%以上が理解する『やさしい日本語』の構造」 --- 5
佐藤 和之 弘前大学 教授

【第二部】伝わる文書の作成に向けて 14:40－16:00

(4) ポスターセッション：

- (4.1) 「産業日本語研究会・ライティング分科会活動報告」 ----- 17
佐野 洋 東京外国語大学 教授／
産業日本語研究会 ライティング分科会 主査
- (4.2) 「産業日本語研究会・文書作成支援分科会活動報告」 ----- 21
橋田 浩一 東京大学 教授／
産業日本語研究会 文書作成支援分科会 主査
- (4.3) 「産業日本語研究会・特許文書分科会活動報告」 ----- 27
谷川 英和 IRD 国際特許事務所 所長・弁理士／
産業日本語研究会 特許文書分科会 主査
- (4.4) 「特許ライティングマニュアル（改訂版）紹介」 ----- 33
白土 博之 一般財団法人日本特許情報機構
特許情報研究所 調査研究部研究企画課 課長
- (4.5) 「システム開発文書品質研究会（ASDoQ）の取り組み紹介」 ----- 37
塩谷 敦子 合同会社イオタクラフト 代表社員／
システム開発文書品質研究会 幹事
- (4.6) 「文章読解・作成能力検定（呼称「文章検」）」 ----- 41
八田 香里 公益財団法人日本漢字能力検定協会 常任理事

- (4.7) 「みんなでニューラル機械翻訳 (NMT) を育てよう！
～「みんなの自動翻訳@TexTra®」、「翻訳バンク」～」 ----- 45
隅田 英一郎

国立研究開発法人情報通信研究機構
先進的音声翻訳研究開発推進センター
副研究開発推進センター長

- (4.8) 「Japio 世界特許情報全文検索サービス紹介」 ----- 51
高橋 幸生 一般財団法人日本特許情報機構 営業推進部 部長

【第三部】 AI 活用時代の日本語データの高度活用事例 16:00—17:30

- (5) 招待講演：「法制執務業務の効率化に向けて (e-LAWS^{イーローズ}) の整備)」 ----- 61
小高 久義

総務省 行政管理局行政情報システム企画課
情報システム管理室長

- (6) 招待講演：「大量日本語データから得られる知見の産業応用
ー内容分析から筆者の性格推定までー」 ----- 73
那須川 哲哉

日本 IBM 株式会社 東京基礎研究所 主席研究員

- (7) 招待講演：「NTT データの AI 研究事例 ーチャットボットから AI 記者まで」 ----- 93
小間 洋和

株式会社 NTT データ 技術開発本部 主任

【まとめ・クロージング】 17:30—18:00

- (8) 全体討論 (含む、講演者への質疑応答) :
進行役：井佐原 均

産業日本語研究会 世話人会 代表／
豊橋技術科学大学 情報メディア基盤センター長・教授

- (9) 閉会挨拶 :
井佐原 均

産業日本語研究会 世話人会 代表／
豊橋技術科学大学 情報メディア基盤センター長・教授

※懇親会・意見交換会

シンポジウム終了後、丸ビル 5 F 小岩井プレミナールにて懇親会・意見交換会
(定員：先着 50 名、会費：3,000 円) の開催を予定しております。

【第一部】

グローバル時代の使える日本語文書スキル

基調講演

「大学におけるライティング教育

―《考える》学生を育てる―」

2010 年頃より、日本では、1 年生にライティング授業を履修させる大学が増えた。学生一人ひとりに対するライティング支援を行う大学も急速に増えた。大学ライティング教育は、何を指すべきか。どのような学生を育成して社会に送り込む使命を持つのか。《書く》とは《考える》ことであり、《考える》ことを鍛えるのがライティング教育の目指すところである。

佐渡島 紗織

早稲田大学 教授

大学におけるライティング教育

—《考える》学生を育てる—

早稲田大学 佐渡島紗織

1. 講演の趣旨

デジタル時代に入り、学生たちは、一日中スマホを手を過ごしている。大学の授業で課されるレポートや発表の準備においても、多くの時間をインターネットに掲載された情報を読むために使っている。

筆者が担当する初年次生向けのアカデミック・ライティング授業では、年々、学生の提出する文章の内容が学生間で似通ったものになる傾向がある。課されたテーマについて、まずインターネットで論点を拾い上げ、それから文章を書くという学習過程を多くの学生がとっていると推測される。論点同士が矛盾していたりし、明らかに複数のサイトから繋ぎ合わせたことが分かる文章もある。

インターネットで世界中と繋がり、多種多様な情報を手に入れることができる時代となったのに、学生たちは《考える》ことをしなくなっているのではないかという危機感をもつようになった。多種多様な情報を容易く手に入れることができるようになったからこそ学生たちが《考える》ことをしなくても済むようになってしまったのであるなら、それは残念なことである。

本講演では、こうしたデジタル時代においていかに《考える》学生を育てるかという、筆者の日頃の苦悶を披露する。ライティング教育という一領域において、大学は何ができるのか。《考える》学生を育てるためのライティング指導や支援とはどのようなものか。その取り組みを紹介する。

2. 文章を通してみる学生の様子

多種多様な情報に容易く触れることができるため、学生は自己の意見をじっくりと創り上げることができない傾向にある。「スルー・プット」という言葉に表されているように、様々な情報を繋ぎ合せて文章を作成し、それをもって自分の意見としてしまう。他者の書いた言葉や数字と自分の考えとの区別が曖昧である。

また、どこかに答えがある、物事にはすべて正解があるという前提をもっている学生が多い。助けを求めれば、インターネットが応えてくれる、大学が支援してくれる、と考えている学生は多い。

3. 《考える》学生を育てるライティング授業

初年次生の約半数(約 4500 人)が履修するライティング授業では、8 編の文章を書く課題を出す。インターネットから論点を拾っただけでは書けない作文課題を考案するようにしている。また、他者の文章から引用させ、引用箇所と自分の論との関係を自覚するように促している。接続表現など、自分の論の展開が自覚できるような言葉の使い方を学ばせる。書いている内容の抽象度を自覚し調節しながら書く技術も学ばせている。語句の使い方を検討することによって《考え》が明確になるような指導に取り組んでいる。

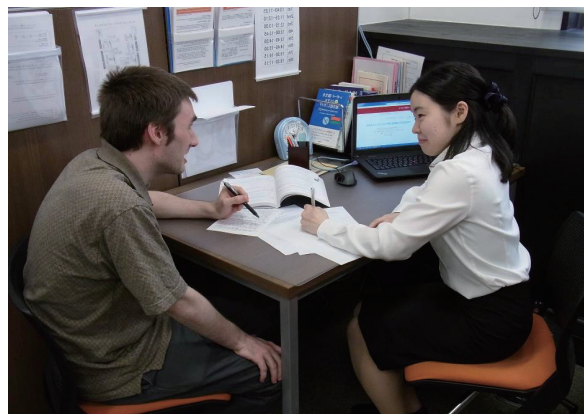
早稲田大学における 10 年間の取り組みを紹介する。本授業は、オンデマンド授業で多数の履修者が一度に受講できるが、文章指導をする大学院生を育て、大学院生が 8 編すべての文章に個別にコメントを付けて返却している。専門領域を超えて志願してくる大学院生たちは、ライティング指導を通して成長し、社会人となっている。



4. 《考える》学生を育てるライティング支援

「ライティング・センター」という、文章作成の個別支援機関では、「自立した書き手の育成」を目指して支援を行っている。学生や教員が文章や文章課題を持って訪れると、専門的な訓練を受けた文章支援者が、対話をしながら一緒に文章を修正したり構想したりする支援をする。添削をするのではなく、書き手が独りで文章の問題点を見つけ、修正法を考え出すことができるように助言をする。

早稲田大学における 15 年間の取り組みを紹介する。添削をせずに、書き手が自ら気づき考えるような対話を展開するには、高度なテクニックが必要である。大学院生である支援者たちに、このテクニックを身につけさせるために行なってきた研修や審査を紹介する。ライティング支援の技術を身に付けた院生たちも、社会に出てそれぞれの領域で活躍している。



基調講演

「日本に住む外国人の 80%以上が理解する

『やさしい日本語』の構造」

「やさしい日本語」という表現法がある。災害への備えで、行政やコミュニティ FM が導入を進める。日本に住んで 1 年以上の外国人なら、漢字圏か非漢字圏かに関わらず、等しく 80%以上が理解できるようにした日本語である。

外国人に日本語だけど読んでみよう、聞いてみようと思わせる表現規則と情報の配列規則に基づく。命に関わる情報を可能な限り簡潔に、やさしく言い換えて伝える。シンポジウムではそのような「やさしい日本語」の構文を紹介する。

検索エンジンで「弘前大学 やさしい日本語」と入力ください。事前に詳しく知ることができます。

佐藤 和之

弘前大学 教授

第9回産業日本語研究会（東京・丸ビルホール）

日本に住む外国人の80%以上が理解する 「やさしい日本語」の構造

（予稿集用原稿）

佐藤和之（弘前大学）
2018年3月5日

「やさしい日本語」内容梗概

「やさしい日本語」という表現法がある。災害への備えで、国や地方行政、コミュニティFMなどが導入を進める。

2016年11月22日朝の「すぐ にげて」「つなみ！にげて！」と伝えたテレビでの津波警報が象徴的である。

同時刻にスマートフォンから伝えられた「やさしい日本語」による緊急地震速報や緊急津波速報は2015年から行われている。

「やさしい日本語」は、日本に住んで1年以上の外国人なら、出身が漢字圏か非漢字圏かに関わらず、等しく80%以上が理解できる文法表現で伝える日本語のことである。

外国人に、日本語だけど読んでみよう、聞いてみようと思わせる表現規則と情報の配列規則に基づく。命に関わる情報を可能な限り簡潔に、やさしく言い換えて伝える。

シンポジウムではそのような「やさしい日本語」の構文を紹介する。

大規模災害発生からの72時間と被災者に伝えるべき情報の表現

災害発生からの72時間で伝える情報は、受け手が的確に理解し、また伝える側も正確かつ迅速に作れる表現でなければならない。

災害発生下での情報はつぎつぎに生じ、すぐに変化する。さらに日本にはいろいろな国のことばを話す外国人も住む。それら情報を外国語にする人手も時間もないのが災害である。

このことへの対処策として、おおむね2000の語を12の規則にそって伝える災害発生後72時間のための日本語表現を規則化し提案した。2017年6月末の活用数は、国や47都道府県を合わせ700事例である。

- 1) 大規模災害下を生き延びる情報を保障する
- 2) 多言語支援を妨げない
- 3) 誤訳が少なく迅速な多言語化への元文となる
- 4) リアルタイムで行政が外国人住民に情報を伝えられる
- 5) 日本人への情報伝達を補完する
- 6) 外国人住民に地域復興の力となってもらえることばである

日本に住んで1年以上・出身国が漢字圏か非漢字圏かに関わらない80%以上の外国人が理解する「やさしい日本語」の表現とは

【津波からの避難指示】

津波が くるかもしれません。
津波は 海から 来る 高い 波です。
海や 川の 近くへ 行かないで ください。
津波は 何回も きます。すぐ 逃げて ください。

【避難所への誘導】

近くの 避難所に 行って ください。
避難所は みんなが 逃げる ところです。
外国人も 使うことが できます。
ぜんぶ 無料です。お金は いりません。

災害発生直後は、防災無線や市町村役場の広報車で伝える。
放送文は1分あたり、すべてひらがなで書いて360文字を読む。
この読み方だと外国人にも日本人にも受け入れられ、よく理解される。

注意 Attention 주의 Attenção

もらうことが できます



あたたかい 食べ物を
もらうことが できます
お金は いりません

(ところ) _____

(時間) 午前・午後 _____ 時 _____ 分から
午前・午後 _____ 時 _____ 分まで

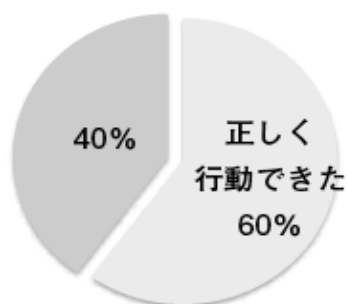
(作った 日) _____ 年 _____ 月 _____ 日 (作った ところ) _____ 93

- ① 避難している外国人が注目するように居住外国人の母語で見出しを付ける
- ② 見出しは大きく、太く書く
見出し語は動作を表すことばにする
- ③ 「やさしい日本語」で伝えている情報を補うイラストを載せる
イラストは簡潔に、重要な要素だけにする
- ④ 伝えたい情報や行動を指示する内容を「やさしい日本語」で表現する
- ⑤ 漢字にはふりがなを付ける
- ⑥ 必要に応じて地図を載せる
地図は最小限の情報で、単純化する
- ⑦ 場所の説明は大きく、目立つように書く
この漢字へもふりがなを付ける
- ⑧ 時間は午前・午後を使った12時間表記にする
- ⑨ 情報の新旧と信頼度を伝えるため、掲載物を作った年月日と機関の名称を書く
年は西暦を使う
□内に掲示資料番号を付ける

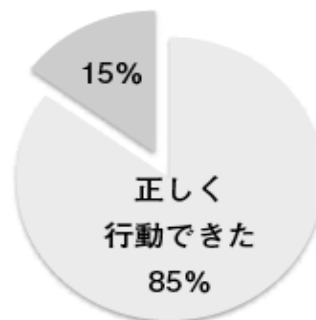
避難所に移ってからは、生活のための情報を掲示物で知らせる。そこには日本人用の掲示物もたくさん貼り出されている。

その中にあって外国人の目を引くもの、そして日本語だけど読んでみようという気にさせる、さらに、日本人にも受け入れられる情報の書き方と表現にしている。

やさしい日本語の信頼性



普通の日本語による行動指示
(N = 43)



「やさしい日本語」による行動指示
(N = 42)

読みことばと書きことばを使った「やさしい日本語」の有効性を検証した

日本語能力が初級後半から中級前半の、漢字圏と非漢字圏出身の外国人による二つのほぼ等しいグループによって実験。読みことば（放送）と書きことば（掲示物）による同じ内容の指示を一方は普通の日本語で、他方は「やさしい日本語」で伝えた。

- 「やさしい日本語」の方が指示された通りの行動をしていた
その理解率（指示の達成率）は最小で85%、最大で91%だった
- 普通の日本語による理解率は29% から良くて60% だった

「やさしい日本語」文法12の規則

災害時に外国人が的確な行動を起こせる表現のための「やさしい日本語」は12の文法規則から成る

- (1) 難しいことばを避け、簡単な語を使う
日本語能力検定試験の旧4級と3級の約2000語を使用
(「火災が発生しました」を、「火事がおきました」に言い換える程度)
- (2) 1文を短くする
(平仮名だけで書いたとして、24文字以内を目指す)
- (3) 災害時によく使われることばや知っておいた方がよいことばは、そのまま使う
(たとえば津波、震度、余震、避難所等には説明を加えて使う)
- (4) カタカナ外来語はできるだけ使わないようにする
ダイヤル ……原語とは意味が違うので通じないため
ライフライン… 原語とは意味が違うので誤解を招くため
デマ…………… 原語では行われない省略なので意味が通じないため
- (5) 擬態語や擬音語は使わない
- (6) 動詞を名詞化した表現はわかりにくいので動詞文にする
(「壁に割れを見つけたら」のような表現のこと。名詞になっている「割れ」を「割れたところ」と、動詞にして伝える)
- (7) あいまいな表現は避ける
- (8) 二重否定の表現は避ける
(「飲めないことはない」といったような表現のこと)
- (9) 文末表現はなるべく統一する
(指示や注意喚起、可能なことなどを確実に伝えるため)
- (10) ことばのまとまりを認識しやすいよう短いポーズを多く用いる
- (11) 使用する漢字や漢字の使用量に注意する
(一文あたり3から4字が目安)
また、全ての漢字にルビを振る
- (12) ローマ字は使わないようにする

「やさしい日本語」の応用

「やさしい日本語」は、日本語能力が初級の外国人が、災害時に生き延びるための情報を伝えることを目的に考え出された。

災害下では外国人だけでなく、日本人にとっても「やさしい日本語」は効果がある。情報は錯綜し、判断能力が著しく低下する状況にあって「やさしい日本語」による情報は、的確な判断に役立つ。

