

第11回産業日本語研究会・シンポジウム 予稿集

新型コロナウイルスの感染拡大に伴い、本シンポジウムは中止いたしました。
本予稿集の作成にあたり、ご協力と御支援をいただきました皆様に感謝いたします。

令和2年3月6日（金）

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11回産業日本語研究会・シンポジウム

まえがき

令和2年3月

AIが支える産業日本語

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、我が国産業界全体の国際競争力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日本語（「産業日本語」）のあり方を研究しています。この「産業日本語」の研究は、明瞭な日本語文の作成、高品質な翻訳文の作成コスト低減などにつながるものです。

2019年5月、新元号「令和」の時代が幕を開けました。新しい時代の訪れとともに、「人工知能（AI）」技術は完全に人の世界にとけこみ、今やテレビや新聞などでこのキーワードを見かけない日はありません。AIの波はビジネス分野にも浸透し、ニューラルネットワークをはじめとする技術は自然言語処理の中核部分に活用され、多言語での情報発信や大量の文書データ処理に不可欠な技術となっています。

また、現場における日本語研究と組み合わせることでAI技術はさらに進化し、社会にとってより有用で効果的な技術が生まれます。その技術の活用により日本語データの処理はさらに効率化され、新時代における日本語の世界が広がります。AI技術に支えられた産業日本語により、情報発信力と知的生産性の飛躍的向上が期待できます。

このような背景のもと、今回のシンポジウムでは、産業日本語研究やデータ活用の場面で我々を支えるAI技術に関する研究や取組み、そして、新しい時代の中で様々に形を変えていく日本語に関する最新の知見やトピックスなどを広くご紹介いたします。本シンポジウムが、産業日本語の更なる普及につながり、我が国産業界に大いに貢献できる機会になることを期待しております。

産業界、学术界などからの、多くの皆さまのシンポジウムへ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)
清藤 弘晃	(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、
大学共同利用機関法人人間文化研究機構 国立国語研究所、
国立研究開発法人情報通信研究機構、独立行政法人工業所有権情報・研修館、
一般社団法人情報処理学会、一般社団法人人工知能学会、
一般社団法人言語処理学会、一般社団法人日本経済団体連合会、
一般社団法人日本知的財産協会、アジア太平洋機械翻訳協会、
一般社団法人大学技術移転協議会、フジサンケイビジネスアイ

■日時：2020年3月6日（金） 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：AIが支える産業日本語

■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：

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

【第一部】 13:10-15:10

- (2) 招待講演：『元号のリズム－新元号はどうして「令和」だったのか』 ----- 3
窪菌 晴夫 国立国語研究所 教授
- (3) 招待講演：『2020に向けた外国語対応の取り組み』 ----- 9
高村 信 総務省 国際戦略局 技術政策課 研究推進室 室長
- (4) 特別講演：『わかりやすい日本語を考える－国語辞典のわかりやすさとは何か－』 ---- 19
サンキュータツオ 学者芸人・漫才師・コラムニスト

【第二部】 15:10-16:20

(5) ポスターセッション

- (5.1) 『産業日本語研究会・ライティング分科会活動』 ----- 25
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
- (5.2) 『産業日本語研究会・文書作成支援分科会活動』 ----- 33
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
- (5.3) 『産業日本語研究会・特許文書分科会活動』 ----- 37
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
- (5.4) 『特許ライティングマニュアル（改訂版）』 ----- 47
石附 直弥 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
- (5.5) 『システム開発文書品質研究会（ASDoQ）の活動紹介』 ----- 51
栗田 太郎 株式会社ソニー／システム開発文書品質研究会 幹事

- (5.6) 『文章読解・作成能力検定（「文章検」）』 ----- 55
 山田 乃理子 公益財団法人日本漢字能力検定協会 普及部 部長
- (5.7) 『翻訳バンクでニューラル機械翻訳を多分野化する』 ----- 59
 隅田 英一郎 国立研究開発法人情報通信研究機構 フェロー
- (5.8) 『Japio 世界特許情報全文検索サービス紹介』 ----- 63
 長谷川 雅昭 一般財団法人日本特許情報機構 営業推進部 主幹

【第三部】

16:20-17:50

- (6) 『開発現場における暗黙知・属人知の AI 技術活用に向けた取り組みのご紹介』 ----- 73
 西田 公祐 SOLIZE 株式会社 SOLIZE Innovations カンパニー
 Senior Manager
 西鳥羽 二郎 株式会社レトリバ 取締役/CRO
 西村 拓一 産業技術総合研究所 人工知能研究センター
 サービスインテリジェンス研究チーム長
- (7) 『ニューラル機械翻訳と特許ライティングマニュアルを用いた日本語原稿の
 プリエディット』 ----- 79
 奥山 尚一 日本知的財産翻訳協会 理事長
- (8) 『自然言語処理によるニュース記事執筆の現状と課題』 ----- 83
 岡崎 直観 東京工業大学 教授

【クロージング】

17:50-18:00

- (9) シンポジウムの終わりに
 長尾 眞 産業日本語研究会 世話人会 顧問/京都大学 名誉教授
- (10) 閉会挨拶
 小林 明 日本特許情報機構 専務理事

※懇親会・意見交換会

シンポジウム終了後、丸ビル5F カンティネッタ バズにて懇親会・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】

招待講演

『元号のリズム－新元号はどうして「令和」だったのか』

新元号として「令和」が選ばれたが、言語学（音韻論）の知識があれば どのような元号が選ばれるか、ある程度予測できる。歴代の元号を調べてみると、その音韻構造は大きな偏りを示し、基本的に2つの構造（「明治・昭和」タイプと「大正・平成」タイプ）に収束する。本講演ではこの2つの音韻構造が日本語のリズム構造に基づくものであり、赤ちゃん言葉やニックネーム、商品名など様々なところに現れていること、さらには英語などの他の言語とも共通するリズム原理に従っていることを解説する。

窪 蘭 晴夫

国立国語研究所 教授

元号のリズム—新元号はどうして「令和」だったのか？

窪 蘭晴夫（国立国語研究所）

はじめに

昨年4月1日に新しい元号「令和」(Reiwa)が発表された。私が予想していた「啓和」(Keiwa)、「永和」(Eiwa)ではなかったが、語頭の子音が異なるだけなので大体予想通りの結果である。この講演では、言語学（音韻論）の知識があればどのような元号が選ばれるか、ある程度予測できることを解説し、元号の背後にある原理（音韻構造）が日本語のリズムを反映したもので、また英語をはじめとする他の言語にも共通することを指摘する。

元号の韻律構造

「令和」ははじめて漢籍ではなく国書（今回は万葉集）から選ばれたということで新聞等では「元号1300年の転換」などと話題になったが、言葉のリズム（韻律）という点から見るとこれまでの伝統からはずれものではない。むしろ言語学的には予想通りの元号であり、これまでの元号と同じようにリズムカルな構造を有している。

元号のリズムを理解するためには、漢語の構造を理解する必要がある。元号は基本的に漢字2文字で、訓読みではなく音読みである。音読みの漢字には昭和の「昭」のように2拍の長さのものと、「和」のように1拍のものしかない。2字漢語となると、2拍+2拍、2拍+1拍、1拍+2拍、1拍+1拍の4種類しか組み合わせがないのであるが、過去の日本の元号はこの点において大きな偏りを示す。250近い元号を調べてみると、その約7割が2+2（長長）の構造のもので、それに続くのが2+1（長短）の2割強である。これに対し、1+2（短長）の元号（たとえば和同や治承）は全体の7%にしかなく、1+1（短短）に至っては皆無である。

さらに詳しく調べてみると、2+2の元号の大半が慶應、大正、平成のように〔強弱強弱〕(○○○○)のリズムを持つものである（弱の部分は長音（ー）や撥音（ん）、促音（っ）、二重母音後半（い）のような語頭に立ちえない音、すなわち言語学で特殊拍と呼ばれる弱い音である）。また2+1の元号のほとんどが、明治や昭和のように〔強弱強〕(○○○)という構造を持つものである。つまり、過去の元号は2+2と2+1の2種類で全体の9割強を占め、その大半が〔強弱強弱〕か〔強弱強〕というリズムを持っている。

このように過去の元号は初めから構造的な偏りを示している。2+2が多いのは、音読み漢字の多くが1文字2拍であることによるもので、特に不思議な偏りではない（つまり「和」のような1拍の漢字より「昭」のような2拍の漢字の方がはるかに多い）。注目すべきは長短（2+1）と短長（1+2）の差である。明治や昭和のような長短の元号があれば、治明や和昭のように短長の元号も同数あっておかしくないのであるが、実際には両者の間に5倍近い開きがある。全体として2+2と2+1を好み、とりわけ〔強弱強弱〕と〔強弱強〕

の2つのリズムを好むというのが日本の元号の大きな特徴として浮かび上がる。今回最終選考に残った6つの候補を見ても、令和^{ぼんな こうし きゅうか}、万和、広至、久化の4つが〔強弱強〕の長短リズム、英弘^{えいこう}と万保^{ばんぼう}の2つが〔強弱強弱〕の長長リズムであり、〔強強弱〕(つまり1+2の短長リズム)や〔強強〕(1+1の短短リズム)のものは皆無であった。

元号の規則性

元号のもう一つの特徴は、過去150年あまり、上記の2種類の元号の現れ方に規則性が見出されるという事実である。文久(1861~64年)から元治、慶應、明治、大正、昭和、平成と続く中で、長長(文久、慶應、大正、平成)と長短(元治、明治、昭和)の元号が交互に選ばれている。

(長長) 文久 慶應 大正 平成
(長短) 元治 明治 昭和 令和

この流れで行くと、平成に続く元号は明治や昭和と同じ長短(強弱強)という構造を持つことが予想された。「令和」はまさにこの予想通りの元号である。また、これはおそらく偶然であろうが、長短の元号の「短」の部分は「元治、明治、昭和」と続いていた。この流れでいくと新しい元号は「〇和」となることが予想されたところである。

日本語のリズム

ところで、〔強弱強弱〕の長長構造と〔強弱強〕の長短構造を好むのは元号だけではなく、日本語の一般的な特徴である。たとえば、赤ちゃん言葉はマンマ、オンブ、ダッコ、クック、バーバ、ジージのような長短(強弱強)の語と、ポンポンやブーブー、ハイハイ、ナイナイのような長長(強弱強弱)の語に二分される。〔強弱強〕は出てきても、〔強強弱〕は出てこないのが赤ちゃん言葉の際立った特徴である(たとえばバーバはあるがババーはない)。また発音の変化を見ても、詩歌(しいか)や富貴(ふうき)、三つ(みつつ)、四つ(よつつ)などは母音を伸ばしたり促音(っ)を入れたりして、短短(強強)の構造から長短(強弱強)の構造を作り出している。漫画ドラゴンボールで魔人ブウの生まれ変わりがウブではなくウーブと呼ばれるのも同じ現象である。野球の声援でも、阿部のような2拍の名前は「かつとばせえ あべー」ではなく「...あーべ」と前の母音を長くして長短(強弱強)の構造が作り出される。