日本人にとっての「やさしい日本語」はまた、子供から高齢者までの日本語を母語とする人たちにとって、もっとも得意な外国語となる。

だからと言って、もし「やさしい日本語」の文法規則を拡大解釈、あるいは日本人にとってより自然な日本語表現にしまうと、それは初級外国人が理解できる「やさしい日本語」でなく、日本人にとってわかりやすい日本語に変質してしまうことに注意しなければならない。

このことを理解した上での「やさしい日本語」の応用が大切である。そこで研究室と研究会は、災害下の外国人を対象にした「やさしい日本語」には「 」を付し、それ以外の用途のわかりやすい日本語には「 」を付さないやさしい日本語にする提案もしており、災害下の外国人住民の安全を担保するようにしている。

「やさしい日本語」を学ぶ初期段階で 必要な資源の情報とアドレス

弘前大学の社会言語学研究室と「やさしい日本語」研究会は、「やさしい日本語」で被災外国人に情報を伝えようとしている機関やボランティアに向けて「やさしい日本語」資源を無料で公開している

- 「やさしい日本語」について詳しく知りたいとき
弘前大学人文学部社会言語学研究室・減災のための「やさしい日本語」研究会
<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ3mokuji.htm>
<http://human.cc.hirosaki-u.ac.jp/kokugo/ejpamphlet2.pdf>
佐藤和之(2016)「外国人被災者の負担を減らす『やさしい日本語』」
『わかりやすい日本語』くろしお出版
- 「やさしい日本語」で使える語彙と漢字について知りたいとき
「やさしい日本語」語彙・漢字表
<http://human.cc.hirosaki-u.ac.jp/kokugo/ej-p8.html>
<http://human.cc.hirosaki-u.ac.jp/kokugo/ej-p7.html>

- 「やさしい日本語」の有効性と信頼性について知りたいとき
公開実験「みんなで減災2005 in ひろさき ～ 災害情報を「やさしい日本語」で」概要
<http://human.cc.hirosaki-u.ac.jp/kokugo/EJ10honjikkenngaiyou.htm>
馬場康維(2007)「実験による検証」『「やさしい日本語」が外国人の命を救う』「やさしい日本語」研究会
- 「やさしい日本語」にする12の規則を知りたいとき
『〈増補版〉「やさしい日本語」作成のためのガイドライン』
<http://human.cc.hirosaki-u.ac.jp/kokugo/ej-gaidorain.pdf>
- 「やさしい日本語」文の作り方を知りたいとき
<http://human.cc.hirosaki-u.ac.jp/kokugo/zouhomanual-pdf/1-annbunn/zouho35-46EJtukurikata.pdf>
- 「やさしい日本語」へ独学で翻訳できるようになりたいとき
「Eラーニング版 わかる！伝わる！はじめての『やさしい日本語』」
<http://human.cc.hirosaki-u.ac.jp/kokugo/e-learningtop.html>

- 外国人住民や外国人児童へスマートフォンを使って緊急連絡したいとき
 - さくさく作成！「やさしい日本語」を使った緊急連絡のための案文集
～災害時における学校や自治体からのお知らせ編～
<http://human.cc.hirosaki-u.ac.jp/kokugo/sakusaku.html>
 - さくさく作成！「やさしい日本語」を使った緊急連絡のための案文集②
～災害時におけるスマートフォンメールでの連絡編～
<http://human.cc.hirosaki-u.ac.jp/kokugo/sakusaku2-smart-top.html>
- 「やさしい日本語」の具体例を体系的に入手したいとき
『増補版 災害が起こったときに外国人を助けるためのマニュアル』
<http://human.cc.hirosaki-u.ac.jp/kokugo/zouhomanual-top.html>
- 防災無線や広報車、コミュニティFMなどでの放送に使う案文を知りたいとき
 - 放送用 時系列案文
<http://human.cc.hirosaki-u.ac.jp/kokugo/zouhomanual-pdf/1-annbunn/zouhohosokuanbun-jikeiretu.pdf>
 - 放送用 情報内容別案文
<http://human.cc.hirosaki-u.ac.jp/kokugo/zouhomanual-pdf/1-annbunn/zouhohosokuanbun-naiyou.pdf>
 - 外国人にも日本人にも伝わる「やさしい日本語」放送文の読み方スピード
<http://human.cc.hirosaki-u.ac.jp/kokugo/onnseikennsyoukekka-bunki.html>

- 「やさしい日本語」を使った掲示物の具体例について知りたいとき
 - ポスターやビラで知らせる
<http://human.cc.hirosaki-u.ac.jp/kokugo/poset-ikkatubannngounasi.pdf>
 - 新しくポスターやビラを作る
<http://human.cc.hirosaki-u.ac.jp/kokugo/zouhomanual-pdf/2-keizibutu/151-184poster-tukuru.pdf>
- 「やさしい日本語」を周囲に説明するパンフレットが欲しいとき
 パンフレット・「やさしい日本語」が外国人被災者の命を救います。
<http://human.cc.hirosaki-u.ac.jp/kokugo/ejpamphlet2.pdf>
- 「やさしい日本語」化支援システム（YAsashii Nihongo Slen System）ソフトが欲しいとき
 東北大学 大学院工学研究科 通信工学専攻 伊藤彰則研究室
<http://www.spcom.ecei.tohoku.ac.jp/~aito/YANSIS/> （パソコン用）
<https://play.google.com/store/apps/details?id=util.ejadvisor3app>
 （アンドロイド用）
- 「やさしい日本語」についてのその他の資料や情報が欲しいとき
 弘前大学人文社会科学部社会言語学研究室サイトマップ
<http://human.cc.hirosaki-u.ac.jp/kokugo/sitemap.htm>

「やさしい日本語」活用の広がり

「やさしい日本語」を使って外国人住民に災害時の情報を伝えるという考え方は、内閣府での施策にも反映されている。

内閣府では、日本語能力に不安のある日系定住外国人に対しての情報提供の在り方を検討しているが、「やさしい日本語」は速やかで正確な情報提供が可能なことから、通常時のみならず災害発生時等にも大きな効果を発揮することが期待されており、内閣府のみならず全国的にも普及させる必要がある。（『定住外国人施策ポータルサイト掲載におけるやさしい日本語の活用に関するPlain English（平明な英語）についての調査』）

「やさしい日本語」研究に携わる協働研究者

「やさしい日本語」研究は、弘前大学の学生とさまざまな機関の研究者、行政職員、消防職員、医療者、コミュニティFMおよびNPO職員との協働で行われている。参加している研究者等の所属機関は以下の通り（五十音順）。

NPO法人CAST・京都工芸繊維大学・コミュニティFMアップルウェーブ・国立国語研究所・さかもとともみクリニック・佐藤内科医院・大東文化大学・統計数理研究所
 東京農工大学・東北大学・弘前消防事務組合・ひつじ書房・弘前大学・藤盛医院

【第二部】

ポスターセッション

伝わる文書の作成に向けて

「産業日本語研究会・ライティング分科会活動報告」

佐野 洋

東京外国語大学 教授／

産業日本語研究会 ライティング分科会 主査

ライティング分科会の活動報告

【活動目的と成果】

ビジネス分野をはじめ、広く一般に公共性の高い産業日本語の書き方を普遍財として普及していく。本年度、伝わる文書を作成するための日本的アプローチと欧米的アプローチを明らかにした。

- 文書には構造がある ■ 文章には「語る姿勢」がある（納得してもらう or 説得する）

文書と文章、段（パラグラフ）

文書	【役割】内容の確約（コミット）		主題
	【形式】儀礼（プロトコール）と文書構造		発信者と受信者
	文章	【語る姿勢】合理的な心構え（表明の戦略）	共感と納得，意志と説得
		【形式】筋書き（アウトライン）	「起・承・結」
		段（パラグラフ）	因果性の表現
		【ことば遣い】書き手の態度	話題と支持
		【形式】字下げと話題文，支持文	

文書を書く時に点検

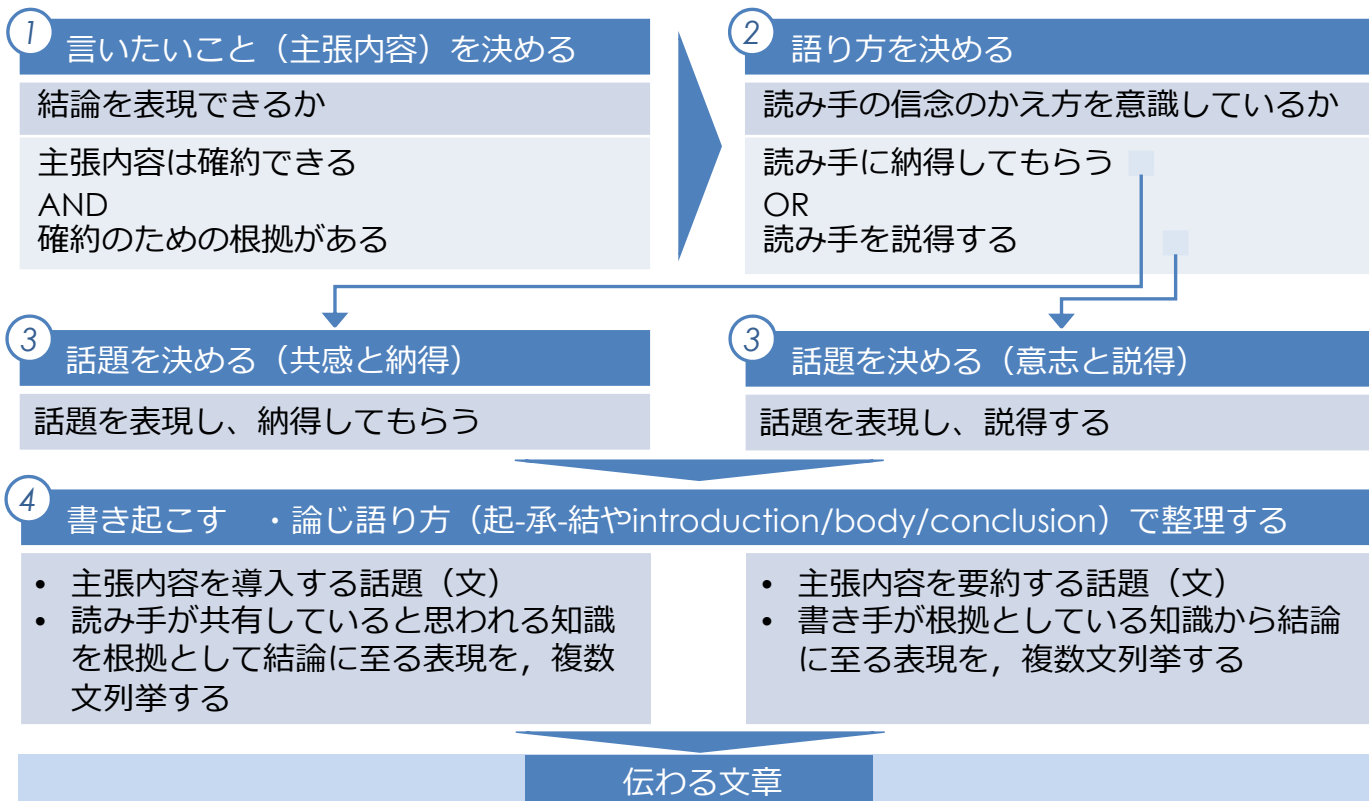
- 発信者と受信者を確認する
- 読み手との共有知識を確認する
- 文書書式を確認する
- 文書儀礼を確認する



- ✓ 両者の関係性を理解しているか
- ✓ 読み手の知識を具体的に意識できているか
- ✓ 適切に把握しているか
- ✓ 適切に記述できているか

伝わる文章の書き方

①、②、③で文群を想起して、④でパラグラフ群にまとめる



今後の課題：効果的な支持文の表現調査，主旨文の書き方

「産業日本語研究会・文書作成支援分科会活動報告」

橋田 浩一

東京大学 教授／

産業日本語研究会 文書作成支援分科会 主査

文書作成支援分科会

2018-03-05

多くの受験生の文書読解能力はAIに劣る

- 国立情報学研究所等の「ロボットは東大に入れるか」プロジェクト(2011～)
 - ◆さしあたり答えはNO (2016)
 - *入試の問題文が複雑
 - ◆2016年の全国統一模試

	国語	数学IA	数学IIB	英語 R/W	英語L	物理	日本史	世界史	合計
満点	200	100	100	200	50	100	100	100	950
人間	96.8	54.4	46.5	92.9	26.3	45.8	47.8	44.8	437.8
AI	96	70	59	95	14	62	52	77	525

2

多くの人々の文書読解能力はAIに劣る

- 生徒は限られた試験時間内に効率的に得点を稼ぐための受験テクニックを使っている？
- しかし実務でも時間は限られているから、大人も同様のテクニックを使
- 米国の大学では在学中に批判的思考等の能力がほとんど向上しない
 - ◆ Richard Arum and Josipa Roksa (2011)
Academically Adrift: Limited Learning on College Campuses. University of Chicago Press.
- 日本の大学教育の効果が米国のそれを上回るとは考えにくい
- まとまった内容の文書の作成・読解に日常的に携わっているのは勤労者のほんの一部

3

構造化文書の図式表現

- 従来型の文書の作成・読解の訓練を多くの人々に日常的に課するのは無理
- 文書自体を作成・読解しやすくする必要あり
- 図式による意味構造の明示
 - ◆ わかりやすい
 - ◆ 作りやすい
 - ◆ 質が高い
 - ◆ 頭が良くなる
 - ◆ 批判的思考力がつく

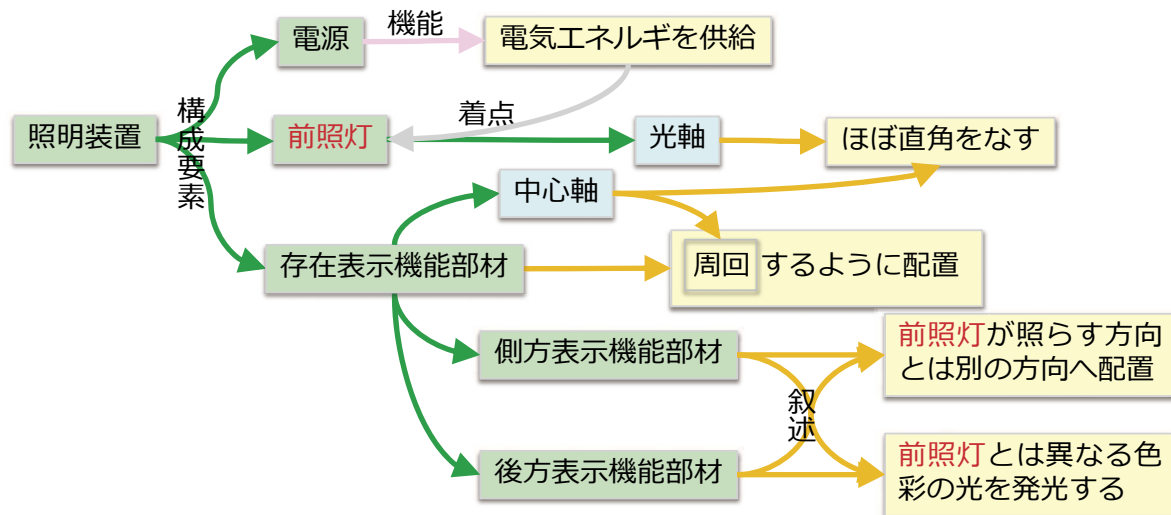
4

特許の請求項

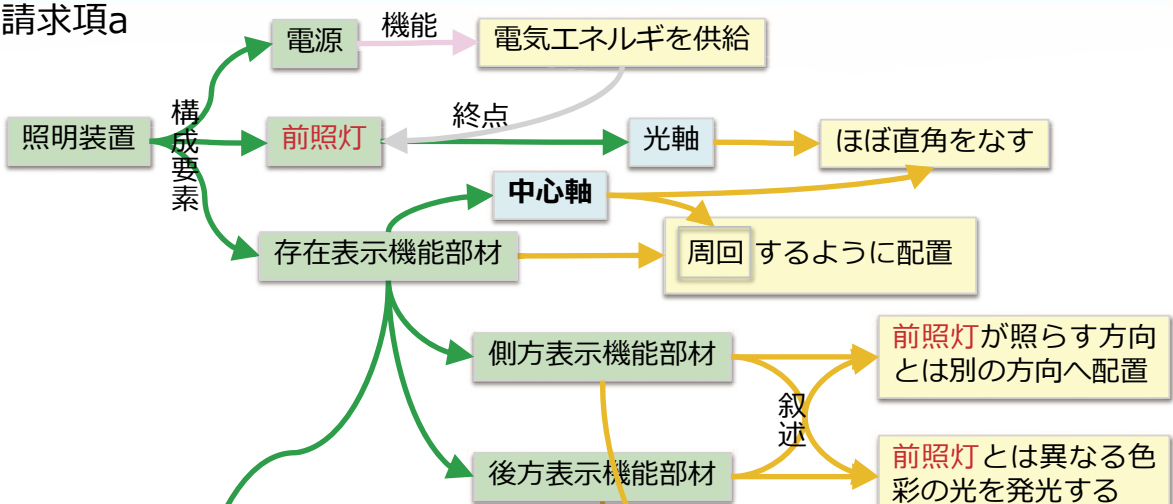
電源と、その電源から電気エネルギーを供給される前照灯と、その前照灯の光軸とほぼ直角をなす中心軸を周回するように配置された存在表示機能部材と、を備え、