一方、女王の発音がジョオーからジョーオーへ変化しつつあるのは短長(強強弱)から長長(強弱強弱)への変化である。〔強弱強弱〕のリズムも日本語では好まれており、たとえばピコ太郎のPPAP(ペンパイナッポーアッポーペン)は見事なまでに〔強弱〕の連続である。このように、日本語にはいたるところに〔強強弱〕を避けて〔強弱強〕や〔強弱強弱〕を作り出す力が働いている(拙著『通じない日本語』平凡社新書)。ちなみに〔強弱強〕や〔強弱強弱〕のリズムを好むのは日本語だけではない。英語や他の言語にも同じ傾向が見られ、たとえば英語のA and Bの表現でJerry and TomやPaul, Peter and Mary(PPM)ではなく、

Tom and Jerry、Peter, Paul and Mary という語順が好まれるのも同じリズム原理による（拙著『ネーミングの言語学』開拓社）。

このように見ると長短（強弱強）という構造を持つ「令和」は、これまでの元号の歴史と構造にも、また日本語のリズムにもとてもうまく合致していることがわかる。マスコミでは漢籍ではなく国書に由来するという点がことのほか強調されているが、言語学的に見ると新元号はこれまでの伝統を忠実に守っており、日本語が好むリズム構造と一致しているのである。

日本固有の文化を尊重するというのであれば、同じ漢字2文字でも、思い切って中国語の発音に基づいた音読みだけでなく日本語独自の読み方（訓読み、和語読み）を模索するのも一つの道かもしれない。たとえば「大和」という漢字をヤマトと和語読みすると、これまでの音読みの元号とは一味違った新鮮な響き——短短短（強強強）のリズム——が出てくる。

令和のアクセント

最後に新元号のアクセントについて述べる。テレビを見ていると「令和」を「明治」と同じ頭高（高低低）のアクセントで発音している人と、「昭和」のような平板（低高高）のアクセントで発音している人に二分できるようである。政府やマスコミ関係者は前者が多く、一般市民は後者が多いような印象を受ける。

（頭高アクセント）	レ	（平板アクセント）	イワ
	イワ		レ

言語学的にはどちらが正しいということはないが、一般に2字漢語はその構造によってアクセントが決まるという傾向が見られる。元号に限らず2+1の構造を持つ3拍漢語の約8割は頭高アクセントで発音され（たとえば政府、文化、詩歌など）、一方2+2の漢語の多くは平板アクセントで発音されている（たとえば構造、伝統、傾向など）。この傾向は過去の元号の発音にも現れており、「大化、元治、明治」など2+1のものは頭高がほとんどで、一方「応仁、大正、平成」など2+2のものは平板が多い。その例外となるのが「昭和」であり、これは「大化」や「明治」と同じ2+1の構造でありながら平板で発音されている。

なぜ「昭和」が例外的なアクセントを持つのか、この問いに答えるのはむずかしいが、一つの可能性として考えられるのが「和」である。普通名詞を見ても「平和、温和、緩和、漢和」のように〇和の語は平板アクセントが多く、また過去の元号でも「明和、永和、弘和」のように〇和のものは頭高と並んで平板も許容するようである。一般に標準語のアクセントは最後の要素によって決まるという原則があり、たとえば「色」で終わる語（赤色、オレンジ色、らくだ色など）や「病」で終わる語「心臓病、肝臓病、腎臓病など」はすべて平板アクセントである。「色」や「病」は平板化形態素と呼ばれているが、「和」もその一つであると考えられる。平板化形態素であれば「昭和」も「平和」や「温和」と同じように平板アクセントになっておかしくない。「令和」もまた、「昭和」と同じアクセントで発音されても

不自然ではないのである。

標準語の名詞は馴染み度が高くなるにつれて平板化する傾向があるとされている。そうであれば、「令和」という元号も人々の中に定着するにつれて徐々に頭高アクセントから平板アクセントへ推移していく可能性が高い。

招待講演

『2020に向けた外国語対応の取り組み』

総務省では、世界の「言葉の壁」をなくし、グローバルで自由な交流を実現するため2014年に「グローバルコミュニケーション計画」を策定し、国立研究開発法人情報通信研究機構（NICT）が開発した多言語音声翻訳技術の研究開発及び社会実装を推進してきました。本講演では、2020年に向けた取り組みとともに、2025年を見据えた多言語音声翻訳技術のさらなる飛躍に向けた取り組みについてもご紹介します。

高村 信

総務省 国際戦略局 技術政策課 研究推進室 室長

2020に向けた多言語対応の取組み

令和2年3月6日
総務省 国際戦略局 研究推進室長
高村 信

「グローバルコミュニケーション計画」（2014年4月 総務省）

1

～多言語音声翻訳システムの社会実装～

ミッション (Mission)

世界の「言葉の壁」をなくす

ビジョン (Vision)

(1) グローバルで自由な交流の実現

様々な会話を高精度に翻訳できる多言語音声翻訳システムにより世界の「言葉の壁」をなくし、世界中の誰もが国境を越えて自由に交流する社会を我が国の技術によって実現する。

(2) 日本のプレゼンス向上

医療機関での会話の支援や多言語での災害情報の配信など、多言語音声翻訳システムを世界に先駆けて社会実装することにより、外国人が暮らしやすい国を実現し、日本の価値と魅力を高める。

(3) 東京オリンピック・パラリンピックでの「おもてなし」

2020年の東京オリンピック・パラリンピックにおいて、多言語音声翻訳システムにより世界から集う選手、観客等を「言葉の壁」を感じさせることなく「おもてなし」する。

行動 (Action)

関係する企業や関係省庁等と連携、協力しながら、まずは6年間のロードマップを共有して取り組む

(プロジェクト1) 病院、商業施設、観光地等における社会実証【国家戦略特区等における活用】

産学官の連携により、大規模プロジェクトとして多様なアプリケーションを集中的に整備して社会実証を実施する。

(プロジェクト2) 多言語音声翻訳の対応領域、対応言語を拡大するための集中的な研究開発投資

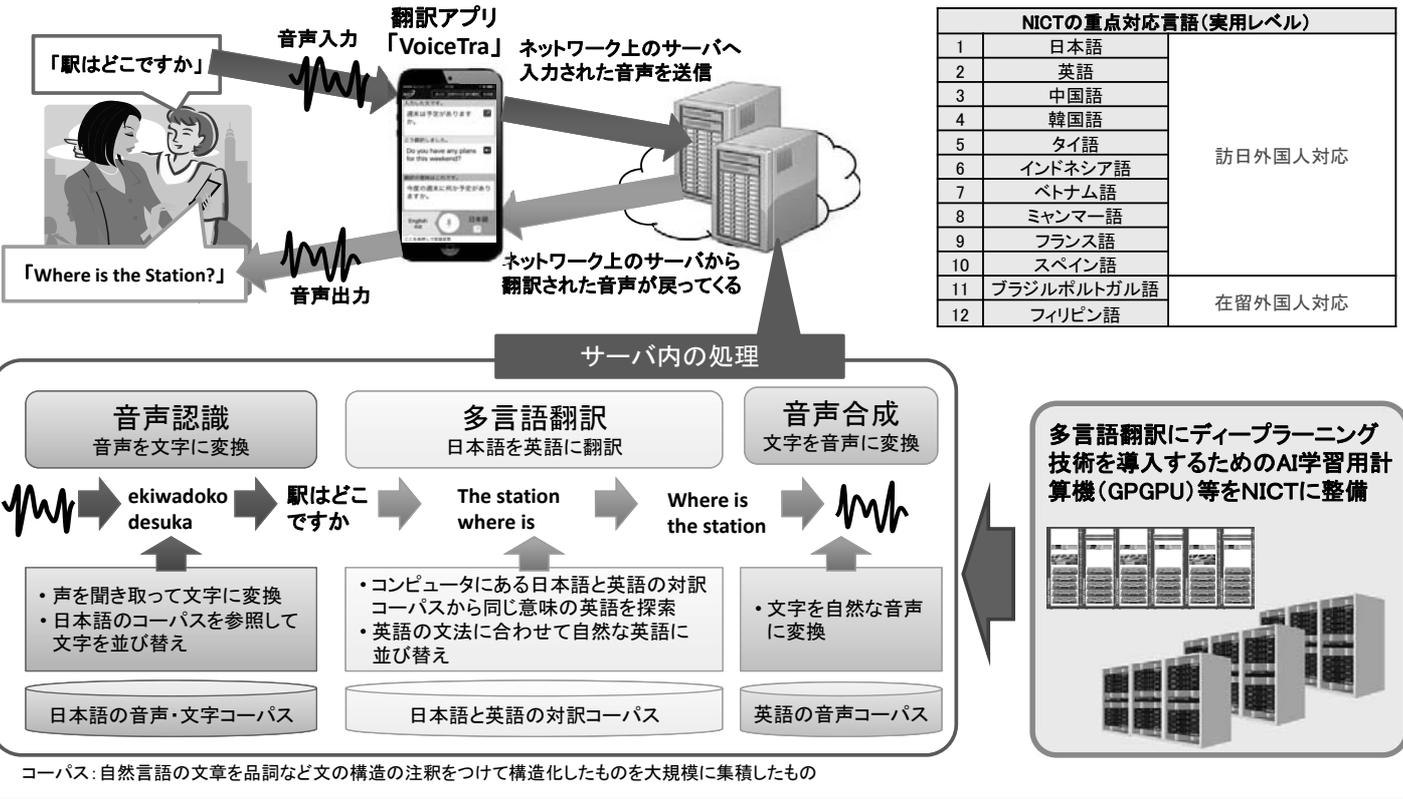
旅行会話のほか、防災・減災分野、医療分野、生活分野を重点分野として翻訳精度を向上するとともに、対応言語数を拡大する。また、多人数の同時翻訳や同時通訳の実現などの更なる高度化に向けて集中して研究開発投資する。

(プロジェクト3) 2020年東京オリンピックにおける社会実装

世界からの注目が集まる東京オリンピックを機会に、「言語の壁」がなくなる社会をショーケースとして世界に発信する。

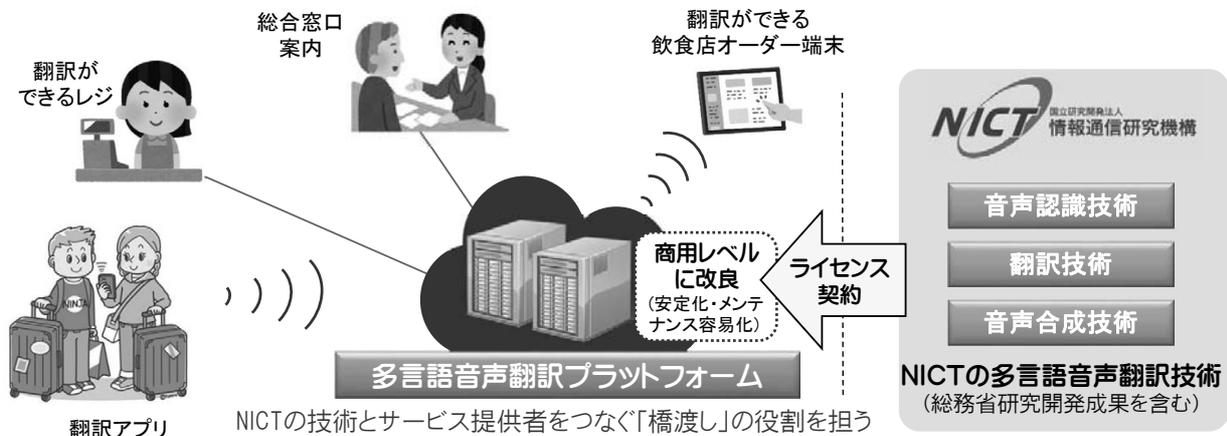
多言語翻訳技術の研究開発及び社会実装の推進

- 「グローバルコミュニケーション計画」に基づく取組により、国立研究開発法人情報通信研究機構(NICT)の多言語音声翻訳技術は、AI(ディープラーニング)技術も導入し、翻訳精度を実用レベルまで向上させ、重点対応言語を12言語まで拡大。



多言語音声翻訳プラットフォーム

- NICTが開発した多言語音声翻訳技術をサービス提供者がより簡単に利用できる環境を整備するため、ライセンス契約により民間企業に広く利用を開放する「多言語音声翻訳プラットフォーム」を本年4月に構築。
 - サービス提供者は、サーバの構築・運営・管理等の技術が不要となり、端末・アプリの開発に集中することができる。
- ⇒ 民間サービスの実用化・普及を促進



技術をより使い易くする

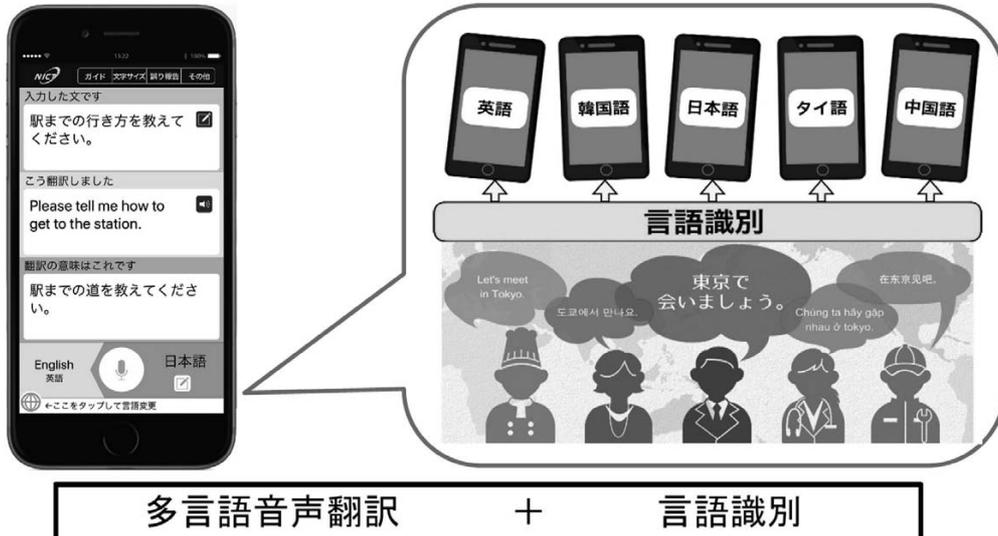
従来、サービス毎に翻訳サーバを立ち上げる必要があったが、ネット経由で簡単に翻訳機能の提供が可能に

低コスト化の推進

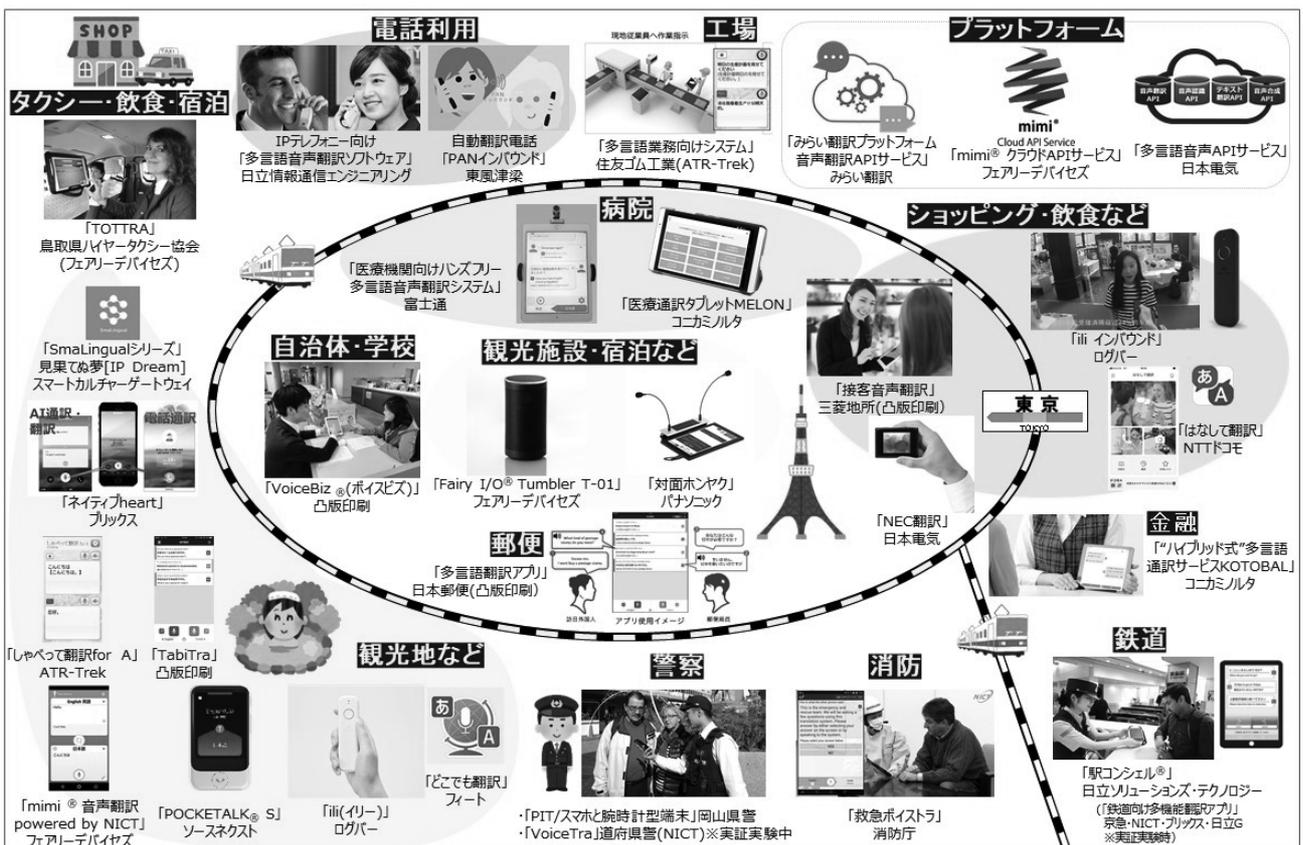
翻訳クラウドサーバが共用可能となることで、運用コストを低減し、翻訳技術の低廉な活用が可能に

相手方言語の自動識別機能の実装

- 翻訳装置では、何語で話すのかをあらかじめ設定する必要があるため、相手が話している言語が何語か分からない場合には使用が困難。
 - NICTの「VoiceTra」では、入力された相手方言語が何語なのかを自動識別する機能を開発。8言語(※)に対応した同機能を2019年10月1日から実装。
- (※) 8言語: 日本語、英語、中国語、韓国語、タイ語、インドネシア語、ベトナム語、ミャンマー語



多言語翻訳技術の社会実装



<p>多言語音声翻訳機</p> <ul style="list-style-type: none"> ●「POCKETALK^oS」 提供元：ソースネクスト(株) 	<p>多言語音声翻訳サービス</p> <ul style="list-style-type: none"> ●「駅コンシェル^o」 提供元：(株)日立ソリューションズ・テクノロジー 	<ul style="list-style-type: none"> ●「NEC 翻訳」 (音声翻訳端末/アプリ) 提供元：日本電気(株) 	<ul style="list-style-type: none"> ●「iili インバウンド」 提供元：(株)ログバー
<p>スマートフォンアプリ</p> <ul style="list-style-type: none"> ●「どこでも翻訳」 日英・日中・日韓 提供元：(株)フィート ●「はなして翻訳」 提供元：(株)NTTドコモ ●「mimi^o 音声翻訳 powered by NICT」 提供元：Fairy Devices(株) 	<ul style="list-style-type: none"> ●「SmaLingual シリーズ 多言語音声翻訳サービス」 提供元：見果てぬ夢 (IP Dream) スマートカルチャージドウェア(株) ●「多言語音声翻訳ソフトウェア」 (電話音声翻訳ソフト) 提供元：(株)日立情報通信エンジニアリング ●「Fairy I/O^o Tumbler T-01」 (多言語音声翻訳サービス向け機器) 提供元：Fairy Devices(株) 	<ul style="list-style-type: none"> ●「対面ホンヤク」 提供元：パナソニック(株) ●「VoiceBiz」 提供元：凸版印刷(株) 	<ul style="list-style-type: none"> ●「医療通訳タブレット MELON」 提供元：コニカミノルタ(株) ●「KOTOBAL」 (コトバレ) 提供元：コニカミノルタ(株)
<p>音声翻訳APIサービス</p> <ul style="list-style-type: none"> ●「みらい翻訳プラットフォーム」 提供元：(株)みらい翻訳 ●「多言語音声APIサービス」 提供元：日本電気(株) ●「mimi^o Cloud API Service」 提供元：Fairy Devices(株) 			

統合イノベーション戦略2019 (令和元年6月閣議決定)

統合イノベーション戦略 2019のポイント

- 1 Society 5.0の社会実装 (スマートシティの実現) 創業/政府事業のイノベーション
- 2 研究力の強化
- 3 国際連携の抜本的強化
- 4 最先端(重要)分野の重点的戦略の構築

<p>知の源泉</p> <ul style="list-style-type: none"> ● Society 5.0データ連携基盤整備の本格化/研究基盤データ整備/EBPMの促進 ● スマートシティ等のアーキテクチャー構築 <p>知の創造</p> <p>イノベーション・エコシステムの創出</p> <ul style="list-style-type: none"> ● 研究力強化・若手研究者支援総合パッケージの策定 ● 大学・国研の共同研究機能等の外部化 ● 大学ガバナンスコードの策定、将来ビジョンの提示 ● 初等中等からリカレントまでの人材育成改革 <p>戦略的な研究開発の推進</p> <ul style="list-style-type: none"> ● 破壊的イノベーションを目指したムーンショット型研究開発 ● 社会実装を目指した研究開発 (SIP、PRISM) 	<p>知の社会実装</p> <p>Society 5.0の実装 (スマートシティ)</p> <ul style="list-style-type: none"> ● 政府一体の取組と本格的実施 ● 官民連携プラットフォームの創設 <p>創業環境の徹底強化</p> <ul style="list-style-type: none"> ● エコシステム拠点都市形成 (大学 (起業家教育)、民間組織 (アクセラレーション) 等) <p>政府事業・制度等におけるイノベーション化の推進</p> <ul style="list-style-type: none"> ● 政府事業イノベーション化拡大 (公共事業から他分野への展開) ● 公共調達ガイドラインの普及・実践 	<p>知の国際展開</p> <p>SDGs達成のための科学技術イノベーションの推進</p> <ul style="list-style-type: none"> ● ロードマップ策定の国際議論を主導 ● プラットフォームの構築 <p>国際ネットワークの強化</p> <ul style="list-style-type: none"> ● 国際スマートシティ連合の枠組み構築 ● 国際研究開発拠点の形成 (バイオ、量子) 	<p>強化すべき分野での展開</p> <p>基盤的技術分野</p> <ul style="list-style-type: none"> ● AI技術 <ul style="list-style-type: none"> ● 全高校生がデータサイエンス・AIのリテラシーを習得 ● AI研究開発ネットワークの構築 ● AI社会原則の国際枠組み構築 →具体的な施策は「AI戦略」に基づき実施 ● バイオテクノロジー <ul style="list-style-type: none"> ● 市場領域を絞ったロードマップの策定 ● データ基盤統合化/国際バイオ都市圏形成 ● 量子技術 <ul style="list-style-type: none"> ● 「量子技術イノベーション戦略」策定 <p>応用分野</p> <ul style="list-style-type: none"> ● 環境エネルギー <ul style="list-style-type: none"> ● 「革新的環境イノベーション戦略」の策定 ● 安全・安心 <ul style="list-style-type: none"> ● 技術ニーズとシーズのマッチングの仕組みの構築 ● 重要技術分野への資源の重点配分 ● 農業・宇宙・海洋
---	---	--	---