前記の存在表示機能部材は、前記の前照灯が照らす方向とは別の方向へ配置された側方表示機能部材および後方表示機能部材を備え、

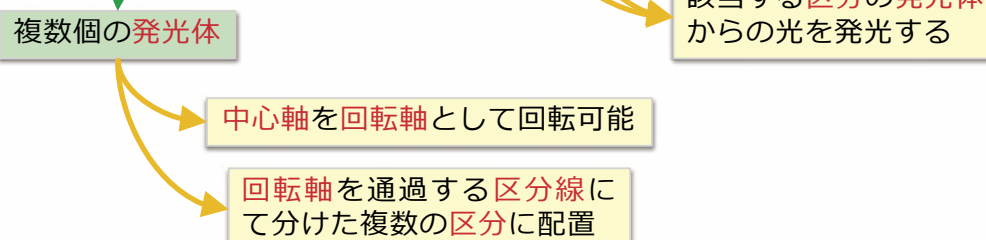
前記の側方表示機能部材および後方表示機能部材は、前記の前照灯とは異なる色彩の光を発光するように形成した照明装置。



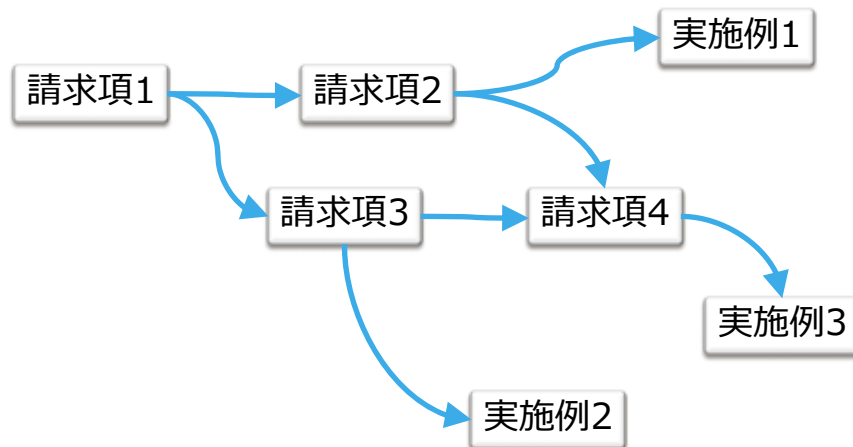
請求項a



請求項b



請求項と実施例の間の派生関係



文書の意味構造の明示

- 第一義的には人間のため
 - ◆ 人間同士の意味共有と合意形成の精度と効率 ↑
 - 社会全体の生産性 ↑
- 機械のためにもなる
 - ◆ データの1次利用
 - * 個票データを1回のサービスで活用
 - ◆ データの2次利用
 - * 機械学習等
- 人間と機械の間の意味共有と合意形成

「産業日本語研究会・特許文書分科会活動報告」

谷川 英和

IRD 国際特許事務所 所長・弁理士／

産業日本語研究会 特許文書分科会 主査

特許文書分科会

2018/3/5

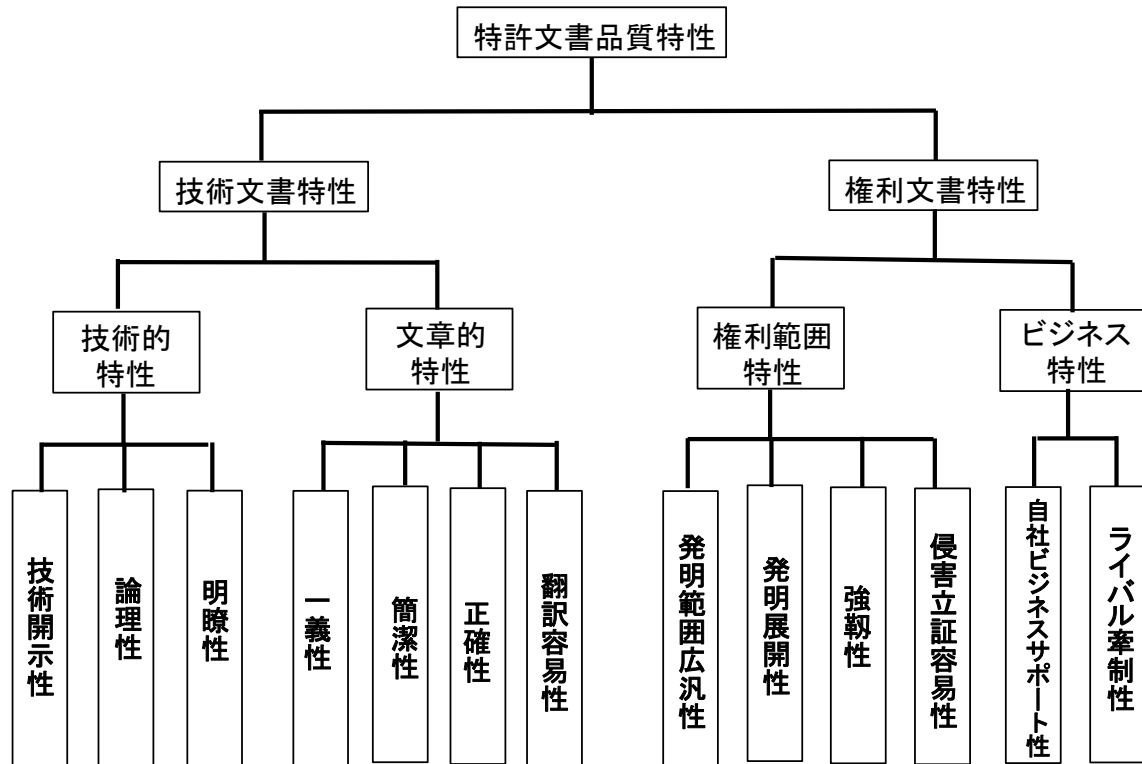
1

(1) 取り組み内容

特許文書の品質とは？

2

(2) 特許文書品質特性



(3) 特許文書品質特性の定義

品質特性			品質特性の説明
大分類	中分類	小分類	
技術文書特性	技術的特性	技術開示性	発明が十分に説明されている度合い 特許法が要求しているサポート要件、実施可能性を満足している度合い
		論理性	論理的である度合い
		明瞭性	発明の説明が明瞭である度合い
	文章的特性	一義性	多義的に解釈できない度合い
		簡潔性	簡潔な文の度合い
		正確性	誤記・技術的誤りが出現しない度合い
		翻訳容易性	翻訳しやすい度合い
権利文書特性	権利範囲特性	発明範囲広汎性	発明の本質が抽出されており、無用な限定が無い度合い。権利範囲の広さの度合い
		発明展開性	発明が十分に展開されている度合い。
		強靱性	拒絶、無効になりにくい度合い
		侵害立証容易性	侵害の立証が容易である度合い
	ビジネス特性	自社ビジネスサポート性	自社製品をカバーしている度合い
		ライバル牽制性	ライバルを牽制できている度合い

(4) 特許文書品質(検討事項)

各特性の ①定性的な評価方法 ②定量的に評価する場合のパラメータ(数, 割合)

小分類	品質特性の説明	評価方法(定性的)	パラメータ(定量的)	
			数量	割合
技術開示性	発明が十分に説明されている度合い。	・特許請求の範囲において、全ての実施形態が含まれるように記載されているか？ ・請求項に記載の発明特定事項(発明の本質)について、明細書中に十分な記載があるか？	・実施の形態の数、実施例の数 ・図面数	・実施例でサポートされている請求項の割合 ・請求項の数値範囲をカバーするデータの割合
論理性	論理的である度合い	・背景・従来技術・課題・解決手段・効果のストーリーの筋が通っているか？ ・明細書の文章全体の構造が明確に把握できるか？	・クレームツリーの枝にぶら下がっている請求項に番号飛びが存在する場合の飛びの数	・発明特定事項と効果の組が記載されている割合 ・請求項の発明特定事項の例示の記載率
明瞭性	発明の説明が明瞭である度合い	・一般名称、専門用語が正しい意味で使用されているか？ ・造語や略語は、定義が記載されているか？	・造語や略語の定義の数	・請求項の造語の定義率 ・請求項の発明特定事項の例示の記載率
一義性	多義的に解釈できない度合い	・定義の無い多義的な用語・文が使用されていないか？	・多義的な用語の数 ・多義的な文の数	・単文率または複文率 ・能動態または受動態の使用率
簡潔性	簡潔な文の度合い	・用語の統一性が図れているか？ ・冗長用語、冗長表現(例: 制御を実行する)が使用されているか？	・長文、複文の数 ・重複記載の箇所の数	・冗長用語、冗長修飾詞の使用率 ・能動態の使用率
正確性	誤記・技術的誤りが出現しない度合い	・用語が統一されているか？ ・明細書中の符号と図面の統一が統一されているか？	・誤記、誤字、脱字の数 ・技術的誤りの数	・用語の統一率 ・誤記、誤字、脱字の発生率
翻訳容易性	翻訳しやすい度合い	・主語があるか？ ・長文や複文が多いか？	・主語の無い文、長文、複文の数 ・助詞の誤使用の数	・長い複合詞の使用率 ・日本語特有の不明確になり易い用語の使用率
発明範囲広汎性	権利範囲の広さの度合い	・上位概念の用語が用いられているか？ ・明細書中に、発明特定事項の定義が限定的に記載されていないか？	・独立項の発明特定事項数、文字数、発明特定事項に対する修飾数、格成分数 ・1つの請求項についての課題や効果の数	
発明展開性	発明が十分に展開されている度合い。	・独立項が多いか？(独立項が多いと、訂正要件を満たす可能性が高くなる。) ・クレームに包含される例が豊富であるか？	・クレーム数 ・クレームツリーのネストレベル(深さ)	
強靱性	拒絶、無効になりにくい度合い	・クレーム文言が「明瞭でありクレーム文言と実施例との間に齟齬がないか？ ・クレームに包含される例が豊富であるか？	・背景技術の先行文献数 ・実施例数	
侵害立証容易性	侵害の立証が容易である度合い	・発明特定事項を実施する主体が複数になっていないか？ ・クレームに包含される例が豊富であるか？	・格成分数	
自社ビジネスサポート性	自社製品をカバーしている度合い	・自社製品、自社サービスに対応した特許請求の範囲、図面、フローチャートが存在するか？ ・収益を守るための請求項が記載されているか？	・実施報償に関する情報、製品の売上情報等の自社固有の管理情報	
ライバル牽制性	ライバルを牽制できている度合い	・他社製品・他社サービスと同一または近似した特許請求の範囲、図面が存在するか？ ・他社製品・他社サービスと同一または近似した用語が使用されているか？	・契約情報(ライセンス契約を締結している等)	

(5) 今後

- (1) 評価方法(定性的)、パラメータ(定量的)の検討
- (2) 評価目的・評価シチュエーション別の重要度の検討
- (3) 特許文書品質の普及に向けた方策の検討

「特許ライティングマニュアル（改訂版）紹介」

白土 博之

一般財団法人日本特許情報機構

特許情報研究所 調査研究部研究企画課 課長

産業日本語「特許ライティングマニュアル」を改訂します！

産業・技術情報を人に理解しやすく、かつ、機械にも処理しやすいものとする「産業日本語」を実現するための「特許ライティングマニュアル」の見直しを行い、以下7つのテーマにまとめました。

テーマの再構成と、例文の追加・修正を行いました。

テーマ名及びその説明文から、概要を理解できるようにしました。

【初版目次】

- A 異文「分りにくい」を避ける。
(ルール 章A条の1～4)
- B 不整列「雑然とした並び」を避ける。
(ルール 章B条の1～4)
- C 省略「語とし穴」を避ける。
(ルール 章C条の1～5)
- D 多義「どの意味なの？」を避ける。
(ルール 章D条の1～3)
- E 非論理的「番号がない」を避ける。
(ルール 章E条の1～7)
- F 非均質並立表現「ばらつき」を避ける。
(ルール 章F条の1～2)
- G 冗長「贅言しすぎ」を避ける。
(ルール 章G条の1～4)
- H 難解「？？」を避ける。
(ルール 章H条の1～2)



【改訂版目次】

文レベル	1. 短文にする	• 一文を短くして複雑な係り受けをなくすることで、人間の理解が容易になり、機械翻訳時の誤訳を減少できる。ただし、短文化する際には、意図しない内容とならないよう注意する。
節・句レベル	2. 省略しない	• 主語や目的語等がないと文の内部構造が不明になり、誤解を生じたり、外国語に訳しにくい。主語や目的語を明示して、文の構造を明らかにする。
	3. 表現を統一する	• 同じ意味の表現を使い、統一することで、読み手が理解しやすくなる。
	4. 構文を明確にする	• 構文が不明な表現は、構文を明確にする。
	5. 表現を簡潔にする	• 冗長な表現を避け、簡潔な表現を用いる。
	6. 簡潔に	• 冗長な表現を避け、簡潔な表現を用いる。
語レベル	7. 言い換える	• 多義的な表現やあいまいな表現は、明確・具体的な表現に改める。また、日本語独特の表現は、訳しやすい表現に改める。

ポスターにて
紹介します！

※上記の内容は今後変更される可能性があります。



「システム開発文書品質研究会（ASDoQ）の
取り組み紹介」

塩谷 敦子

合同会社イオタクラフト 代表社員／

システム開発文書品質研究会 幹事

あすどっく システム開発文書品質研究会
(ASDoQ : Association of System Documentation Quality)

■ 設立 2011年7月11日

■ 目的

1. 開発文書の品質の定義
2. 開発文書の品質を測る方法の研究
3. 開発文書の品質を向上させる方法の研究

■ 活動

- 大会(1回/年), ワークショップ(1回/年), 研究会(3回程度/年)
2017大会テーマ: 文書品質にとって技術とはなにか〜文書品質から開発技術を見直そう〜

■ 特徴

- 参加費無料
- 会員の多様性. 業務や専門領域は異なるが, 全員が, 文書品質に関心を持つ
- 会員数

- ・ 法人: 20社
- ・ 個人: 157名

(2017.11現在)

代表幹事	山本雅基	名古屋大学
アドバイザー	高田広章	名古屋大学
アドバイザー	田丸喜一郎	IPA/SEC
アドバイザー	山本修一郎	名古屋大学
アドバイザー	清水吉男	(株)システムクリエイツ
幹事, 事務局長	藤田悠	長野工業高等専門学校
幹事	栗田太郎	ソニー(株)
幹事	小林直子	エプソンアヴァシス(株)
幹事	塩谷敦子	(同)イオタクラフト
運営委員	粕渕清孝	(株) SCREENアドバンスシステムソリューションズ
運営委員	奈良慶之	(株)ベリサーブ
運営委員	森川聡久	(株)ヴィッツ
会計監査	中澤達夫	長野工業高等専門学校

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

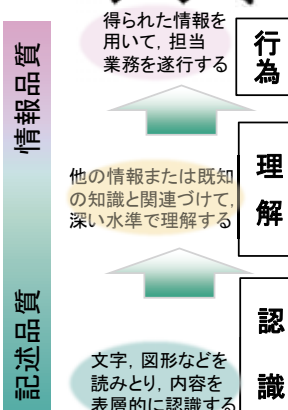
<E-mail> secretariat@asdoq.jp (事務局)

1

システム開発文書品質の定義

読み手が開発文書を取り扱う段階に応じて, 品質特性を定義.