第6期科学技術基本計画の本格検討開始 / イノベーション司令塔機能のさらなる強化

今後の研究開発重点項目	個別項目	具体的取組内容	達成時期
4-3. 人間共生型AI	言葉の壁を越える、翻訳・通訳ができるAI	ストレスなく実利用可能な以下の翻訳技術を段階的に実現 ・ 特定場面（医療、行政手続き、日常生活や旅行、ビジネス等）で利用可能（会話レベル） ・ 周囲の状況や文化的背景も考慮し、話者の意図を補足しながら利用可能（議論レベル） ・ シビアな交渉場面でも利用可能（交渉レベル）	2020年度 2025年度 2030年度
	汎用多言語自動翻訳・同時通訳技術	以下の基盤技術開発と音声認識・合成を組合せ、高精度と遅延の最小化を両立する実用レベルの同時通訳の実現 ・ 対訳が無い又は少ない条件下でも少ない性能劣化で、対話、SNS、論文、新聞などあらゆる分野で日本語のみで受発信可能な汎用多言語多分野自動翻訳 ・ 一文を超えた情報の取り込みにより、実用可能な反応速度内で高精度化を達成する技術	2025年度

外国人材の受入れ・共生のための総合的対応策（改訂）

（令和元年12月 外国人材の受入れ・共生に関する関係閣僚会議決定）

II 施策

3 生活者としての外国人に対する支援

(1) 暮らしやすい地域社会づくり

① 行政・生活情報の多言語化、相談体制の整備

【具体的施策】

○ 多言語対応の基礎となり得る自動翻訳については、多言語自動音声翻訳技術を更に簡便に利用できる基盤となる「自動音声翻訳プラットフォーム」を民間事業者が立ち上げ、官民を問わず、自動音声翻訳技術を役務として享受可能な環境が整備されたことを踏まえ、利用促進のための周知活動を実施する。

さらに、多言語自動音声翻訳技術については、2025年大阪・関西万博も見据え、日常生活・行政手続・観光等の場面に加え、ビジネスや国際会議等での議論の場面も含め、日本人と外国人及び外国人同士でストレスなく十分なコミュニケーションを可能とするため、AIによる同時通訳の実現に取り組むとともに、今般の入管法の改正も踏まえ、特定技能外国人を含め、在留外国人に対応する観点から強化対象言語を追加し、併せて翻訳精度の向上を図る。〔総務省〕《施策番号44》

○（前略）また、同相談窓口における通訳の配置・多言語翻訳アプリの導入による多言語対応（11か国語以上）等の相談体制の整備・拡充の取組を交付金により引き続き財政的に支援する。（後略）〔法務省〕《施策番号39》

○ 多言語自動音声翻訳の利用促進の観点も踏まえ、一元的相談窓口をはじめ、外国人と接する機会が多い行政機関の相談窓口においては、自動翻訳アプリ等を活用しながら、外国人の相談ニーズに適切に対応できる多言語対応を進める。〔全省庁〕《施策番号45》

(2) 生活サービス環境の改善等

① 医療・保健・福祉サービスの提供環境の整備等

【具体的施策】

○ 電話通訳及び多言語翻訳システムの利用促進、外国人患者受入れに関するマニュアルの整備、都道府県内の多様な関係者が連携し地域固有の事情を共有・解決するための対策協議会の設置等を通じて、全ての居住圏において外国人患者が安心して受診できる体制の整備を進める。〔厚生労働省〕《施策番号55》

② 災害発生時の情報発信・支援等の充実

【具体的施策】

○ 外国人からの119番通報や外国人のいる救急現場での活動等に迅速・的確に対応できるよう、電話通訳センターを介した同時通訳の体制整備を進める。

外国人のいる救急現場での活動等に迅速・的確に対応できるよう、外国人傷病者とのコミュニケーションを支援する多言語音声翻訳アプリの消防本部への導入の促進を図る。〔総務省〕《施策番号67》

外国人材の受入れ・共生のための総合的対応策（改訂）

（令和元年12月 外国人材の受入れ・共生に関する関係閣僚会議決定）

10

II 施策

3 生活者としての外国人に対する支援

(2) 生活サービス環境の改善等

③ 交通安全対策、事件・事故、消費者トラブル、法律トラブル、人権問題、生活困窮相談等への対応の充実

【具体的施策】

○ 外国人からの110番通報に迅速・的確に対応できるよう、全都道府県警察において整備している三者通話システムの活用を推進するとともに、事件・事故等の現場における外国人との円滑なコミュニケーションを支援するため、多言語翻訳機能を有する装備資機材を令和元年度中に導入し、運用を開始する。また、外国人が刑事手続の当事者となった場合において、引き続き、適切な通訳の確保を図る。加えて、検察庁に来庁等する外国人との円滑なコミュニケーションを支援するため、多言語自動音声翻訳機器の整備を検討する。〔警察庁、法務省〕《施策番号69》

(4) 外国人の子供に係る対策

【具体的施策】

○ 公立学校において、令和8年度には日本語指導が必要な児童生徒18人に対して1人の教員が基礎定数として措置されるよう、公立義務教育諸学校の学級編制及び教職員定数の標準に関する法律（義務標準法）の規定に基づいた改善を着実に推進する。また、各地域における関連部署・団体等による支援の状況等も踏まえつつ、日本語指導補助者や母語支援員の活用等の指導体制の構築や、日本人と外国人が共に学び理解し合える授業の実施、きめ細やかな指導を行うための多言語翻訳システムや遠隔教育といったICTを活用した支援等、各地方公共団体が行う外国人児童生徒等への支援体制の整備に対する支援を実施する。その際、母語・母文化の重要性に配慮するとともに、各地方公共団体におけるNPOや企業等を含む幅広い主体との連携も促進する。

また、中央教育審議会において、これら現状の施策を踏まえつつ、増加する外国人児童生徒等への教育の在り方について検討を進める。〔文部科学省〕《施策番号93》

○ 言語、母国の教育制度や文化的背景や家庭環境に留意し、適切に障害のある外国人の子供の就学先の決定が行われるよう、地方公共団体への周知を行うとともに、就学先の相談に当たって多言語化に対応した翻訳システムの活用を推進する。（後略）〔文部科学省〕《施策番号99》

(6) 適正な労働環境等の確保

② 地域での安定した就労の支援

【具体的施策】

○ 多言語コンタクトセンターの対応言語について、3か国語を新たに追加し14か国語とすることで機能強化を図る。また、通訳員を配置しているハローワークについて、各地域の実情を踏まえ、対応言語の追加の検討を行うほか、多言語翻訳システムについて試行的に導入しその効果を測定することにより、ハローワークにおける相談体制等の更なる整備を図り、円滑な就職支援を実施する。〔厚生労働省〕《施策番号129》

今後の研究開発目標

11

- 訪日外国人及び在留外国人は年々増加し、外国人との交流の機会は今後ますます増える見込み。2025年には大阪・関西万博も控え、ビジネス・国際会議の場面も含め、多言語翻訳技術の更なる高度化が必要。
※訪日外国人数（年間）：約3,119万人(H30)、在留外国人数：約273万人(H30.12月)
- 「統合イノベーション戦略」及び「AI戦略」では、2025年度までに「同時通訳」、「議論に利用できる翻訳技術」の実現などを目指す目標設定。
- 総務省としては、多言語翻訳技術の更なる推進のため「グローバルコミュニケーション計画」の次期計画の検討を進め、令和2年度からAIによる同時通訳等を実現するための研究開発を実施。【令和2年度予算案：14億円】

現在 短文の逐次翻訳

※今年度中に12言語を実用レベルで実現見込み



今後 文脈に応じた語彙の統一

(例:「携帯会社」→「モバイルキャリア」→「携帯電話事業者」)

話者の意図の補完

(例:「奈良時代」→「8世紀」)

同時通訳

※対象言語の拡大(総合的対応策を踏まえた言語追加)も目指す



今後の研究開発目標（対応言語）

- 令和2年度以降の研究開発により、NICTの多言語翻訳技術については、「外国人材の受入れ・共生のための総合的対応策」に位置づけられた言語もすべてカバーした15言語の達成を目指す。

		総務省研究開発 による対応言語	国家戦略として対応が求められている言語			
			訪日外国人 上位15か国	在留外国人 上位10か国	外国人材の受入れ・共生のための 総合的対応策	
					11か国語	特定技能の在留 資格による受入先 9か国語
1	日本語	○			○	
2	英語	○	○	○	○	
3	中国語	○	○	○	○	○
4	韓国語	○	○	○	○	
5	タイ語	○	○	○	○	○
6	インドネシア語	○	○	○	○	○
7	ベトナム語	○	○	○	○	○
8	ミャンマー語	○				○
9	フランス語	○	○			
10	スペイン語	○			○	
11	ブラジル ポルトガル語	○	○	○		
12	フィリピン語	○	○	○	○	○
13	ネパール語	×→○	○	○	○	○
14	クメール語	×→○				○
15	モンゴル語	×→○				○

特別講演

『わかりやすい日本語を考える

－国語辞典のわかりやすさとは何か－』

新しい言葉を入れることが国語辞典の使命ではありません。現在使われている日本語を、わかりやすく説明してくれているのが国語辞典という書物です。しかし「わかりやすさ」の考え方は、作っている側にとっても使っている側にとっても多様です。類語と比較した説明を「わかりやすい」とするのか、平易な言葉で説明するのが「わかりやすい」のか、用例を豊富に扱っているのが「わかりやすい」なのか。辞典の「わかりやすさ」について考えます。

サンキュータツオ

学者芸人・漫才師・コラムニスト

国語辞典において、わかりやすさとはなにか。

国語辞典において、わかりやすさとはなにか。

辞書の個性がうたわれているが、ユーザーはそれを理解せずに手に取ることが多い。収録語数の多いもの、説明が丁寧なもの、用例が豊富なもの、新語に強いもの、または保守的なものなど、さまざまある。しかしどれも「わかりやすさ」があったほうがいい。では、辞書において「わかりやすさ」とはなんだろうか。国語辞典を引き比べることで、これを考える足掛かりとしたい。

▼知っている言葉の意味や使い方を、より正確にわかりやすく教えてくれる。

【水】

＜三省堂国語辞典 7＞

- ①自然界に多くあり、われわれの生活にはなくてはならない、すき通ったつめたい液体。
〔水素二、酸素一の割合の化合物〕
「川の一・水道の一・一の事故 [=水難]」(↔湯)
- ②液状のもの。「ひざに一がたまる」
- ③新しい環境(カンキョウ)。「プロの一」
- ④〔すもうで〕←水入り。
⇒お水

＜新明解国語辞典 7＞

- ⊖泉からわき川を流れ海にたたえられたり雨となって降って来たりする、冷たい液体。化学的には、水素と酸素の化合物としてとらえられる。〔化学式 H_2O 〕〔きれいなものは無色透明で飲食に適し、生物の生存に不可欠。熱したものは「湯」、蒸発したものは「水蒸気」、凍ったものは「氷」と言う〕
「寝耳に一 [=⇒寝耳] / 湧(ワキ) 一・飲み一・雨(アマ) 一・一争い ③ 一飢饉(キキン)」
- ⊖液状の物。水分。「膝(ヒザ)の関節に一がたまる / はな一・一あめ・一ようかん・一おしろい・一物・一ぼうそう」
かぞえ方 ⊖は、一滴・一杯

▼ぼんやりとは知っているが、類語との弁別がわかりやすい

※「水素二、酸素一の割合で化合した無色無味の液体。地球の大部分を覆う。」と記述されていた時代から、三省堂国語が「われわれの生活にはなくてはならない、すき通ったつめたい液体。」と記述したことから、生活に密着した語積の在り方や、わかりやすさを指向する辞典が求められはじめた。

【丘】

＜新明解国語辞典 7＞

見る人に威圧感を与えることもなく、散歩がてらのぼることが出来る程度の高さの、ゆるやかに隆起した土地。「小高い一 / 愁ひつつ一にのぼれば花茨(イバラ)」⇒山
表記 「岡」とも書く。

かぞえ方 一丘（イッキュウ）・一座

「山」と「丘」の違いを、「威圧感」「散歩がてら」におく。

※辞書をひく人は言葉の意味を知りたいというよりも、似た言葉とのちがいやニュアンス、あるいは用例などを知りたいのではないだろうか。

▼知らない言葉を、わかるように説明する

【ガレット】

<三省堂国語辞典 7>

- ①フランスのブルターニュ地方で作る、そば粉のクレープ。
- ②フランスの焼き菓子（ガシ）。パイ生地（キジ）を使って、丸く平たい形に作る。ガレット デ ロワ。
- ③〔料〕丸く平たい形にした料理。「じゃがいものー」

※はじめて聞いた言葉を立項し、説明するのも辞書の役割。

▼なんという言葉かわからないものを調べる

【りんご】

<明鏡国語辞典 2>

多く紅色・黄緑色の甘酸っぱい果実を食用とするバラ科の落葉高木。また、その果実。生食のほかジャム・ジュース・酢・酒などの原料にする。紅玉・陸奥（むつ）・ふじ・つがる・王林・ゴールデンデリシャス・スターキングデリシャスなど、栽培品種が多い。四、五月ごろ、枝頂に白または淡紅色の五弁花をつける。

※日本人で「りんご」をひく人はなにを求めているのか。意味ではなく用途や品種、あるいは作り方などと考えた結果のもの。

このように、他の辞書には見られない強みを持った各辞典を、「わかりやすさ」という軸をもとに比較検討し議論の材料とする。論点を明確にするのが本講演の目的である。

【第二部】ポスターセッション
ポスター概要説明

(1) 『産業日本語研究会・ライティング分科会活動』

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

2019年度活動報告

2020年3月6日

JAPIO ライティング分科会

ライティング分科会の活動報告

【活動目的】

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きます。

活動指針

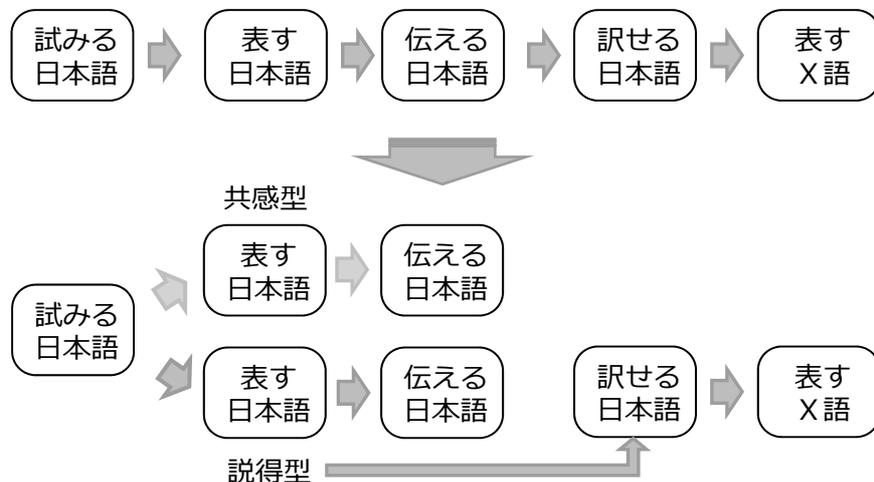
- ・ 思考の道具（知覚、情動、知性の顕在化）として日本語を分析する（道具論）
- ・ ライティングプロセスモデル（横井）を基礎として思考と書き方の連繋を見出す
- ・ 日本語を使いこなすためのマニュアルを作成する
- ・ 対象分野はビジネス文書（筋書きに沿って主張を伝える文書）とする

2019年度 【5回の会議を実施】

- 1 : ■ 各委員による具体事例（日本語文章）の分析（原因・理由-結果の使い方の妥当性）
- 2 : ■ 各委員による具体事例（日本語文章）の分析（原因・理由-結果の使い方の妥当性）
 (■ 原因・理由が結果に結びつかない、惟作法が違いそうだとの議論を行った)
- 3 : ■ ものの実体表現の違いの検討（主張表現の背後の思惟作法の違いを含む）
- 4 : ■ ものの实在性と動きの表現の再検討（主語と主題、動詞と述語の違いの検討）
- 5 : ■ ものの实在性と動きの表現の再検討（筋書きの作り方の違いの再検討）

検討の結果

- ・ ライティングプロセスモデルの修正



ものの言語表現上の实在性

モノのそれらしさ（価値指向？主観？）

唯心的見方，一元的捉え方，総合判断

〔モノである〕

共感型文章

モノのかたち（事実指向？客観？）

唯物的見方，二元的捉え方，分析判断

〔モノがある〕

説得型文章

- ・ 過去の議論も含め「モノのそれらしさ，モノのかたち」の観点から再整理しました。

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きます。本年度、主張（価値と事実／立場と態度）を含む文章の段（パラグラフ）構成と、主張を効果的に伝えるための二つの論じ方を明らかにしました。

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

文書と文章、段（パラグラフ）

文書	【役割】 確約（コミット）するコト		主張
	【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（主張の戦略）	二つの戦略
		【形式】 筋書き（アウトライン）	二つの論じ方
		段（パラグラフ）	【ことば遣い】 正確な語彙や表現
【形式】 字下げと話題文，支持文	表現の型		

文書を書く時に点検

- 主張（価値と事実）を意識する
- 発信者と受信者を確認する
- 文書儀礼を確認する
- 読み手との共有知識を確認する



- ✓ 価値（立場） or 事実（態度）？
- ✓ 両者の関係性を理解しているか
- ✓ 適切に記述できているか
- ✓ 読み手の知識を具体的に意識できているか

〔主張〕

〔価値の主張〕

価値を主張し立場を確約する（結論）

読み手に納得してもらおう立場（共感）

根拠から結論に至る関係群がある
 関係の存在の蓋然性が高いこと（因果であること）を表す
 （関係の実在度合が結論を支持する）

〔関係である〕

共感とは関係の実在度合を担保するもの

関係の実在 ▣ 「～である」でモノを表す

〔モノである〕

共感型文章：皆の見解を統一する

一般的・恒常的叙述に偏向する（モノが関係的・機能的に在るから）

〔事実の主張〕

事実を主張し態度を確約する（結論）

読み手を説得する態度（意志）

根拠から結論に至る関係連鎖がある
 根拠の存在の蓋然性が高いこと（因果があること）を表す
 （根拠の実在度合が結論を支持する）

〔根拠がある〕

意志とは根拠の実在度合を担保するもの

根拠の実在 ▣ 「～がある」でモノを表す

〔モノがある〕

説得型文章：相手の見解を変える

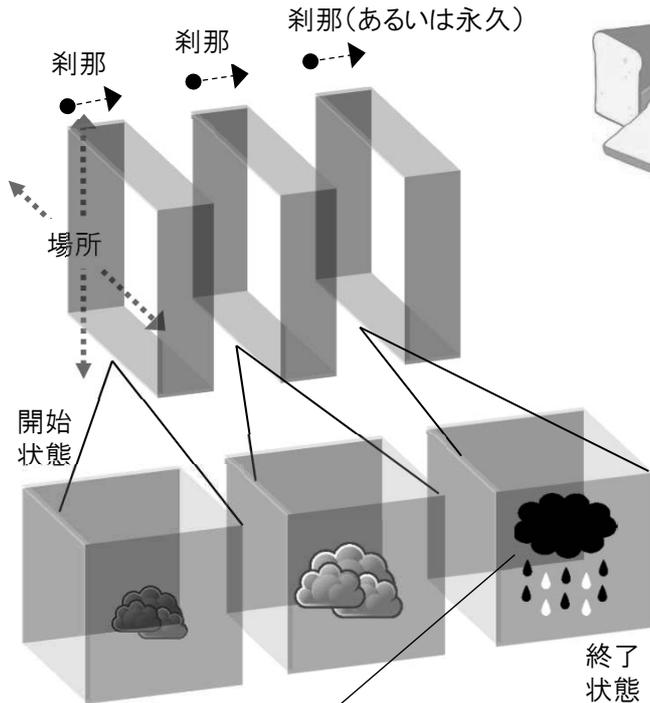
個別的・特定の叙述に偏向する（モノが知覚的・具体的に在るから）

【論じ方とモノの見方（思惟特徴）】

語る姿勢（主張の戦略）に結び付いている二つの論じ方は、時間経過の創造の仕方（不変性と不動性）の違いから生まれます。モノの見方（～である／～がある）と動きの表し方（状態変化／位置変化）が違ふことを意識しましょう。