品質モデルの
利用シーンを
ポスターにて
紹介しています



システム開発文書品質モデル(Ver.1.0a)

品質特性	説明	品質副特性	測定項目 (例)
(1)完全性	開発に必要な十分な情報が記載されていること	合目的 正確 妥当	読み手の明示 技術が正しく実現可能 許容範囲で開発可能
(2)論理性	論理的に整合がとれていること	無矛盾 一貫 構造	項目間の矛盾がない 理由または根拠の明記 階層が合理的
(3)理解容易性	理解しやすいこと	非曖昧性 関係	複数の解釈が成り立たない 参照先を正しく明示
(4)可読性	読みやすいこと	簡潔 統一 表記工夫	一文一義 視点の統一 箇条書きを適切に使用
(5)規範適合性	記述が文法や規則に則していること	文法適合 記法適合 基準適合	誤字脱字がない 箇条書きのルールに適合 テンプレートに準拠

■ システム開発文書品質モデルのシート (ポスター展示エリアにて配布中, ダウンロード可能)

➢ <https://asdoq.jp/research.html>

■ 品質モデル概要と試行した課題の掲載論文

➢ 「システム開発文書品質モデル活用への取組み」 情報処理学会デジタルプラクティス Vol.8 No.2 (Apr. 2017)
<http://www.ipsj.or.jp/dp/dp-index.html> (閲覧無料)

2

「文章読解・作成能力検定（呼称「文章検」）」

八田 香里

公益財団法人日本漢字能力検定協会 常任理事

●公益財団法人 日本漢字能力検定協会について

日本語・漢字を学ぶ楽しさを提供し、豊かな社会の実現に貢献

公益財団法人日本漢字能力検定協会は、内閣府より認定を受けた公益財団法人です。

日本語・漢字に関する「普及啓発・支援活動」「調査・研究活動」「日本語能力育成活動」を3つの柱とし、社会生活に必要な日本語・漢字の能力の向上と、広く日本語・漢字に対する尊重の念と認識を高める活動に邁進しています。

普及啓発・教育支援活動

- ・博物館・図書館の運営
- ・漢字・日本語キュレーション
サイト・検索サイトの運営
- ・機関誌の発行
- ・講演会・講座の開催
- ・漢字文化啓発イベント
- ・教材の発行 など



「今年の漢字®」

調査・研究活動

- ・漢字文化研究
- ・研究論文の顕彰
- ・教育研究への助成 など



「漢字ミュージアム」

日本語能力育成活動

- ・検定・テストの実施



「漢字カフェ」



公益財団法人 日本漢字能力検定協会

THE JAPAN KANJI APTITUDE
TESTING FOUNDATION

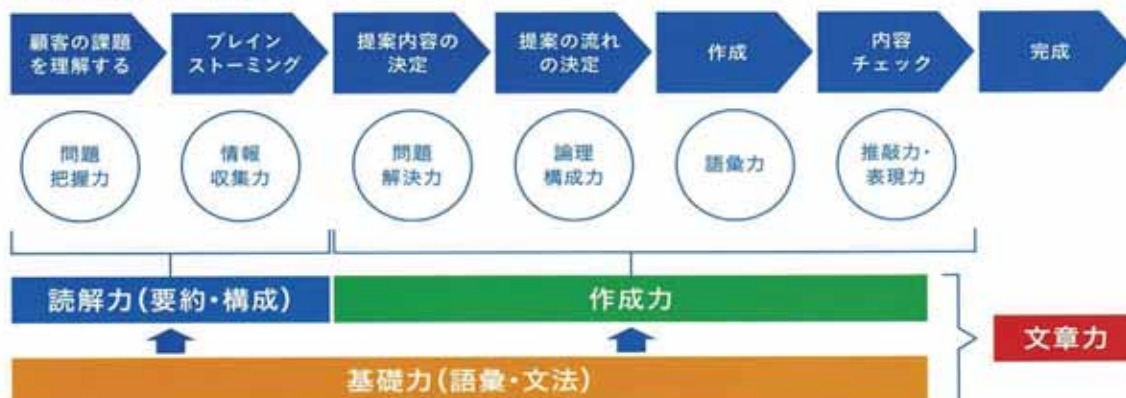
●文章検のコンセプト

文章は相手に自分の思いや考え、あるいは、重要な情報を伝達する手段です。的確に伝えるには、相手や場面を考えて、最も効果的な表現を考える必要があります。

「聞く」ことも「話す」こともコミュニケーションですが、コミュニケーションのトレーニングには文章を書く練習が効果的です。

「文章検」は「文章能力」を「基礎力」「読解力」「作成力」の3つの力に分類し、それぞれの知識・能力を測定します。

【例】提案書を作成する



公益財団法人 日本漢字能力検定協会

THE JAPAN KANJI APTITUDE
TESTING FOUNDATION

●文章検・各級の社会人活用例

	対象者例	課題例
準2～2級	営業・管理部門・管理職	社内外に対する論理的な企画書・決裁書が求められる。コミュニケーションの質と効率を高めたい。部下の指導力を高めたい。
3級	支店長・SV・エンジニア	複数の情報をとりまとめて本部などに報告する際、冗長でポイントがわかりにくい。説得力のあるまとまった書面を作成したい。
4級	接客スタッフ・技能工・事務一般職	引き継ぎ・改善案など、短い業務日報を起票するが、起きた事象と主観が混在してわかりにくい。簡潔で正確な報告書を作成したい。

●文章検2級・社会人結果データ

受験者数	合格者数	合格率
1043	505	48.4%

N=41社（2017年2月～12月）

	文章読解（input）			文章作成（output）		合計
	問 1	問 2 - 1	問 2 - 2	問 3	問 4	
配点	30	10	30	50	80	200
平均点	24.2	9.3	21.9	37.7	40.9	134.1
得点率	80.7%	93.4%	72.9%	75.4%	51.2%	67.0%

文章検

漢検

公益財団法人 日本漢字能力検定協会

THE JAPAN KANJI APTITUDE
TESTING FOUNDATION

「みんなでニューラル機械翻訳（NMT）を育てよう！
～「みんなの自動翻訳@TexTra®」、「翻訳バンク」～」

隅田 英一郎

国立研究開発法人情報通信研究機構
先進的音声翻訳研究開発推進センター
副研究開発推進センター長

みんなで ニューラル機械翻訳(NMT)を 育てよう！

～ 「みんなの自動翻訳@TexTra®」、「翻訳バンク」～

NICT
隅田 英一郎

2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

1

「みんなの自動翻訳@TexTra®」
<https://mt-auto-minhon-mlt.ucri.jgn-x.jp>



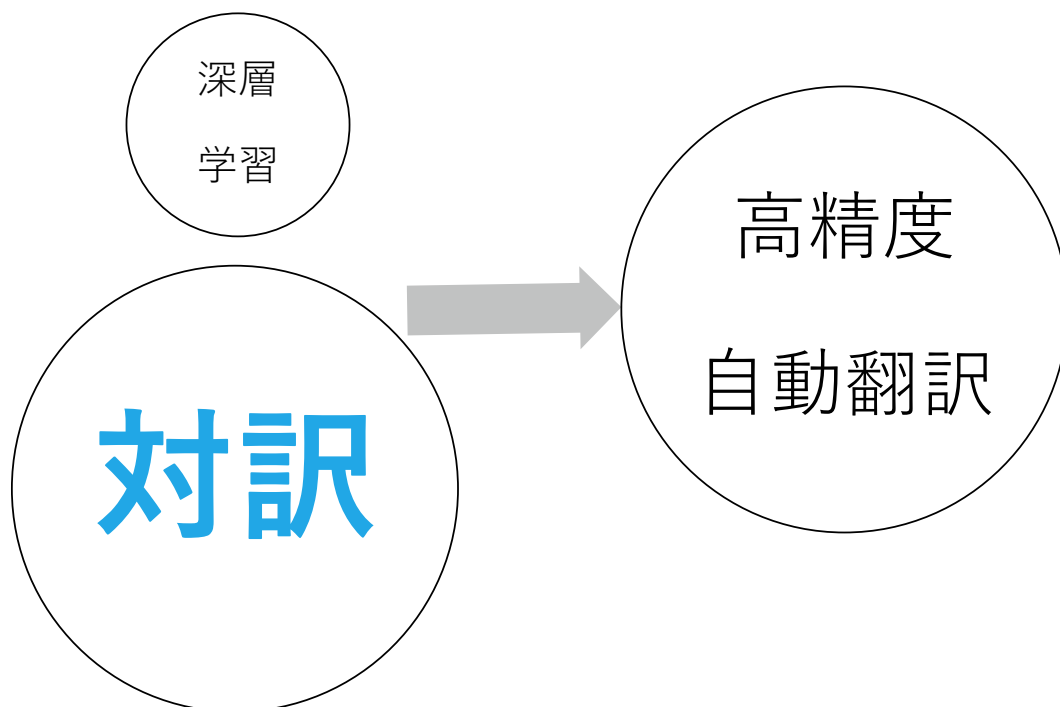
高精度の
ニューラル機
械翻訳(NMT)
が使えます

2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

2

どんな文でも高精度に翻訳するNMTをみんなで作る方法



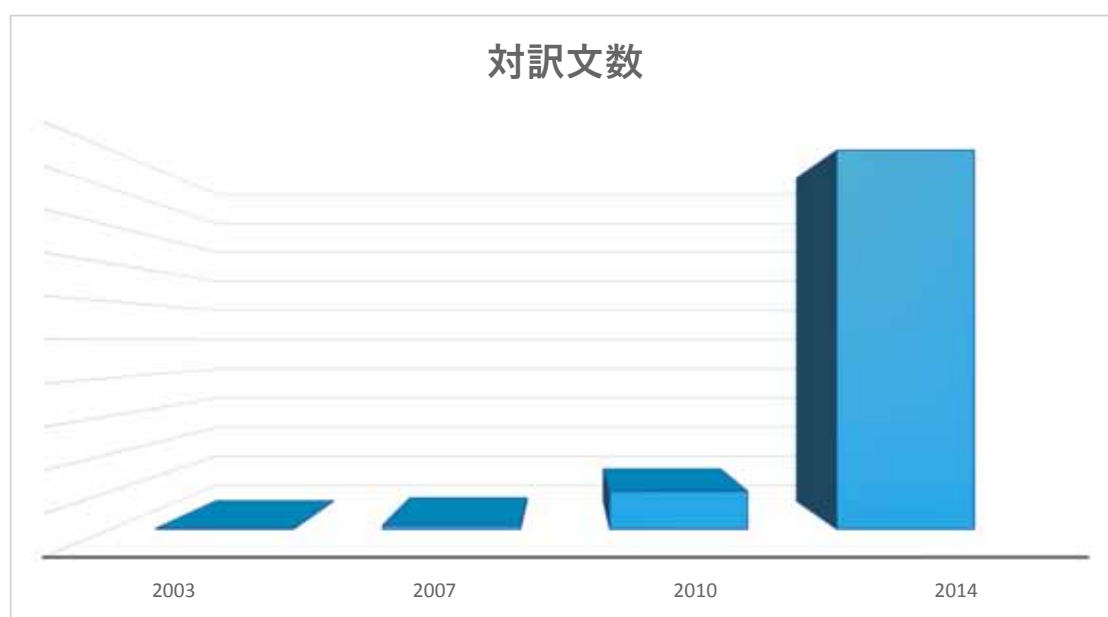
2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

3

対訳データの指数関数的成長

「対訳文数は4年で10倍」 @NICT

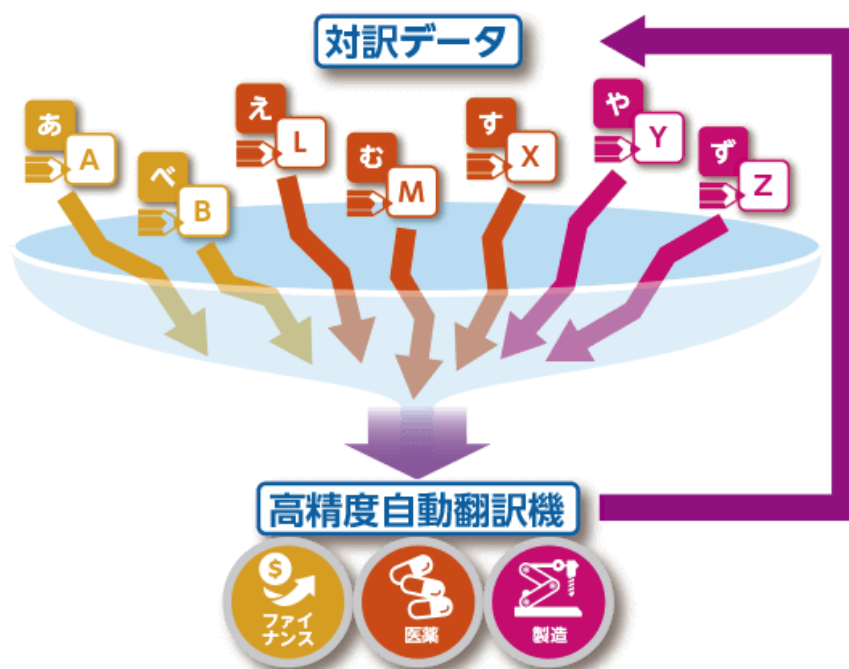


2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

4

「翻訳バンク」 <http://www2.nict.go.jp/ais/h-bank.html>



2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

5

1 番簡単な方法と2番目に簡単な方法

～webサイトの利用～



～翻訳データ提供に関する
契約の締結～

情報提供に関するご承諾依頼について（案）

国立研究開発法人情報通信研究機構（以下「弊機構」という。）は、弊機構が進める「多言語音声翻訳の研究開発」（以下「本研究開発」という。）に関連し、〇〇〇〇〇〇〇〇〇〇〇〇（以下「△△△△」という。）から提供を受ける情報について以下の事柄をご承諾頂きたく、茲許ご依頼申し上げます。

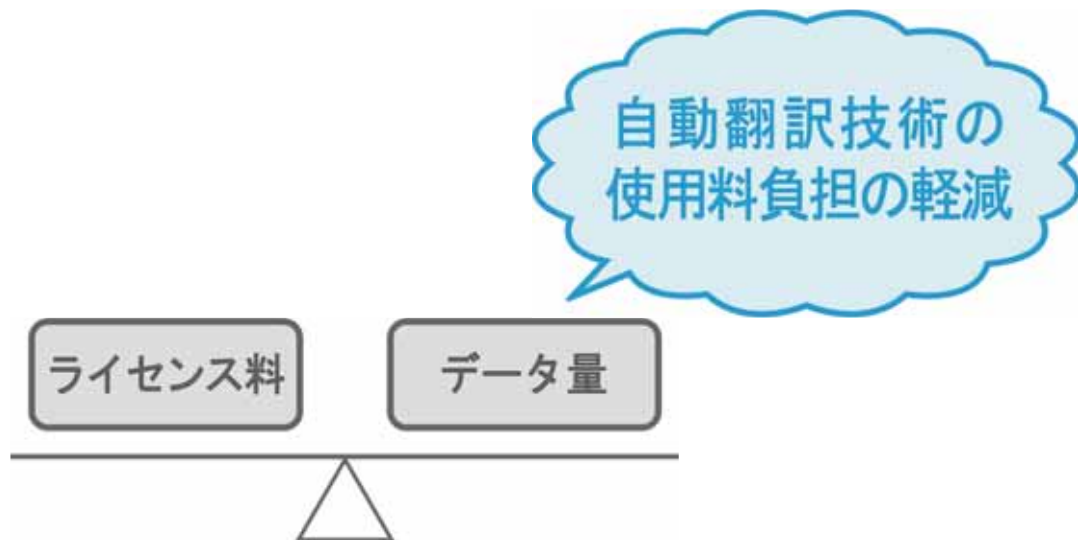
2018/3/5 14:40-16:00

【第二部】 伝わる文書の作成に向けて

6

3 番目の方法

～自動翻訳技術使用ライセンス契約に伴う
翻訳データの提供～



NO NMT,
NO LIFE.