〔モノである〕（食パン一枚モデル）

動き：モノは状態を変える～状態変化
（離散時間，相対空間）



食パン一枚モデル

不変性があると見做す
〔終了状態が既知〕

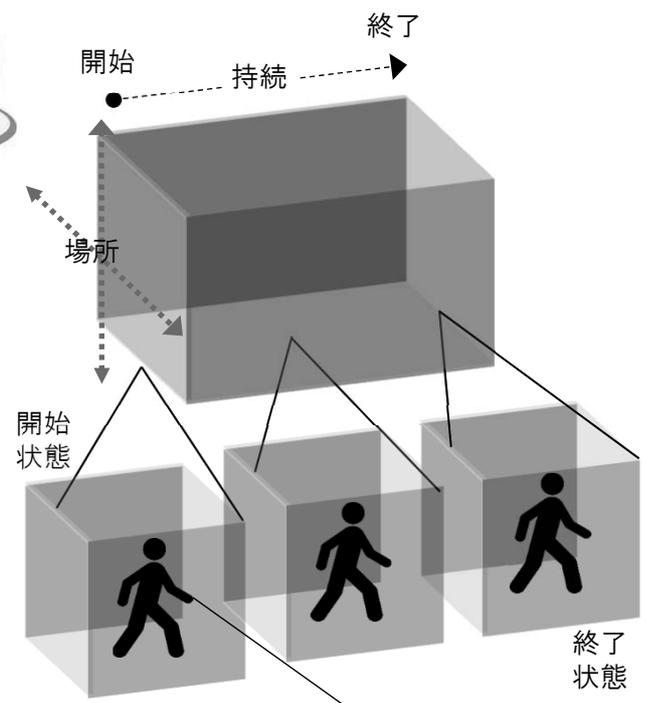
不変状態の結果が既知

結果が在る。ある原因・理由（開始状態）を前提とすると結果である

関係の实在度合が結論を支持
思惟方向は、これより前（確率論的）

〔モノがある〕（食パン一斤モデル）

動き：モノは位置を変える～位置変化
（連続時間，絶対空間）



食パン一斤モデル

不動性があると見做す
〔開始状態が既知〕

不動状態の原因・理由が既知

原因・理由が在る。この原因・理由ならば結果（終了状態）が起こる

根拠の实在度合が結論を支持
思惟方向は、これより後（決定論的）

仮説と演繹思惟

推論・思惟

直接的な演繹思惟



〔尤も確からしい〕



〔尤も確からしい〕

把握（結果）
仮説（原因・理由）
結論（関係）

$$D$$

$$p(H_i|D)$$

$$\therefore H_i \rightarrow D$$

〔関係の实在度合が結論を支持する〕

〔根拠の实在度合が結論を支持する〕

前提（関係）
把握（原因・理由）
結論（結果）

$$H \rightarrow D$$

$$H_i$$

$$\therefore D$$

ライティング分科会の活動報告

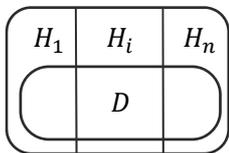
【語る姿勢と叙述の特徴】

語る姿勢（主張の戦略）に結び付いている二つの論じ方は、思惟特徴である総合的な判断の主張なのか、分析的な判断の主張なのかの違いに分かれます。関係の实在度合を述べるのか根拠の实在度を述べるのかを意識しましょう。

価値・立場主張〔納得型文・パラグラフ〕

関係の实在度合が結論を支持する

D 「英語が苦手だ」
 $p(H_i|D)$ 「英語が苦手である条件で日本人だ」
 $\therefore H_i \rightarrow D$ 「日本人が英語が苦手である」
 (関係の实在度合)



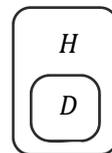
総合判断：分析判断でない判断を述べる

「日本人は英語が苦手だ。」

事実・態度主張〔説得型文・パラグラフ〕

根拠の实在度合が結論を支持する

$H \rightarrow D$ "Japanese people (are/be) poor at English"
 H_i "the/a Japanese (is)"
 $\therefore D$ "(be) poor at English"
 (根拠の实在度合)



分析判断：概念のうちに必然的に含まれている性質を取り出して述べる

"The/A Japanese is poor at English."

パラグラフ

論じ語り方（起-承-転-結，起-承-結 (introduction/body/conclusion)）で整理する

- 主張内容を導入する話題（文）
- 読み手が共有していると思われる知識を根拠として結論に至る表現を、複数文列挙する

- 主張内容を要約する話題（文）
- 書き手が根拠としている知識から結論に至る表現を、複数文列挙する

起 導入文（価値や立場を示す）

起 導入文（事実や態度を示す）

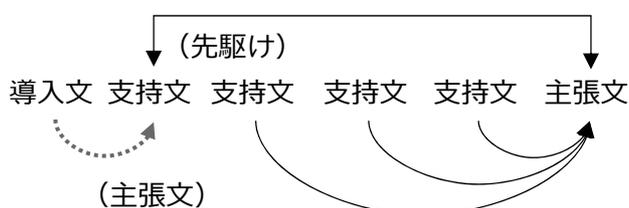
承 支持文（主張の先駆け）
 支持文*（関係に説得力を持たせる）

承 主張文（根拠の实在性を述べる文）
 支持文*（根拠に説得力を持たせる）

転 支持文（視点を変えて問い直す）

結 支持文（主張の畳みかけ）

結 主張文（関係の価値を主張する文）



書き方規則（メタ規則）

- 公の（共有された）論考か、個人の論考（思い）かを区別する（情動注意）
- 一般的な類例から具体的な類例の順序で列挙する（確率論的，離散時間）
- 高確度の結果に結び付く根拠と原因を考える（個別への配慮しすぎに注意する）

- 誰もが知覚できるのか、個人の知覚（経験）かを区別する（独善注意）
- 出来事は時間順序や一般から具体への順序で列挙する（決定論的，連続時間）
- 根拠と原因は一因でなく多面的視点が必要（根拠の単純化し過ぎに注意する）

【伝わる書き方で文章を書こう！】
文章は「語る姿勢」と「形式（起承転結か起承結）」を決定し、「健全な態度」に沿って書くことが必要です。「語る姿勢」と「健全な態度」には、文章化する際に、二つの効果的な表現技法があります（共感技法と意志技法）。

価値・立場主張	【語る姿勢】合理的な心構え...（共感，意志）	事実・態度主張
<ul style="list-style-type: none"> ■ 共感による納得は、共通項を探すことであり（調和），多視点で客観的な事実や主観的な信念を担保づけること 		<ul style="list-style-type: none"> ■ 意志による説得は、他の選択肢を排除することであり（対立），単視点で客観的な事実や主観的な信念を裏付けること

価値・立場主張	【健全な態度】根拠の考え方...（共感，意志）	事実・態度主張
<ul style="list-style-type: none"> ■ 根拠（共有知識）を明確にして表現する ■ 公の見解と私的意見を明確に区別する ■ 安易な判断と断定は思考を止めるので避ける 		<ul style="list-style-type: none"> ■ 根拠（一般事実）を明確にして表現する ■ 追認された出来事と経験を明確に区別する ■ 安易な判断と断定は思考を止めるので避ける

共感と納得（の例）

意思と説得（の例）

日本人は、外国語が苦手である。その苦手を克服するために努力しなければならないことを様々な観点から議論してきた。議論の中には、日本語が世界の言語の中でも際立って特殊な言語であり、本来、日本人には、外国語は向かないという悲観論を説くものもあった。確かに、その悲観論も、欧米の諸言語の有様を調べてみると、頷けないことはない。しかし、本稿は、すべての国々の人々にとって、自国語以外の言語を習得する困難さには、さしたる違いはないという観点に立つ。その観点から、苦手とする外国語を習得するために日本人が心得なければならない要点は、外国語を使う環境に身を置くこと、日本語の能力を鍛えること、自分独自の考えや主張を持つことである。



英語に関する意識調査は、日本人が、未成年・成人ともに6割超が英語に対して苦手意識を持っていること示している。その苦手意識を克服するポイントが、様々な観点から議論されている。その議論のひとつに悲観論がある。例えば、日本語と中国のイー族の彝（い）語が音節文字を持つ言語である。さらに日本語だけが表意文字と表音文字を使う。したがって日本語は際立って特殊な言語であるから日本人に外国語習得は難しい。一方に、楽観論がある。その議論は、第二外国語学習の観点から、自国語以外の言語を習得する困難さに共通点があることを主張する。本稿は、楽観論に賛成する立場である。この立場から日本人が心得るべき学習の要点を提案する。その要点は、(1) 外国語を使う環境に身を置くこと、(2) 自らの日本語の能力を鍛えること、(3) 自分独自の考えや主張を持つことである。

納得技法	※ 段（パラグラフ）作文の技法		説得技法
提示の仕方	結論に関する価値・立場を述べる／結論に至る事実・態度を述べる		
支持の仕方	共感を通じて立場を納得してもらおう（結論から根拠を押し量る）	意志をもって態度を示し説得する（根拠から結論を導く）	
根拠への注意	個人的な信念を根拠にし過ぎない	個別的な出来事を根拠にし過ぎない	
書き手の態度	根拠（共有知識と一般事実）を積み上げて正しさを主張する		
	信念（理性的な情感・情意）で価値を主張する		
表現	「彼は家に困っている。」等・「何々は（どんな）だ」，「何々が～に（どう）だ」，「何々が（どう）なる」	「革新技术が生活を変える。」等・「何（々）が（何（々）をどう動かす」，「誰（々）が（何（々）をどう為す」	

(2) 『産業日本語研究会・文書作成支援分科会活動』

橋田 浩一

東京大学 教授／

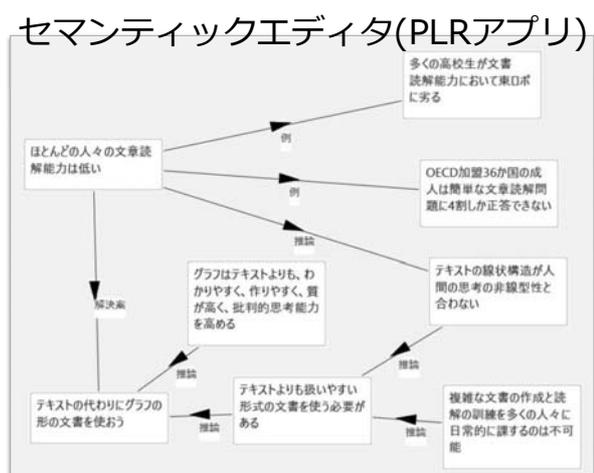
産業日本語研究会 文書作成支援分科会 主査

文書作成支援分科会

2020-03-06

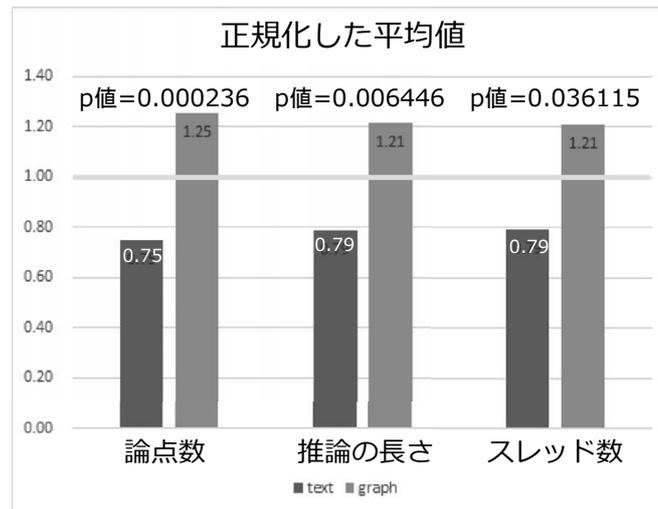
グラフ文書による生産性の向上?