「Japio 世界特許情報全文検索サービス紹介」

高橋 幸生

一般財団法人日本特許情報機構

営業推進部 部長



2018年3月5日



Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio世界特許情報全文検索サービス (Japio-GPG/FX)概要

Japio世界特許情報全文検索サービス(Japio-GPG/FX)は、
日本、米国、欧州、中国、韓国、独国、仏国をはじめとして、
 Japioが特許公報を収集した国や機関の
公報全文、および、DOCDBなどに含まれる
 世界の特許情報を、
機械翻訳技術を活用して**日本語と英語**で
横断的に一度に検索できるようにしたサービスです。



Japio-GPG/FX概要 1. 蓄積データ

収録国	蓄積範囲 (公報発行年)	使用データ ※データ種類により蓄積期間は異なります。
中国(CN)	1985～	中国公報データ、中国公開特許と文抄録テキストデータ、機械翻訳データ、代表図面、全図面、DOCDB、FI・Fターム
日本(JP)	1983～	日本公開特許公報データ、公表公報データ、日本特許公報データ、日本実用新案公報データ、機械翻訳データ、代表図面、全図面、PAJ、DOCDB、整理標準化データ(FI・Fターム)
アメリカ(US)	1976～	米国出願公開特許データ、米国登録特許データ、米国公開特許と文抄録データ、米国特許と文抄録テキストデータ、機械翻訳データ、代表図面、全図面、DOCDB
欧州(EP)	1978～	欧州特許公開出願公報データ、欧州登録特許公報データ、欧州公開特許と文抄録テキストデータ、機械翻訳データ、代表図面、全図面、DOCDB
PCT(WO)	1978～	PCT国際出願データ、機械翻訳データ、代表図面、DOCDB
韓国(KR)	1999～	韓国公報、機械翻訳データ、代表図面、DOCDB
ドイツ(DE)	1976～	DOCDB、代表図面、機械翻訳データ(DE公報独英、DE公報独英→日MT)
フランス(FR)	1976～	DOCDB、代表図面、機械翻訳データ(FR公報仏英、FR公報仏英→日MT)
その他 (BR、RU、IN等) 88か国・地域、4機関	1976～ (注:国により異なります。)	DOCDB

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX概要 2. 検索項目

- 国、公報種別 : 特許・実用新案、公開・登録
- 番号 : 出願番号、文献番号、優先権主張番号
- 日付 : 出願日、公報発行日、優先権主張日
- 分類 : IPC、FI・Fターム(日本のみ)、CPC、ECLA
- 出願人 : 出願人(日本語・英語・中国語・他)
- 発明者 : 発明者(日本語・英語・中国語・他)
- 技術用語 : 発明の名称、要約、クレーム、
詳細な説明(日本語・英語・中国語・他)

* 利用可能制限・範囲あり

Japio-GPG/FX概要 3. 主な機能

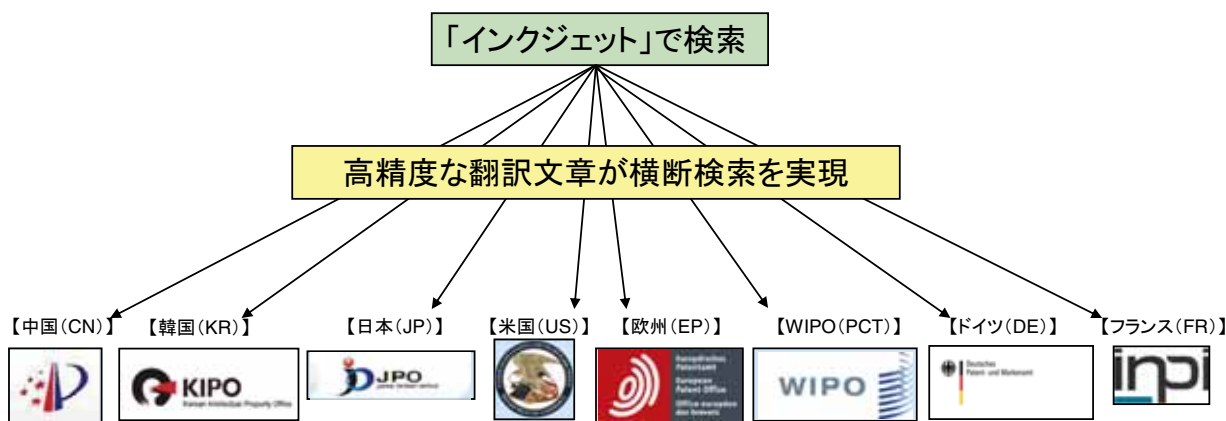
- ダウンロード機能
 - ⇒公報PDF、書誌・テキストダウンロード、サマリーダウンロード
- しおり機能
 - ⇒最大500件まで可能
- 検索式保存機能
 - ⇒最大300件まで可能 コメント入力も可能
- 用語検索支援機能
 - ⇒Japio独自の同義語テーブルが利用可能
 - ⇒英語・日本語の異表記を参照可能
- 出願人検索支援機能
 - ⇒日本語・英語・中国語表記の参照可能

5

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX特徴 1. 横断検索

- 同一検索式で主要国(中国・韓国・日本・米国・欧州・WIPO(PCT)・ドイツ・フランス)の特許情報(全文)を日本語、英語(中国・韓国を除く)で一括でテキスト検索(横断検索)が可能です。



6

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX特徴 2. 翻訳(英⇒日)

■ 検索用の日本語は、**約1億文対のコーパス**を使用し、
高品質の英日機械翻訳文を作成いたしました(SMT)。

Japio-GPG/FX特徴 3. 翻訳(中⇒日)

■ 検索用の日本語は、**約5,200万文対のコーパス**を使用し、
高品質の中日機械翻訳文を作成いたしました(SMT)。

7

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX画面

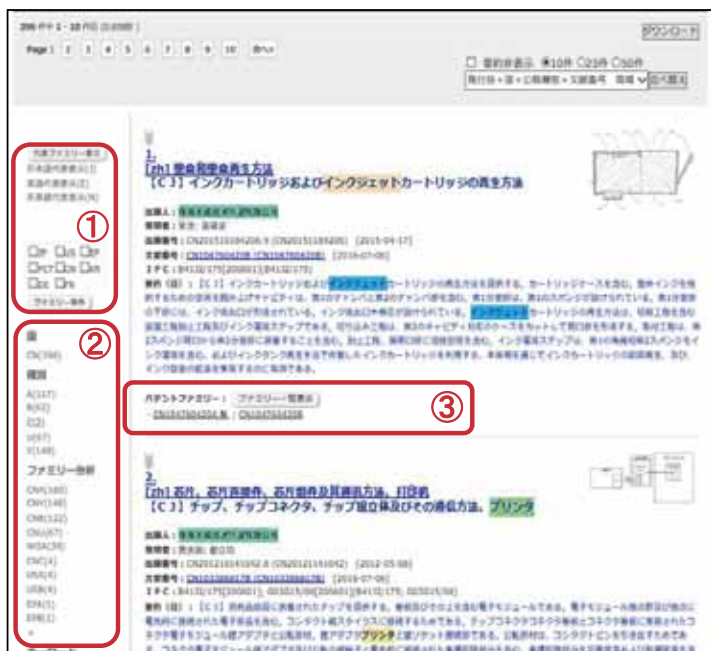
1. 検索画面

- ① 検索対象国を簡単に選択
- ② 検索したい用語を日本語・英語で入力
 - ・日本語と英語の混在可能
 - ・AND、OR検索可能
- ③ 異なる分類を入力
 - ・分類(IPC、ECLA、CPC、FI、Fターム)
 - ・OR検索可能
- ④ 詳細検索入力も用意

8

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX 画面 2. 一覧表示



①パテントファミリー文献のうち、代表となる文献のみを表示する機能とファミリーから除外する機能

②ファセット機能で1次検索後の絞り込みが容易

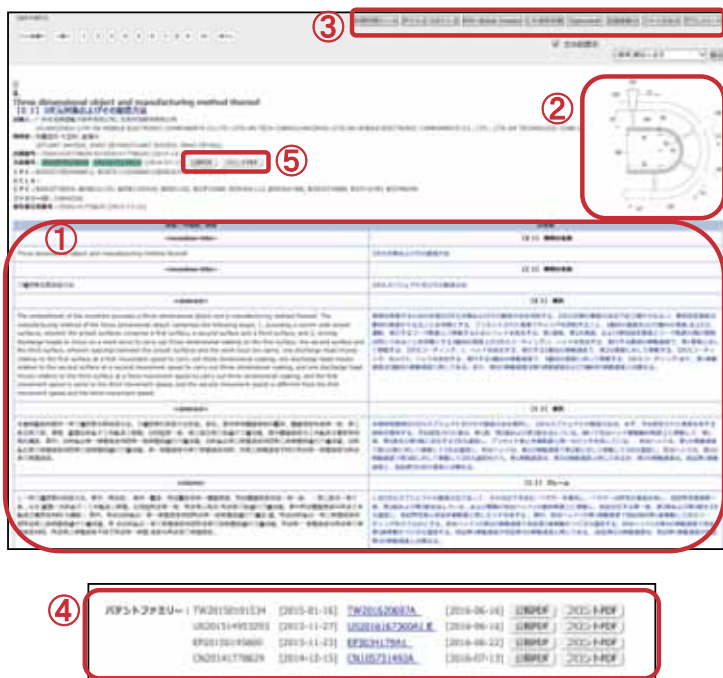
③各国ファミリーの確認も容易

※その他、ハイライト、公報リンク機能などを実装

9

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

Japio-GPG/FX画面 3. 詳細表示



①発行国言語(英語、中国)、日本語を全文対訳表示

②代表図表示

③JPT/JE、UST/JE、全図面表示等

④パテントファミリー表示

⑤公報PDF出力機能

10

Copyright (C) 2016, Japan Patent Information Organization & Hatsumei-Tsushin Co., Ltd. All Rights Reserved

【第三部】

A I 活用時代の日本語データの 高度活用事例

招待講演

「法制執務業務の効率化に向けて（e-LAWS^{イーローズ}の整備）」

法制執務業務は、国会対応や予算関連業務と並ぶ国家公務員の業務効率化の重要な鍵。

「霞が関で働く女性有志」の提言を踏まえ、業務省力化・平準化の観点から、ICTを活用し、一連の法案等作成業務を支援する「法制執務業務支援システム（e-LAWS^{イーローズ}）」を総務省行政管理局において整備。今後、法案等作成業務の省力化や業務の平準化が期待されるところ。

また、民間ビジネス等での活用を想定したオープンデータ化を推進するため、「法令標準XMLスキーマ」を策定し、このスキーマを活用して整備した法令データを平成29年6月からインターネットでの公開を開始。

本講では、e-LAWSの仕組み、整備に当たって苦労した点、今後の展望等について説明します。

小高 久義

総務省 行政管理局行政情報システム企画課

情報システム管理室長

第9回産業日本語研究会・シンポジウム

法制執務業務の効率化に向けて (e-LAWSの整備)

平成30年3月5日

総務省 行政管理局 行政情報システム企画課
情報システム管理室長 小高 久義

はじめに（総務省行政管理局での法令DBの整備経緯）



総務省組織令（平成十二年政令第二百四十六号）

（行政管理局の所掌事務）

第五条 行政管理局は、次に掲げる事務をつかさどる。

一～三 （略）

四 行政機関が共用する情報システムの整備及び管理に関すること。

五～十一 （略）

年 月	経 緯
昭和54年 4月 (1979年)	法令検索システム（各府省への提供） ※ホストコンピュータ
平成11年 1月 (1999年)	霞が関WAN（現：政府共通ネットワーク）経由での提供 ※ホストコンピュータ → サーバ
平成13年 4月 (2001年)	法令データ提供システム（総務省が整備した法令データの公開） ※e-Gov（インターネット）経由で提供
平成28年10月 (2016年)	e-LAWS（法制執務業務支援システム）運用開始 ※「霞が関で働く女性有志」の提言を踏まえ、システム整備
平成29年 6月 (2017年)	所管府省が確認・認証した法令データの公開 ※e-LAWSで整備した法令データをe-Gov（インターネット）経由で提供

1 「霞が関で働く女性有志」の提言



持続可能な霞が関にむけて（抜粋）（平成26年6月 霞が関で働く女性有志）

（提言7）法律等立案作業の改善

- ▶ 直接携わる者の範囲は広くないものの、**法律等立案作業は、深夜・早朝・休日まで及ぶことの多い非常に負担の大きい業務の一つ**です。特に法案提出に向けた準備が集中する年末～3月頃にかけては、法制局審査等により連日深夜に及ぶ多大な負担がかかっています。
- ▶ 最終的な法律等の条文については、政策担当者自らが検討することが必須であり、また、国民の権利義務等へ直接的影響を与えるものであることから、**最大限の精査を欠かすことができません**。

出典： <http://www.cas.go.jp/jp/gaiyou/jimu/jinjikyoku/files/teigen1.pdf>

2

2 法令改正作業（従前）



【現行条文の確認】

改正対象の現行法令を確定するためには、現行日本法規（注）と読み合わせを行い、内容の正確性の確認が必要。

（注）法務省大臣官房司法法制部が国の事業として編集。

現行日本法規にも反映されていない最近の改正情報は官報で補っている。

例えば、「特定商取引に関する法律」を改正する場合、91ページに渡る条文から改正箇所を確認する必要



【参照する法令の収集】



法案作成では、関係する条文をまとめて法令集を作成し、推敲にあたって参照する（参照条文）。

改正対象の現行法令と同様に、正確性が要求されるため、原本として現行日本法規を手作業でコピー、編集を行っていた。

（例）

改正に関係する100法令分の該当冊子、ページ特定

⇒ ヒモを外し、B5の黒本を拡大コピー

⇒ 改正後未施行の条文がある場合には、コピーした紙にその旨を記述

⇒ これを切り貼りして「参照条文」の出来あがり

※この作業に約5時間半かかった法案も

3

2 法令改正作業（従前）



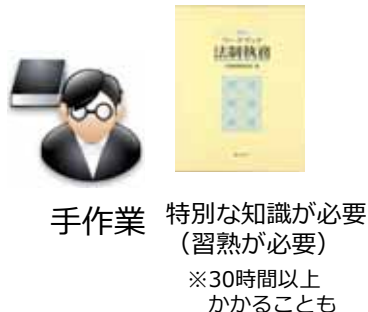
【新旧対照表の作成】

確認した現行条文を下欄（現行）に
改正条文を上欄（改正案）に記載した
新旧対照表を作成

正確な電子データがないと
作成に時間がかかる

【法案（改め文）の作成】

新旧対照表を基に法案（改め文）
を作成



児童扶養手当法（昭和三十六年法律第二百三十八号）の一部を次のように改正する。
第一条中「父」の下に「又は母」を加え、「生活」を「生活等」に改め、「もつて児童の」を削る。
第二条第二項を削り、同条第三項を同条第二項とする。

「改める文」については、いわば改正のための「道具」ですが、その独特のルール等ゆえ、作成・審査に莫大な労力を要しています。

4

3 法令改正（改め文）のしくみ



一部改正法令（改め文）

公布：平成13年6月22日
施行：平成15年1月 1日

法律第五十九号
国民の祝日に関する法律及び老人福祉法の一部を改正する法律
第一条 国民の祝日に関する法律（昭和二十三年法律第七十八号）の一部を次のように改正する。
第二条 海の日の項中「七月二十日」を「七月の第三月曜日」に改め、同条敬老の日の項中「九月十五日」を「九月の第三月曜日」に改める。

国民の祝日に関する法律

【現行（平成13年6月22日時点）】

第二条 「国民の祝日」を次のように定める。
元日 一月一日 年のはじめを祝う。
・・・
海の日 七月二十日 海の恩恵に・・・
敬老の日 九月十五日 多年にわたり・・・
・・・

【改正後（平成15年1月1日時点）】

第二条 「国民の祝日」を次のように定める。
元日 一月一日 年のはじめを祝う。
・・・
海の日 七月の第三月曜日 海の恩恵・・・
敬老の日 九月の第三月曜日 多年に・・・
・・・

5

4 「改め文」とは？



対象となる法令のどの部分をどのように改めるかを、次の6種類の動詞を利用して逐語的に記述（逐語的改正方式）。

改正点が明確であり、かつ簡素に表現できるというメリットがあるため、法改正の方法として定着。

1	改める	第五条中「公団」を「機構」に改める。
2	加える	第四十七条中「理事」の下に「、経営管理委員」を加える。
3	削る	第十七条を削る。
4	繰り上げる 繰り下げる	第八項から第十一項までを一項ずつ繰り上げる。 第九号から第十一号までを二号ずつ繰り下げる。
5	付する	第十六条の次に次の章名を付する。
6	とする	第六条を削り、第五条を第六条とし、 第四条の次に次の一条を加える。

6

1 「霞が関で働く女性有志」の提言（続き）



持続可能な霞が関にむけて（抜粋）（平成26年6月 霞が関で働く女性有志）

（提言7）法律等立案作業の改善

- ▶ 直接携わる者の範囲は広くないものの、**法律等立案作業は、深夜・早朝・休日まで及ぶことの多い非常に負担の大きい業務の一つ**です。特に法案提出に向けた準備が集中する年末～3月頃にかけては、法制局審査等により連日深夜に及ぶ多大な負担がかかっています。
- ▶ 最終的な法律等の条文については、政策担当者自らが検討することが必須であり、また、国民の権利義務等へ直接的影響を与えるものであることから、**最大限の精査を欠かすことができません**。
- ▶ 一方で、現行の条文と改正後の条文とをつなぐ「改める文」については、いわば改正のための「道具」ですが、その**独特のルール等ゆえ、作成・審査に莫大な労力**を要しています。また、**習熟が容易でない技術であることから、ミスが発生しやすい**実情にもあります。
- ▶ 一時期、この「改める文」の要否について検討がなされた経緯ありますが、改めて検討経緯の確認等を行い、仮に**廃止が困難なのであれば**、各担当における作成は「新旧対照表」（現行の条文・改正後の条文の表）までとし、**「改める文」の作成については、習熟した職員が集中的に処理する体制を設けること等を検討すべき**と考えます。

出典： <http://www.cas.go.jp/jp/gaiyou/jimu/jinjikyoku/files/teigen1.pdf>

7

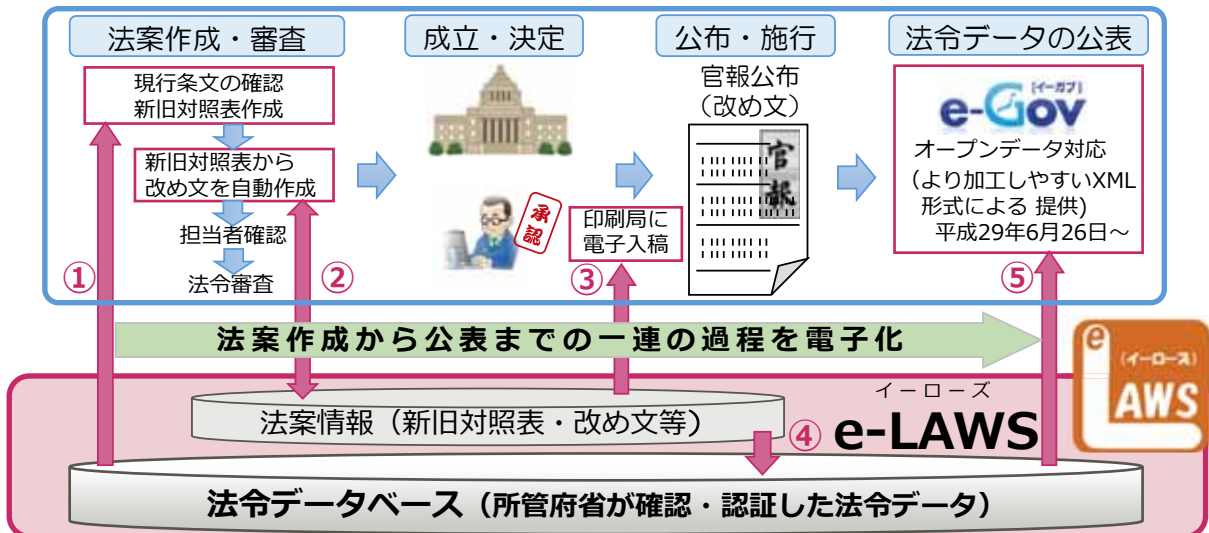
5 「法制執務業務支援システム（e-LAWS）」とは？



「霞が関で働く女性有志」の提言を踏まえ、業務省力化・平準化の観点から、ICTを活用し、法案等関係資料の作成支援等を行うシステムの開発を行うことを決定

総務省行政管理局において、一連の法案等作成業務を支援する「e-LAWS」を開発し、平成28年10月3日から各府省において運用を開始

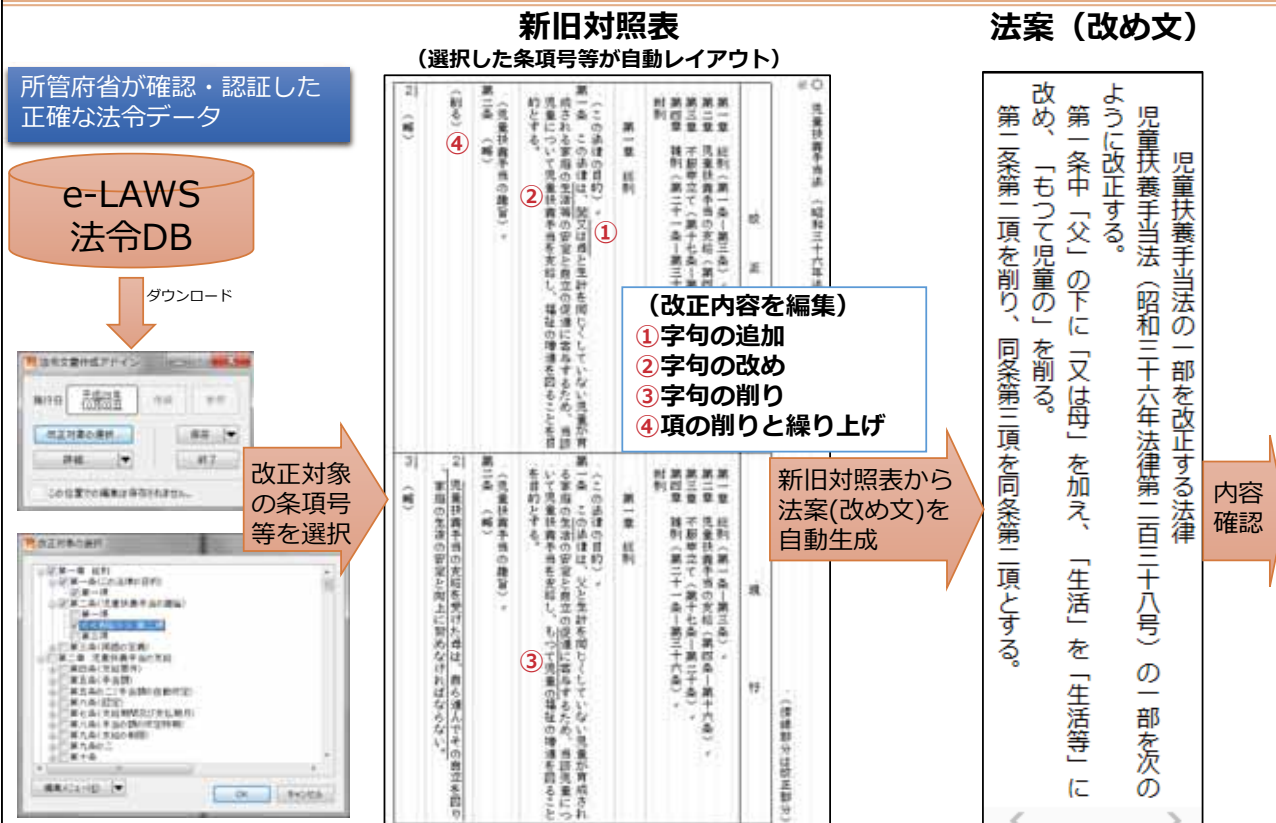
- ①所管府省が確認・認証した法令データを行政及び国民へ提供
- ②新旧対照表から改め文を自動作成するなど、法案担当者の負担を軽減



e-LAWSとは、法制執務業務支援システム(e-Legislative Activity and Work Support System)の頭文字を取った略称

8

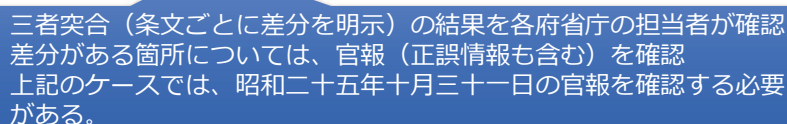
6 e-LAWSを利用した法案等関係資料作成



* 本資料の改正内容は、機能を明示するためのサンプルです。実際の改正内容とは異なります。

9

- 所管府省が確認・認証
した法令データを整備



10

**e-LAWSで作成する
新旧対照表の形式を統一（標準化）**

各府省において、文字数、条文の追加・削りの表現等が異なっていた。

11

9 法令標準XMLスキーマの策定



法令標準XMLスキーマ

要素数：134

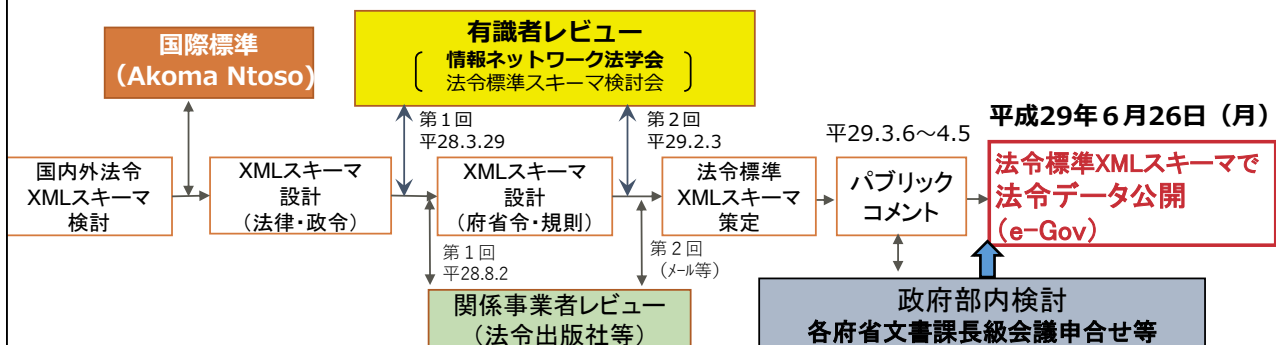
属性数：34

パブリックコメント

<http://search.e-gov.go.jp/servlet/Public>

案件番号：145208896

構造		属性	
日本語	英語	属性名	英語
法令	Law	元号	Era
		年号	Year
		番号	Num
法令番号	LawNum		
法令本体	LawBody		
題名	LawTitle		
本則	MainProvision		
条	Article	番号	Num
項	Paragraph	番号	Num
号	Item	番号	Num



12

XML文書化された法令文書の例



電子政府の総合窓口
[e-Gov]

法令検索 電子申請 行政手続案内検索 パブリックコメント FAQ

条件指定画面へ 検索結果一覧へ 別画面で表示 XML形式ダウンロード

このページのリンク: <http://elaws.e-gov.go.jp/search/>

児童扶養手当法

```
<?xml version="1.0" encoding="UTF-8"?>
<Law Year="36" Num="238" LawType="Act" Lang="ja" Era="Showa">
  <LawNum>昭和三十六年法律第二百三十八号</LawNum>
  <LawBody>
    <LawTitle>児童扶養手当法</LawTitle>
    + <TOC>
    - <MainProvision>
      - <Chapter Num="1">
        <ChapterTitle>第一章 総則</ChapterTitle>
        - <Article Num="1">
          <ArticleCaption>(この法律の目的)</ArticleCaption>
          <ArticleTitle>第一条</ArticleTitle>
          - <Paragraph Num="1">
            <ParagraphNum/>
            - <ParagraphSentence>
              <Sentence>この法律は、父又は母と生計を同じくしていない児童が育成される家庭の生活の安定と自立の促進に寄与するた
                め、当該児童について児童扶養手当を支給し、もつて児童の福祉の増進を図ることを目的とする。</Sentence>
            </ParagraphSentence>
          </Paragraph>
        </Article>
        - <Article Num="2">
          <ArticleCaption>(児童扶養手当の趣旨)</ArticleCaption>
          <ArticleTitle>第二条</ArticleTitle>
          - <Paragraph Num="1">
            <ParagraphNum/>
            - <ParagraphSentence>
              <Sentence>この法律は、父又は母と生計を同じくしていない児童が育成される家庭の生活の安定と自立の促進に寄与するた
                め、当該児童について児童扶養手当を支給し、もつて児童の福祉の増進を図ることを目的とする。</Sentence>
            </ParagraphSentence>
          </Paragraph>
        </Article>
      </MainProvision>
    </LawBody>
  </Law>
```

ダウンロード

13

10 e-Gov法令検索（e-LAWSの法令データ）の公開



報道資料

平成29年6月23日

法令データベース「e-Gov法令検索」のリニューアル公開

電子政府・オープンデータの新たなステップとして、法令データをより使いやすく、より身近なものにするため、6月26日、法令データベース「e-Gov法令検索」をリニューアル公開します。

本データベースは、二次利用が容易なオープンデータとして提供するため、各府省が確認した法令データを標準データ形式(XML形式)で提供するとともに、データを利活用しやすくするためのAPI機能なども提供します。

昨年10月に本格運用を開始した「法制執務業務支援システム」、通称「e-LAWS(イーローズ)」において整備された約8,000以上の法令データ(憲法、法律、政令、府省令、規則)を、「e-Gov法令検索」として公開します。

電子政府・オープンデータ推進につなげていくため、各府省が確認した法令データを二次利用可能な標準データ形式であるXML形式で提供(「法令標準XMLスキーマ」により「文法」を定義し、機械判読可能なデータを提供)するとともに、外部アプリケーションからデータを利活用するためのAPI機能や、バルク機能(XML一括ダウンロード機能)も提供します。

(1) 公開時期: 平成29年6月26日(月)午後

(2) 公開場所: e-Gov(電子政府の総合窓口)

URL: <http://www.e-gov.go.jp/>

(3) 参考資料: 「e-Gov法令検索(e-LAWSの法令データ)の公開」(PDF pptx)

14

二次利用が容易なオープンデータでの提供

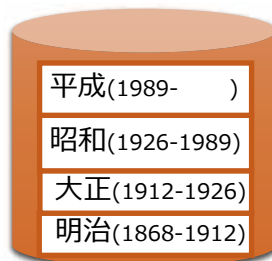


二次利用が容易な形式(XML)で、法令データを提供
さらに、バルクデータ及びAPIの提供により、利用者が必要な法令データを容易に入手可能

【29.4.1時点の全法令】
e-LAWS法令データ
(各府省が確認したデータ)

公開場所: e-Gov(電子政府の総合窓口)
URL: <http://www.e-gov.go.jp/>

憲法、法律、政令、府省令、規則等の現行8000法令について、二次利用を可能とするため法令構造を定義したXML形式でデータ提供



法令の条文 **様式**
XML形式で提供する法令データに加え、府省令・規則等で定めている登録申請様式等を、PDFデータで新たに提供(約10,000ファイル)



検索方法などの
クイックガイドを掲載

① API

外部アプリケーションから直接、法令データを入手できるAPIを提供することにより、必要な法令データを容易に入手可能

② バルクデータ

(XMLを束ねた一塊のデータ)
約8000法令の全てのデータを10年単位で約10分割することにより、10回のアクセスで全法令データを入手可能

二次利用の例

- ・マイ六法
(自分の業務等に必要な法令データだけを集めた法令集)
- ・官報、英文法令、条例等他のデータとの連携
- ・研究用途として活用
構文解析、定義語抽出(目的、罰則規定等)、
条文内容の履歴解析 等

15

11 働き方改革への貢献（テレワーク）



パソコン1台あれば、家からe-LAWSに繋ぎ、
法案作成作業が可能に

必要なもの



必要なもの



パソコン1台のみ



16

おわりに



e-LAWSの開発

- ・ 政府(所管府省)が認証した法令データベースの構築
- ・ 働き方改革に資する法制執務業務の負担軽減

e-LAWSの法令検索をご利用ください。 <http://elaws.e-gov.go.jp/>
意見、要望についても随時受け付けてますので、よろしくお願いします。

今後

法令標準XMLスキーマをe-LAWS以外の官報データ、日本法令の外国語訳への適用等、法令情報の電子データの標準化ができればと考えているところです。

17

招待講演

「大量日本語データから得られる知見の産業応用

ー内容分析から筆者の性格推定までー」

コールセンターにおける顧客対応記録やレビューサイトにおける消費者の書き込みなど大量のテキストデータを分析し、商品やサービスの向上につなげる取り組みが広がっている。近年では、人の性格が数値化でき性格特性が書いたテキストに反映されるという心理学の知見から、テキストの筆者の性格が推定できるようになった。