- 先進国でも多くの人々の読解力は低い
- 日本では読解力の格差が拡大している
- グラフ文書
 - ◆ 各ノードは単文や名詞句
 - ◆ 各リンクは談話関係など
 - ◆ 任意のテキスト文書の内容を表現可能
- 日常生活と業務において大量のグラフ文書が持続的に生成・共有される社会を目指す
 - ◆ 人間による文書の作成・読解およびAIの実行・学習の効率と質を高める
 - ◆ PLRによる安価で安全な文書管理



文書の共同作業に関する実験

- 社会的合意形成や共有知の形成は文書の共同作成
- **テキスト**よりも**グラフ**の方がその生産性が高い
 - ◆ 2人のグループ×9によるグラフとテキストの共同作成に関する実験



3

(3) 『産業日本語研究会・特許文書分科会活動』

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

産業日本語特許文書分科会活動報告

特許明細書の標準化に向けた取り組み

～特許文書品質特性モデル～

産業日本語研究会 特許文書分科会
谷川英和・安彦元・黒川恵・久保田真司・杉尾雄一・的場成夫・清藤弘晃・石附直弥
2020/3/6

1

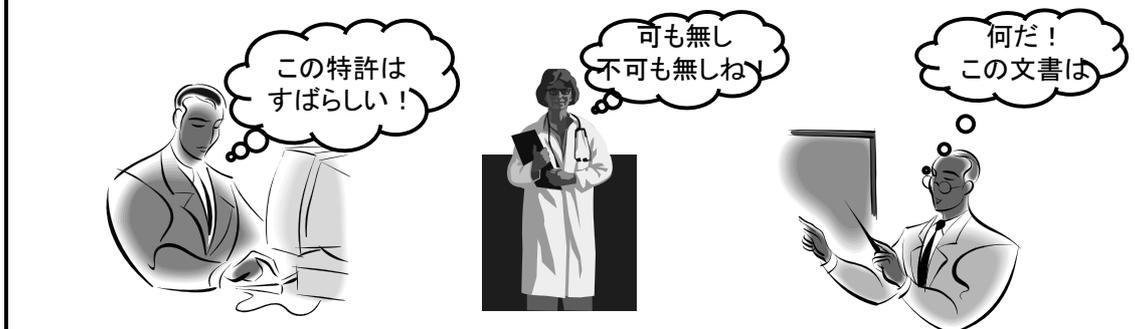
目次

1. 背景
2. 特許文書品質特性の概要
3. 特許文書品質特性を用いた評価
4. 評価目的・評価シチュエーション別の重要度
5. 各特性の事例
6. 今後

2

1. 背景

人、立場、シチュエーション等(例えば、出願前の事務所、中間処理時の出願人、侵害訴訟時の競合企業等)によって、特許文書の評価は異なっている

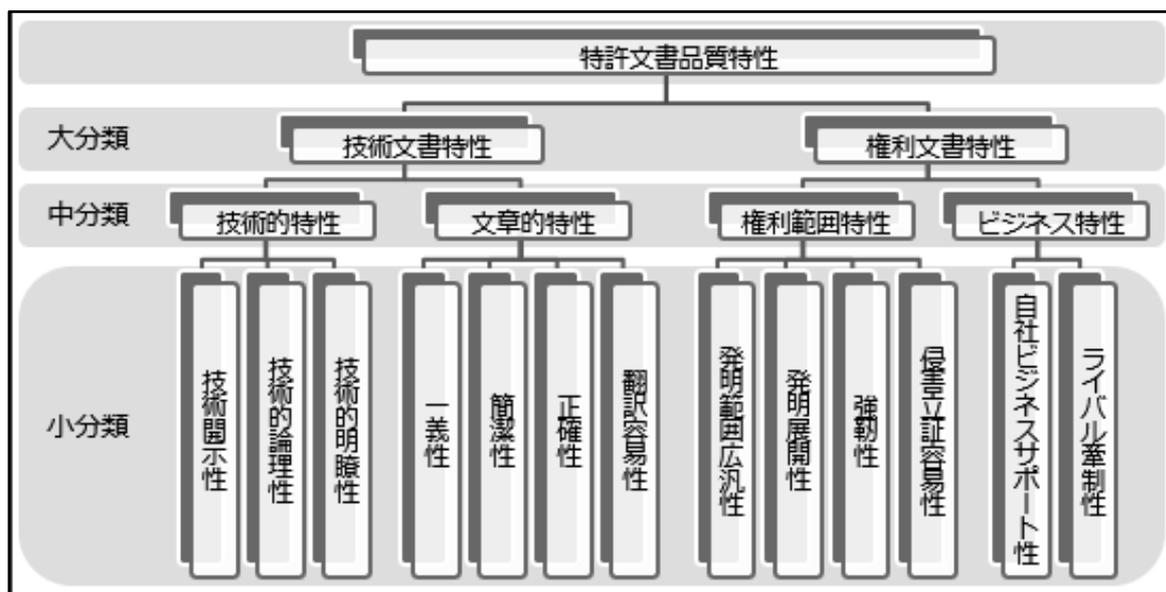


特許文書の品質の標準化が必要

特許に関する活動が加速

3

2. 特許文書品質特性の概要(1)



2. 特許文書品質特性の概要(2)

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、実施可能要件を満足している度合い。
		技術的論理性	論理的である度合い（例えば、クレームと明細書のストーリーとの技術的因果関係が明瞭であること、背景・従来技術・課題・解決手段・効果等のストーリーの筋が通っている度合い）。
		技術的明瞭性	発明の技術的な説明内容が明瞭である度合い。
	文章的特性	一義性	特許文書を構成する文章が一義的に捉えられる度合い。
		簡潔性	特許文書を構成する各文が簡潔である度合い。
		正確性	特許文書を構成する各文に誤りがない度合い。
		翻訳容易性	翻訳のし易さの度合い。
権利文書特性	権利範囲特性	発明範囲広汎	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。

3. 特許文書品質特性を用いた評価①

～ 定性的評価 ～

小分類	品質特性の説明	評価方法(定性的)
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？
技術的論理性	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？
技術的明瞭性	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？
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？
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？
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？
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？
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？
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？
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？
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？
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？
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？

3. 特許文書品質特性を用いた評価② ～ 定量的評価 ～

小分類	品質特性の説明	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。	・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
技術的論理性	論理的である度合い	・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
技術的明瞭性	発明の説明が明瞭である度合い	・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	・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	・長文、複文の数 ・重複記載の箇所数	・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	・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言いやすい用語の使用率
発明範囲 広汎性	権利範囲の広さの度合い	・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。	・クレーム数 ・クレームツリーのネストレベル(深さ)	
強靭性	拒絶、無効になりにくい度合い	・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	・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	・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	・契約情報(ライセンス契約を締結している等)	

7

* 格成分数・・・特許の広さ指標。特許請求の範囲の記載において各構成要素につきどれだけ限定がかけられているかを数値化したもの

4. 評価目的・評価シチュエーション別の重要度

品質特性			評価者 評価対象	特許書類の評価目的・評価のシチュエーション								
大分類	中分類	小分類		出願前			出願後・権利化前			権利化後		
				出願前の知財部のチェック	企業に提出前の所内チェック	審査着手時	中間処理時	中間処理時	侵害訴訟	他社抑制効果	ライセンス契約	
			出願人 (主に企業)	作成者側 (事務所)	審査官	出願人 (主に企業)	作成者側 (事務所)	競合企業	競合企業	競合企業・ライアンス企業		
技術文書特性	技術的特性	技術開示性	明細書	A	B	A	C	C	A	B	C	
		論理性	明細書 請求の範囲	A	B	A	C	C	B	C	C	
		明瞭性	明細書	A	B	A	C	C	A	B	C	
	文章的特性	一義性	明細書 請求の範囲	B	B	A	C	C	A	B	C	
		簡潔性	明細書 請求の範囲	B	A	C	C	C	C	C	C	
		正確性	明細書 請求の範囲	B	A	A	C	C	A	B	C	
		翻訳容易性	明細書 請求の範囲	B	A	C	C	C	C	C	C	
権利文書特性	権利範囲特性	発明範囲広汎性	請求の範囲	A	A	C	A	A	A	A	A	
		発明展開性	請求の範囲	A	B	C	A	A	A	B	B	
		強靭性	請求の範囲	A	B	A	A	A	A	B	C	
		侵害立証容易性	請求の範囲	B	B	C	B	C	A	B	C	
	ビジネス特性	ライバル牽制性	請求の範囲	B	C	C	A	C	A	A	B	
		自社ビジネス貢献性	請求の範囲	A	C	C	A	C	A	A	A	

8

5. 各特性の事例(1)

大分類	中分類	小分類	対象特許	悪例(品質特性に関わる箇所に下線)	良例(品質特性に関わる箇所に下線)	コメント・理由
技術文書特性	技術的特性	技術的開示性	2(タクシー)	図2に示す、顧客Xが顧客端末5011に表示されている空車状況情報から空車Bを選択すると、顧客Xが空車Bを選択した旨の空車選択通知とともに、顧客Xの電話番号およびメールアドレスを含む顧客特定情報が、基地局101のサーバ105の結合コネクション106へ送信され(ステップS605)、これが仲介テーブル107へ記入される。	図2に示す、顧客Xが顧客端末5011に表示されている空車状況情報から空車Bを選択すると、顧客Xが空車Bを選択した旨の空車選択通知とともに、顧客Xの電話番号およびメールアドレスを含む顧客特定情報が、基地局101のサーバ105の結合コネクション106へ送信され(ステップS605)、これが仲介テーブル107へ記入される。結合コネクション106とは、顧客端末5011と車載装置3とが情報を送受信する中継の装置である。結合コネクション106は、例えば、クラウドサーバ、ASPサーバ等問わない。仲介テーブル107は、顧客端末5011と車載装置3とが情報を送受信するための情報が管理されるテーブルである。結合テーブル107は、例えば、RDB、CSVファイル等であり、その構造は問わない。	漢語である「結合コネクション106」「仲介テーブル107」の定義、および例示は記載しないと、技術開示性が低くなる。
			1(半導体装置)	図3は、以上のようにして形成されたMOSTランジスタ100の、動作を説明する図である。(中略)図3(h)においては、ゲート電極112にゲート電圧VGとしてOVが入力されているため、MQS-FEI100はオフしており、電流は流れない。	図3は、以上のようにして形成されたMOSTランジスタ100の、動作を説明する図である。(中略)図3(h)においては、ゲート電極112にゲート電圧VGとしてOVが入力されているため、MQSランジスタ100はオフしており、電流は流れない。	「MOSTランジスタ」の用語が統一されておらず、符号100に対して、「MOS-FET」という異なる用語が使われている。
			視力回復組成物	前記アントシアニンの粒子径が100nm以下にナノ化されたものであることを特徴とする請求項1〜7のいずれか1項に記載の視力回復組成物。	前記アントシアニンの重量平均粒子径が100nm以下にナノ化されたものであることを特徴とする請求項1〜7のいずれか1項に記載の視力回復組成物。	「粒子径」とは、平均粒子径を指すのかどうか不明でない。また、平均粒子径を指したとしても、重量平均粒子径や数平均粒子径などのどれを指すのかどうか不明でない。
	統一性	1(半導体装置)	例えば、図6(d)では、シリサイド化用金属膜として、コバルト(Co)膜20を使用した。これに代えて、より安価なニッケル(Ni)の膜を使用してもよい。	例えば、図6(d)では、シリサイド化用金属膜として、コバルト(Co)膜20を使用した。これに代えて、より安価なニッケル(Ni)の膜を使用することも可能である。また、白金(Pt)の膜を使用してもよい。	より安価な「ニッケル」だけに倣っているのか、白金にも倣っているのか、一般的に解釈できない。	
		2(タクシー)	…すると基地局101は、顧客Xの周辺にいる空車A、空車B、および空車Cを空車位置データベース102から特定して、それらの位置情報を取得する(ステップS202)とともに、これらの車両の属性情報を車両データベース103から取得し(ステップS203)、さらに顧客の周辺の地図データを地図データベース104から取得し(ステップS204)、以上の情報およびデータベースを合成して空車状況情報を作成し(ステップS205)、顧客Xの顧客端末501へ送信する(ステップS206)。	基地局101は、顧客Xの周辺にいる空車A、空車B、および空車Cを空車位置データベース102から特定して、それらの位置情報を取得する(ステップS202)。基地局101は、空車A、空車B、および空車Cの属性情報を車両データベース103から取得する(ステップS203)。基地局101は、顧客の周辺の地図データを地図データベース104から取得する(ステップS204)。基地局101は、取得した位置情報、属性情報、および地図データを合成して空車状況情報を作成する(ステップS205)。基地局101は、顧客Xの顧客端末501へ空車状況情報を送信する(ステップS206)。	複数の処理を一文で繋げよう、分割した方が簡潔である。必要でない接続詞は削除すべき。	
	文書的特性	正確性	3(視力回復組成物)	パソコンやスマートフォン等の普及により、これらの機器を長時間にわたって使用し続けることで目の疲れや痛み、更には視力の低下を訴える人が急増している。	普及したパソコンやスマートフォン等の機器を長時間にわたって使用し続ける人が増えたことにより、目の疲れや痛み、更には視力の低下を訴える人が急増している。	パソコンやスマートフォンの普及が、目の疲れの直接の原因ではない。それらの機器を長時間使用する人がいて、更にはそのトコロの中に、目の疲労を感じる人がいる。ということであるはず。