本講演では、テキストに書かれた内容の分析に加え、その筆者の性格特性を推定する仕組みと、その結果の活用可能性を紹介する。

那須川 哲哉

日本 IBM 株式会社 東京基礎研究所 主席研究員

第9回産業日本語研究会・シンポジウム

【第三部】 AI活用時代の日本語データの高度活用事例

大量日本語データから得られる知見の産業応用
ー内容分析から筆者の性格推定までー

日本アイ・ビー・エム株式会社

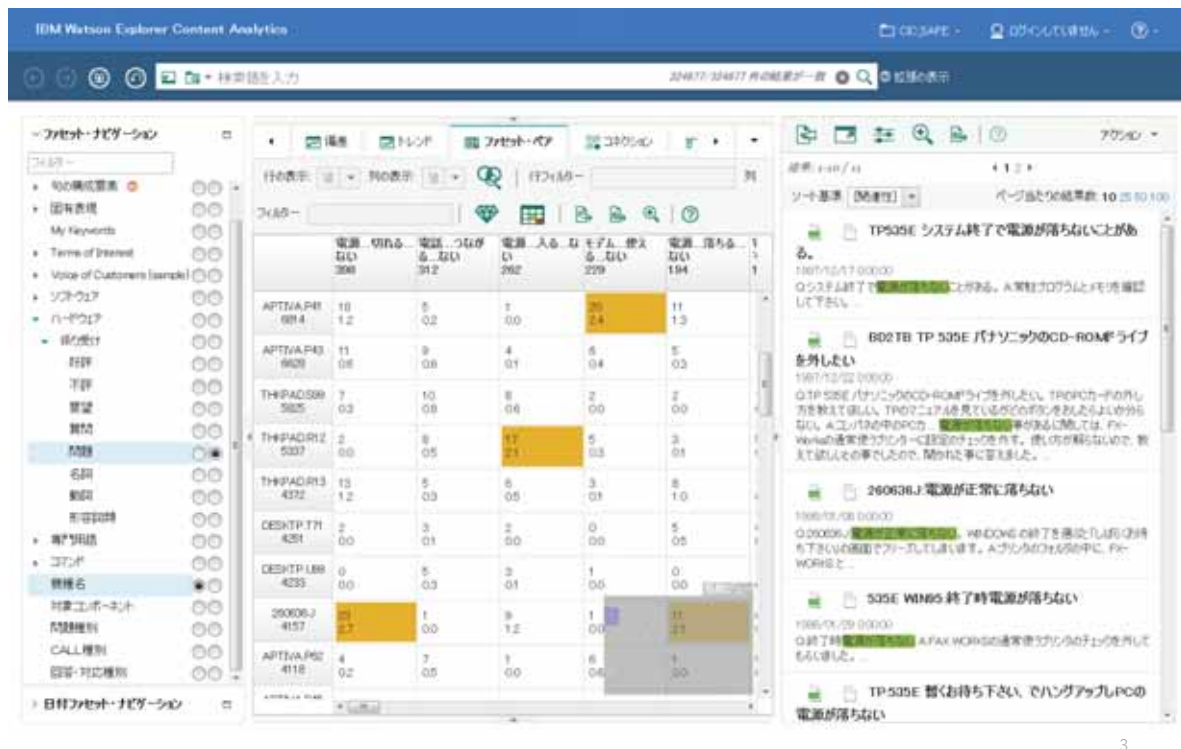
東京基礎研究所

那須川哲哉

大量の日本語データの分析例

- 自動車不具合情報の分析デモ
- PCヘルプセンターへの問い合わせ記録の分析デモ

データ全体を分析すると、特定の機種に特定の問題が集中している状況を見出すことができます。



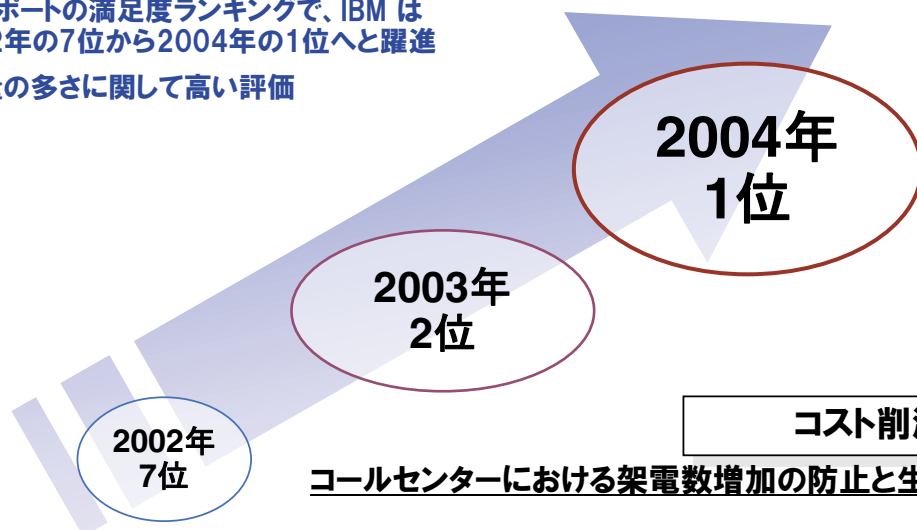
3

テキストマイニングの活用効果

お客様満足度の向上

日経パソコンお客様満足度調査 – PC サポートランキングでの順位向上

- Webサポートの満足度ランキングで、IBM は2002年の7位から2004年の1位へと躍進
- 情報量の多さに関して高い評価



4

 TAKMI Bringing Order to Unstr		A Global Volunteer Network
 The Optimization of Global Railways		Smarter Planet
 A Computer Called Watson		The Globally Integrated Ent
 The Rise of the Internet		The Networked Business Pla
 KAMAC		The Automation of Personal
 Excimer Laser Surgery		The Social Security System
 Magnetic Stripe Technology		The DNA Transistor
 The First Salaried Workforce		Corporate Leadership in En
 Optimizing the Food Supply		The Origins of Computer Sc
 The Floppy Disk		The Apollo Missions
 SAGE		Fractal Geometry
 IBM 1401: The Mainframe		Silicon Germanium Chips
 UPC		Magnetic Tape Storage
 Patents and Innovation		

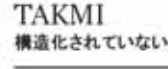
IBMの100年の軌跡 - Icons of Progress

<http://www-03.ibm.com/ibm/history/ibm100/jp/ja/stories/>

 **TAKMI**
Bringing Order to Unstructured Data

<http://www-03.ibm.com/ibm/history/ibm100/jp/ja/icons/takmi/>

IBM 100 年の軌跡

 **TAKMI**
構造化されていないデータに秩序をもたらす





1997年、IBM東京基礎研究所の研究員たちが新しい強力なテキスト分析ツールのプロトタイプを開発しました。膨大なテキストのデータベースの中にある大量の埋もれた知識を効率良く獲得し、利用するための新たな扉を、TAKMI (Text Analysis and Knowledge Mining) と名付けたこのシステムが開いたのです。

この「進歩の象徴」に貢献したえりきぎのチーム・メンバー

- 船橋川哲哉
IBM 主要研究員
- 武田浩一
技術推進、IBM 東京基礎研究所、アナリティクス&インテリジェンス マネージャー、自然言語処理
- 渡辺日出雄
IBM 東京基礎研究所、ナレッジ・メインストラクチャー・グループ マネージャー、自然言語処理
- 荻野繁雄
研究員、自然言語処理
- 村上明子
研究員、ソーシャル・アナリティクス
- 金山博
研究員、自然言語処理(構文解析・意味解析)
- 竹内正寛
研究員、自然言語処理・知能ソフトウェア工学
- 古田一星
研究員、データベース・検索・大規模データ処理
- 坪井祐太
研究員、統計的自然言語処理
- 室岡大介
研究員、検索
- 伊川祥平
研究員、データ工学、テキストマイニング
- 西山新紗
研究員、自然言語処理

日本語データの分析の深化

- 「何を表現しているか」から

「どう表現しているか」へ

→ 表現方法から筆者の性格が見えてくる

7

AI時代におけるPersonalityの重要性:
人を理解し個別対応を可能にしたい

Personalityの測り方

- 人相学
 - 人相から気質や性格を推測
- 骨相学
 - 骨格や頭蓋骨の形から知能や行動を推測
- 血液型性格分類
 - 血液型によって人の性格を分類
- ビックファイブ理論:特性5因子論
 - 人間の多様な性格は5つの要素の組み合わせで構成される
 - 膨大な統計調査や脳科学による科学的裏付けが進んできており、世界標準に

8

性格関連研究の特徴

- ビッグファイブという標準的なモデルが存在し、Personalityが数値化可能に
 - ビッグファイブ

人間の性格を5つの要素の組み合わせで記述

 - Openness to experience
独創的・好奇心が強い・開放的・知的 vs. 着実・警戒心が強い
 - Conscientiousness
手際が良い・勤勉・注意深い・まめ vs. 楽天的・不注意
 - Extraversion
外向的・社交的・エネルギッシュ vs. 内向的・孤独志向・控えめ
 - Agreeableness
人当たり良い・温情あり・協調的 vs. 冷たい・不親切
 - Neuroticism
繊細・神経質 vs. 情緒安定・自信家

9

ビッグファイブの測り方：心理テスト

	全く当てはまらない	当てはまらない	どちらでもない	当てはまる	とても当てはまる
1. 人生を楽しんでいる	○	○	○	○	○
2. 他人にあまり興味がない	○	○	○	○	○
3. 準備を怠らない	○	○	○	○	○
4. ストレスに弱い	○	○	○	○	○
5. 語彙が多い	○	○	○	○	○
6. たくさんしゃべるタイプではない	○	○	○	○	○
7. 人と接することに興味がある	○	○	○	○	○
8. 自分の持っているものを近くに置く	○	○	○	○	○

10

【性格関連研究から得られた知見】

子は親の性格を50％程度引き継ぐ

- Bouchard, Thomas J., and John C. Loehlin. "Genes, evolution, and personality." Behavior genetics 31, no. 3 (2001): 243-273.

11

【性格関連研究から得られた知見】

言葉に性格が反映される

- **文章の特徴に筆者の性格との関連性が見出せる**
 - Mairesse, F., Walker, M.A., Mehl, M.R., and Moore, R.K., Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. In Journal of Artificial intelligence Research, 30: 457-500, 2007.

12

テキストからの性格推定



<https://personality-insights-livedemo.mybluemix.net/>

13

Personality Insightsとは

- ソーシャルメディアへの書き込みや、フリーテキストから、筆者の性格特性を推定するシステム
 - <https://personality-insights-livedemo.mybluemix.net/>
 - ① Personality (性格・個性)
 - ② Needs (欲求)
 - ③ Values (価値観)
- 解析に必要なテキスト
 - 少なくとも3,500単語、理想的には6,000単語以上
- 現在対応している言語
 - 日本語、英語、スペイン語、アラビア語、韓国語

14

Personality Insightsが推定すること

- ビックファイブ/OCEAN の5軸でPersonalityを推定
 - Openness to experience: 好奇心が強い・独創的 vs. 着実・警戒心が強い
 - Conscientiousness: 勤勉・まめな人 vs. 楽天的・不注意
 - Extraversion: 外向的・エネルギッシュ vs. 孤独を好む・控えめ
 - Agreeableness: 人当たりの良い・温情のある vs. 冷たい・不親切
 - Neuroticism: 繊細・神経質 vs. 情緒安定な・自信家の
- 各軸のさらに細かい推定も可能
 - Openness to experience: 活発度、自己主張、明朗性、刺激希求性、友情、社交性
 - Conscientiousness: 大胆性、芸術的関心度、情動性、想像力、思考力、現状打破
 - Extraversion: 達成努力、注意深さ、忠実さ、秩序性、自制力、自己効力感
 - Agreeableness: 利他主義、協調性、謙虚さ、強硬さ、共感度、信用度
 - Neuroticism: 悲観的、自意識過剰、低ストレス耐性、激情的、心配性、利己的

15

Personality Insightsが推定すること： Needs(欲求)

- Kevin FordのUniversal Needs Map に沿った分析 (欲求と社会的価値の関係)
 - 個人の様々な習慣に関係 : ブランドの選択、商品の選択、職業の選択
 - Challenge: 挑戦
 - Closeness: 親密
 - Curiosity: 好奇心
 - Excitement: 興奮
 - Harmony: 調和
 - Ideal: 理想
 - Liberty: 自由主義
 - Love: 社会性
 - Practicality: 実用主義
 - Self-(expression): 自己表現
 - Stability: 安定性
 - Structure: 仕組

16

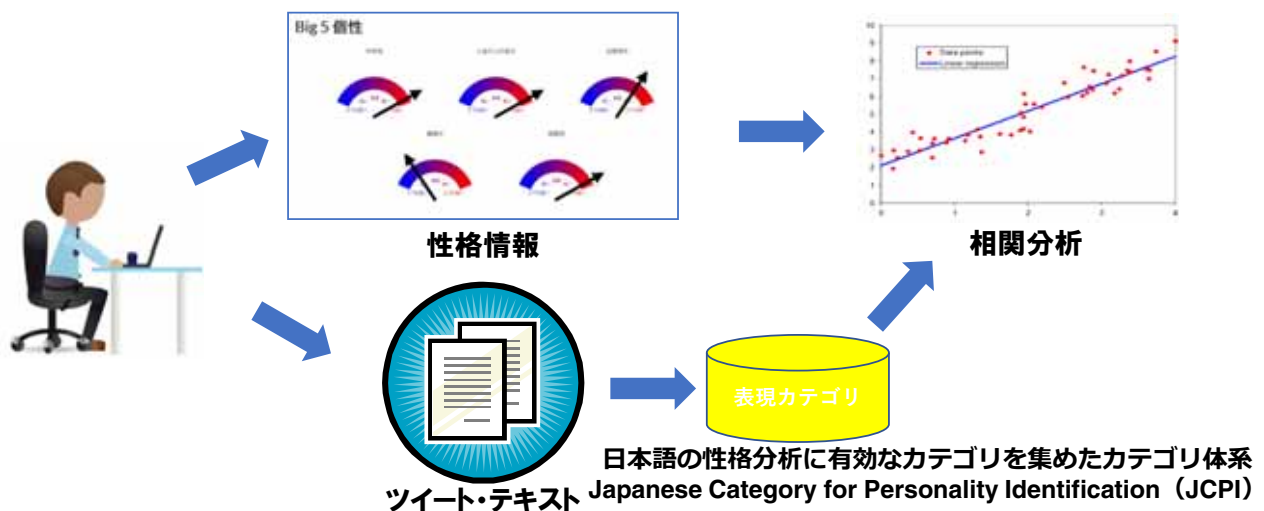
Personality Insightsが推定すること： Value(価値観)

- Schwartzの価値概説 (Schwartz Value Survey) に沿った分析
 - 4つの上位価値と10個の価値によって構成される
 - 4つの上位価値
 - Self-transcendence : 自己超越
 - Conservation : 現状維持
 - Self-enhancement : 自己増進
 - Open to change : 変化許容性
 - 10の価値
 - 博識、善行、調和、伝統、秩序、権勢、達成、快楽、刺激、自決

17

テキストから性格を推定する仕組み

1. ネット上でTwitterユーザに性格診断アンケートを依頼
2. 性格診断アンケートへの回答から推定される性格と回答者のツイートを蓄積
3. ツイート・テキストにおける各表現カテゴリの表現の割合と性格との相関を分析



18

性格診断アンケート

	全く当てはまらない	当てはまらない	どちらでもない	当てはまる	とても当てはまる
1. 人生を楽しんでいる	☐	☐	☐	☐	☐
2. 他人にあまり興味がない	☐	☐	☐	☐	☐
3. 準備を怠らない	☐	☐	☐	☐	☐
4. ストレスに弱い	☐	☐	☐	☐	☐
5. 語彙が多い	☐	☐	☐	☐	☐
6. たくさんしゃべるタイプではない	☐	☐	☐	☐	☐
7. 人と接することに興味がある	☐	☐	☐	☐	☐
8. 自分の持っているものを近くに置く	☐	☐	☐	☐	☐

19

アンケート例 - 確認用質問

- 途中で以下のような質問を何回か入れ込み、まじめに答えていない人（例えば、全て同じところにマークしている人）のデータを排除

ありがとうございます。その調子です。確認のため"どちらでもない"を選んでください。



アンケートの進捗

19.2%

20

アンケート結果例



21

テキスト中の表現の分析

- 特定カテゴリの表現がどの程度含まれているかを分析
- 日本語の性格分析に有効なカテゴリを集めたカテゴリ体系を構築
Japanese Category for Personality Identification (JCPI)
- 下記構成要素からなる約90種類のカテゴリを設定

【構成要素】

- Linguistic Inquiry and Word Count (LIWC)を参考にしたカテゴリ
- 助詞 (〔格助詞〕〔係助詞〕など)
字種 (〔漢字〕〔カタカナ〕など) といった日本語独特カテゴリ
- 〔読書〕〔遊び〕〔イベント〕などの独自カテゴリ

22

JCPIの構成要素：LIWCを参考にしたカテゴリ群

• LIWC : Linguistic Inquiry and Word Count 言語表現の特徴を心理学的観点から整理・体系化し分析する枠組み

I. STANDARD LINGUISTIC DIMENSIONS (言語的特徴)

- 代名詞 (1人称か, 2人称か, 3人称か) ・ 否定形 ・ 同意表現 ・ 他

• II. PSYCHOLOGICAL PROCESSES (心理作用、精神的・知覚のプロセス)

- 好評表現 ・ 不評表現 ・ 視覚表現 ・ 聴覚表現 ・ 他

• III. RELATIVITY (相対的概念)

- 未来への言及 ・ 過去への言及 ・ 空間表現 ・ 他

• IV. PERSONAL CONCERNS (関心の対象)

- 職業関連表現 ・ 学業関連表現 ・ 趣味関連表現 ・ 宗教関連表現 ・ 他

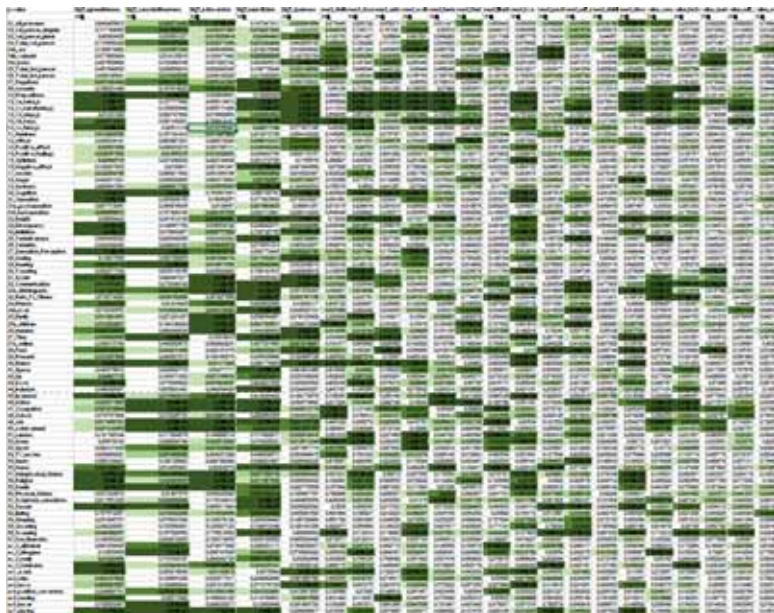
23

JCPIの各カテゴリの表現の 出現頻度と性格特性の分析

		5%水準以上の有意性	1%水準以上の有意性	0.5%水準以上の有意性	0.1%水準以上の有意性	
		big5 agreeableness	big5 conscientiousness	big5 extraversion	big5 neuroticism	big5 openness
JCPI	表現の例	P値	P値	P値	P値	P値
代名詞全体	これ、自分、それ	0.842455815	0.006572465	6.79E-04	0.187047911	0.020041654
一人称代名詞：単数	自分、私、俺	0.717784062	0.001689219	0.028267285	0.016327302	0.5236259
一人称代名詞：複数	我々、私たち、僕ら	0.89593425	0.920691227	0.012050414	0.03464401	0.001365726
一人称代名詞全体	自分、私、俺	0.715662228	0.001857832	0.022377054	0.021823403	0.442395129
俺	俺、おれ、オレ	0.128874083	0.034687343	0.406046669	0.318915648	0.518908003
私	私、わたし、あたし	0.077865434	0.514125816	0.190477489	0.005898906	0.133750754
僕	僕、ぼく、ボク	0.437609884	0.162483168	0.097032723	0.693896514	3.42E-05
二人称代名詞全体	あなた、お前、そちら	0.451549602	0.308659319	0.603990242	0.169776949	0.022198693
三人称代名詞全体	彼、彼女、彼ら	0.946196641	0.010095858	0.346100903	0.215280805	0.005902104
否定形	わからない、知らない、来ない	0.007181493	0.001979067	0.006141266	6.55E-05	0.068792822
肯定表現	OK、認める、了解	0.158221446	0.147016032	0.001622674	9.39E-04	0.002657395
助詞	の、て、が	5.32E-08	8.14E-04	0.226364851	0.15985436	1.32E-10
格助詞	が、に、を、で	1.59E-07	0.137777083	0.458119012	1.51E-07	1.31E-10
格・係・副・終助詞	が、は、に、を	1.72E-07	0.135000448	0.955171438	6.09E-05	1.31E-10
終助詞	か、な、よ、ね	0.91221098	0.082747699	0.273663854	0.006962603	5.57E-05
係助詞	は、も、しか、こそ	0.004214694	0.281142921	0.764013197	3.01E-05	1.31E-10
副助詞	とか、だけ、まで、くらい	8.20E-04	0.23511118	0.016150243	0.03677798	0.577927186
数値	2、1、3	0.013395053	0.187181413	0.011437838	0.312372736	0.854142293

24

JCPIの各カテゴリの表現の出現頻度と性格特性の分析



- 大半のカテゴリが何らかの性格特性と統計的に有意な相関を示している

5%水準以上の有意性
1%水準以上の有意性
0.5%水準以上の有意性
0.1%水準以上の有意性

25

JCPIの各カテゴリの表現の出現頻度と性格特性の分析

JCPI	表現の例	agreeableness 人当たりの良さ		neuroticism 繊細さ		openness 好奇心の強さ	
		R値	P値	R値	P値	R値	P値
僕	僕、ぼく、ボク	-0.019229991	0.437609884	0.009752	0.6938965	0.101884196	3.42E-05
格助詞	が、に、を、で	-0.128299128	1.59E-07	0.128534	1.51E-07	0.257380587	1.31E-10
係助詞	は、も、しか、こそ	-0.07064671	1.72E-07	0.102578	6.09E-05	0.178145598	1.31E-10
副助詞	だけ、まで、とか、くらい	-0.082488225	0.91221098	-0.05164	0.0069626	0.013786148	5.57E-05
終助詞	か、ね、な、よ	0.002732013	0.004214694	-0.06665	3.01E-05	-0.099139128	1.31E-10

- P値と相関係数（R値）が示唆すること
 - 〔僕〕を多用しがちな人は、好奇心が高い傾向
 - 〔格助詞〕を多用しがちな人は、冷静で、繊細で、好奇心が高い傾向
 - 〔格助詞〕を省略しがちな人は、人当たりが良く、自信家で、着実性が高い傾向
 - 〔係助詞〕を多用する人は、繊細で、好奇心が高い傾向

26

JCPIを用いた性格推定システムの実装

- 性格診断アンケート回答者のツイートを利用し、アンケートから得られた性格プロフィールと各カテゴリ表現の出現頻度とを回帰分析
- GloVeの学習済みWord Vector(200次元)のデータの日本語表現部分を利用し、カテゴリ体系に依存しない文字Nグラム($2 \leq N \leq 10$)の表現の情報を併用
- 1630人のデータを用いて、0から1の値をとるBig Fiveの各性格プロフィールに対して、10-分割交差検証により、実測値と推定値とのMean Absolute Error (MAE)を計算し、精度確認

27

性格推定システムの性能 (MAE) 評価結果

	JCPI	GloVe	JCPI+GloVe
協調性	0.1057	0.1027	0.1001
誠実性	0.0962	0.0941	0.0939
外向性	0.1211	0.1158	0.1145
感情起伏	0.1186	0.1147	0.1122
知的好奇心	0.1099	0.1064	0.1063
平均	0.1103	0.1067	0.1054

- カテゴリ体系を用いるより、Word Vectorを用いた方が精度が高いが、カテゴリ体系とWord Vectorを併用した場合に、最も高い精度

28

Personality Insights のチューニング： 結果を正しく解釈するために

- 学習データの増加に応じて、不定期に再学習
 - 同じデータに対して結果が変わる可能性
 - 画面表示・用語も随時変更
- 10万ユーザのツイートデータで分布を正規化
 - 「外向性が90%」という結果の場合
 - 100人中
 - より外向性が高い人が10人程度
 - より内向的な人が90人程度
 - Twitter以外のデータでは百分率は意味を持たない
 - 同じタイプのデータとの比較が重要
 - 同じ業務報告書のテキストでAさんBさんCさんの結果を比較
 - 「誰が最も外交的か」「誰と誰が似ているか」

29

性格推定機能の活用可能性

- 相性が良さそうな人の検出・推薦
- 商品推薦

30

相性が良さそうな人の検出・推薦

<https://your-celebrity-match.mybluemix.net/>



31

国内導入事例

- **金沢工業大学**

卒業生や上級生の履修した授業や成績、課外活動などのデータを基に、約7,000人の学生一人ひとりに的確でタイムリーなアドバイスを提供

- **フォーラムエンジニアリング**

公平で正確かつスピーディーな人材マッチングシステムを構築

- **JAL**

利用者の性格にあったおすすめ情報が提供するほか、性格診断の結果によって画面のデザインを変化

32

ご清聴ありがとうございました

本資料に記載されている内容は、全て、那須川哲哉個人の見解に基づいており、IBMの見解を示すものではありません。

IBM Watson は、International Business Machines Corporation の米国およびその他の国における商標です。

招待講演

「NTT データの AI 研究事例

ーチャットボットから AI 記者まで」

近年、生物の脳構造を参考にして考案された最先端の AI であるディープラーニング技術が実際にビジネスで通用する水準にまで進化してきています。ディープラーニングなどの AI 技術を自然言語処理に適用する事例も増えてきており、NTT データでは、アナウンサーが読み上げる原稿を自動生成する AI 記者や、自然な日本語で接客するチャットボットなどを研究開発しています。本講演では、弊社が AI 技術の発展とともに培ってきた日本語処理技術の応用事例と最新の研究成果をご紹介します。

小間 洋和

株式会社 NTT データ 技術開発本部 主任

NTTデータのAI研究事例 —— チャットボットからAI記者まで

2018年3月5日

株式会社NTTデータ技術開発本部 小間 洋和

© 2018 NTT DATA Corporation

目次

1. NTTデータのR&D重点AI適用領域

2. 顧客接点領域

事例 —— チャットボットによる接客対応

3. 業務高度化領域

事例 —— AI記者によるニュース原稿自動生成

当社のR&D重点AI適用領域

顧客対応等の
フロントオフィス業務

審査等のミドル・
バックオフィス業務

大規模データ分析による
予測や予兆検知

顧客接点領域

業務高度化領域

複合
高度分析領域

業務領域

社会基盤領域

顧客接点領域

ECサイトが抱えている課題

Webサイトを訪問
してくれたお客様を
つなぎとめたい

定型化できる
簡単な問合せ
対応を減らしたい

丁寧なクロージングで
売上増や顧客満足度を
向上させたい

AIチャットボット（ヒトとAIが力を合わせる）

ショッピングサイトやメーカー、金融機関等、ホームページでAIを活用したチャット対応を始めるお客様が増えています

ヒトとボットが連携する仕組みで解決

ヒトとチャットボットが連携することで単純な対応はチャットボットに任せ、ヒトは売上増や顧客満足度向上に注力できます



幅広いリーチには
ボットの接客

有望顧客には
ヒトの接客

- Webチャット
- チャットアプリ

チャットボット



Powered by

corevo



チャット接客



※「corevo」は日本電信電話株式会社の商標です

ボットによるチャット対応のポイント

- お客様の発言は多種多様であるため、単純なキーワードマッチ方式ではお問合せの意図を判断できません
➡ **単語ベクトルを基にした自然言語理解エンジン**
- お問い合わせ以外(雑談など)にも対応する必要があります
➡ **30万ペアの対話DBを基に構築した大規模ルール**

多様な言い回しに対する意図判定技術

NTTグループのAI「corevo™」を用いて、お問合せの「意図」を判定し、予め用意した文章を回答します。

お客様の多様な言い回し

キャンセルはいつまで可能？

キャンセル期限は注文後
何時間ですか？

注文を取消したいのですが、
いつまで可能でしょうか？



corevo

自然言語理解

お問合せの意図を判定

質問の意図

キャンセル期限

商品検索

...

支払方法

業務高度化領域

放送業界のお客様の課題

定型的な原稿を作成する作業
から解放され、取材などの
コア業務に集中したい

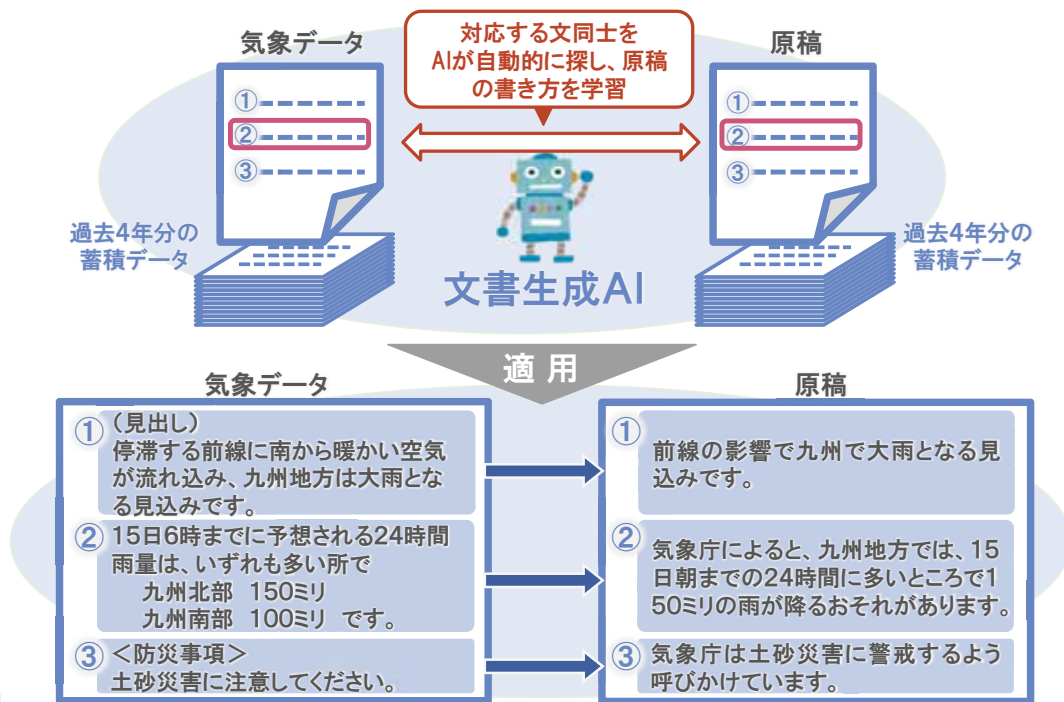
同じカテゴリの文書は書き方を
統一したいがテンプレートや
ルールの作成が困難

過去の大量データから文書生成ルールを自動作成



近年進化を遂げている深層学習によって
文書生成や翻訳／要約の精度が高まっています

深層学習を使用した文書生成アプローチ



© 2018 NTT DATA Corporation

8

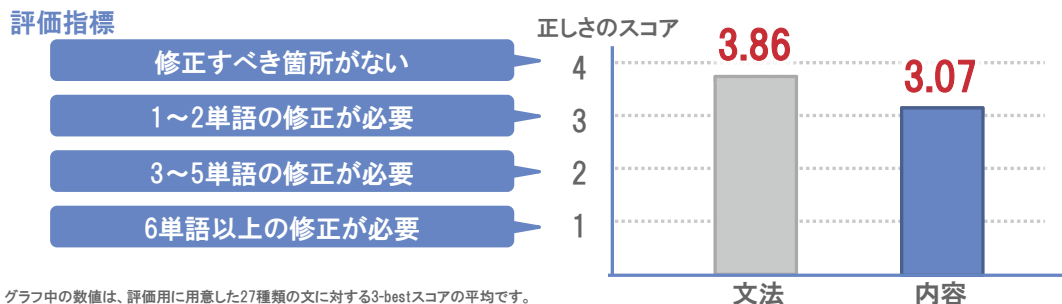
NTT DATA

効果測定と課題

生成原稿の正確性を「文法」と「内容」の観点で4段階評価

正確性評価

生成した文に含まれる誤りの個数でAIの性能をスコアリングしました



課題

- 正確性を高める補正ルールとの併用による高精度化
- 誤り箇所の可視化による人力修正の効率化

© 2018 NTT DATA Corporation

9

NTT DATA



© 2018 NTT DATA Corporation