9

5. 各特性の事例(2)

大分類	中分類	小分類	対象特許	悪例(品質特性に関わる箇所に下線)	良例(品質特性に関わる箇所に下線)	コメント・理由
技術文書特性	正確性	3(視力回復組成物)	[0036]本実施の形態において、アントシアニンは、その粒子径が100nm以下にナノ化されたものが好ましく、80nm以下にナノ化されたものであることがより好ましい。ナノ化することで即効性がより高まる。	[0036]本実施の形態において、アントシアニンは、その重量平均粒子径が100nm以下にナノ化されたものであることが好ましく、80nm以下にナノ化されたものであることがより好ましい。ナノ化することで即効性がより高まる。	「実施例」では、平均粒子径としている。実施例の記載に基づいて、平均粒子径と正確に一致すべき。更に、平均の定義が、重量平均なのか、数平均なのか、その他ののか、不明でない。平均の定義が不明でない結果、侵害疑義容易性も欠如。	
			2(タクシー)	基地局、顧客端末、および車載装置からなるタクシー播送システムにおいて、前記基地局は、前記顧客端末から受信した顧客特定情報と、前記顧客端末から受信した空車選択通知に対応する空車の車両特定情報とを対応させて記憶する仲介テーブルと、…確立する結合コネクションと、…を有することを特徴とする、タクシー播送システム。	基地局、顧客端末、および車載装置からなるタクシー播送システムにおいて、前記基地局は、前記顧客端末から受信した空車選択通知に対応する空車の車両特定情報と、前記顧客端末から受信した顧客特定情報とを、対応させて記憶する仲介テーブルと、…確立する結合コネクションと、…を有することを特徴とする、タクシー播送システム。	「対応する」、「対応させて」が、2回出て、どの要素が対応するか、不明確で解釈に迷う。対応させる要素の修飾関係を、「と」、「を」を用いて明確にする。
	2(タクシー)	前記顧客端末と前記車載装置との通信は、両者間の通信による通信、または電子メールによる通信であることを特徴とする、請求項3に記載のタクシー播送システム。	前記顧客端末と前記車載装置との通信は、両者間の、音声、データ、テキスト、グラフィックのいずれか1つ、または組み合わせによる通信であることを特徴とする、請求項3に記載のタクシー播送システム。	「通信」、「電子メール」は、限定が強い。一般的な技術で記述したほうが良い。		

10

5. 各特性の事例(3)

大分類	中分類	小分類	対象特許	悪例(品質特性に関わる箇所以下線)	良例(品質特性に関わる箇所以下線)	コメント・理由
権利文書特性	権利範囲特性	発明展索性	2(タクシー)	<p>【請求項1】 基地局、顧客端末、および車載装置からなるタクシー捕捉システムにおいて、前記基地局は、複数の空車の車載装置から受信したそれぞれの空車の位置情報を格納する空車位置データベースと、地図を記憶する地図データベースと、前記顧客端末から受信した顧客の位置情報に基づいて、該顧客の周囲に存在する一または複数の空車を前記空車位置データベースから特定して、地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末へ送信するサーバと、を有することを特徴とする、タクシー捕捉システム。</p> <p>【請求項2】 複数の空車の車載装置から受信したそれぞれの空車の位置情報を空車位置データベースへ格納し、顧客端末から受信した顧客の位置情報に基づいて、該顧客の周囲に存在する一または複数の空車を前記空車位置データベースから特定して、地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末に表示する、タクシー捕捉方法。</p> <p>* タクシー捕捉システムの請求項1と、タクシー捕捉方法の請求項2のみが存在するものとする。</p>	<p>【請求項1】 基地局、顧客端末、および車載装置からなるタクシー捕捉システムにおいて、前記基地局は、複数の空車の車載装置から受信したそれぞれの空車の位置情報を格納する空車位置データベースと、地図を記憶する地図データベースと、前記顧客端末から受信した顧客の位置情報に基づいて、該顧客の周囲に存在する一または複数の空車を前記空車位置データベースから特定して、地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末へ送信するサーバと、を有することを特徴とする、タクシー捕捉システム。</p> <p>【請求項2】 前記基地局は、車両の属性情報を記憶する車両データベースをさらに有し、前記サーバは、前記一または複数の空車に関する属性情報を前記車両データベースから読み出して、前記空車状況情報に付加することを特徴とする、請求項1に記載のタクシー捕捉システム。</p> <p>【請求項7】 タクシー捕捉システムの基地局であって、複数の空車の車載装置から受信したそれぞれの空車の位置情報を格納する空車位置データベースと、地図を記憶する地図データベースと、顧客端末から受信した顧客の位置情報に基づいて、該顧客の周囲に存在する一または複数の空車を前記空車位置データベースから特定して、地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末へ送信するサーバと、を有するタクシー捕捉システムの基地局。</p>	<p>下位概念の請求項、サブコンビネーションの請求項、他のカテゴリの請求項が必要である。</p>
			3(視力回復組成物)	<p>一方、実施例3では、アントシアニンに粒径100nm以下となるように高速回転衝撃粉砕機を用いてナノ化処理を施した後、ローヤルゼリーと混合した。なお、ナノ化処理後のアントシアニンの粒径を測定したところ、平均粒径が90nmであった。</p>	<p>一方、実施例3では、アントシアニンに粒径100nm以下となるように高速回転衝撃粉砕機を用いてナノ化処理を施した後、ローヤルゼリーと混合した。なお、ナノ化処理後のアントシアニンの粒径を測定したところ、重量平均粒径が90nmであった。重量平均粒径は、レーザー回折-散乱式粒度分布測定器(●●技製△△装置)を用いて測定した。</p>	<p>・平均の算出法、粒径の測定法が分からない。 ・請求項にパラメータが記載される場合、その測定条件を記載しておくが良い。 ・相対平均粒径？体積平均粒径？ ・球相平均径(光散乱)？2軸平均径？ https://fs.beckmancoulter.co.jp/column/particle/basic/03 https://fs.beckmancoulter.co.jp/column/particle/basic/01</p>
			3(視力回復組成物)	<p>アントシアニンとセサミン及び/又は梅エキスとを混合後に前記アントシアニン及び/又は前記梅エキスの相子径を100nm以下とするナノ化処理を行い、その後ローヤルゼリーをナノ化処理済み混合物に添加する工程を備えることを特徴とする視力回復組成物の製造方法。</p>	<p>アントシアニンとセサミン及び/又は梅エキスの混合物であって、前記混合物の相子径が100nm以下である混合物に、ローヤルゼリーを添加する工程を備えることを特徴とする視力回復組成物の製造方法。</p>	<p>＜一義性、明瞭性、強弱性、技術開示性＞ クレームの「平均粒径」が明確でないから、無効または無効の抗弁が成立した例有。(知財審) *ナノ化処理工程と、ローヤルゼリー添加工程がある。2つの工程が別々の事業者に行われてしまうと、関係侵害が成り立たない場合が想定され、仮に間接侵害が成り立つとしても、その立証は難しい。 *したがって、100nm以下とするナノ化処理工程については省略し、添加する相手として、粒径が100nm以下である混合物を単に規定すること</p>

5. 各特性の事例(4)

大分類	中分類	小分類	対象特許	悪例(品質特性に関わる箇所以下線)	良例(品質特性に関わる箇所以下線)	コメント・理由
ビジネス特性	ビジネス特性	自社ビジネスサポート性	2(タクシー)	<p>【請求項1】 基地局、顧客端末、および車載装置からなるタクシー捕捉システムにおいて、前記基地局は、複数の空車の車載装置から受信したそれぞれの空車の位置情報を格納する空車位置データベースと、地図を記憶する地図データベースと、前記顧客端末から受信した顧客の位置情報に基づいて、該顧客と予め決められた条件を満たすほど近くに存在する一または複数の空車を前記空車位置データベースから特定して、前記地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末へ送信するサーバと、を有することを特徴とする、タクシー捕捉システム。</p> <p>* 地図データベースとして、外部のデータベースサーバ(例えば、GoogleMap)を採用することとなった場合</p>	<p>【請求項1】 基地局、顧客端末、および車載装置からなるタクシー捕捉システムにおいて、前記基地局は、複数の空車の車載装置から受信したそれぞれの空車の位置情報を格納する空車位置データベースと、前記顧客端末から受信した顧客の位置情報に基づいて、該顧客と予め決められた条件を満たすほど近くに存在する一または複数の空車を前記空車位置データベースから特定して、取得した地図上に該顧客の位置と該一または複数の空車の位置とが表示される空車状況情報を作成し、該空車状況情報を該顧客端末へ送信するサーバと、を有することを特徴とする、タクシー捕捉システム。</p>	<p>外部の地図データベースを使用することとなった場合、システムの構成として、地図データベースを含めるべきではない。</p>
			4(自転車ライト)	<p>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交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、</p>	<p>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交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、</p>	<p>必須ではない発明特定事項(電源)は削除すべき。 「存在表示機能部材」は、冗長で長過ぎ。翻訳が容易ではない。</p>

6. 今後

- (1) 特許文書品質特性モデルの普及を目的としたテキスト作成
 - ・4分野の仮想的な特許文書を用いた定性的評価事例
 - ・4分野の仮想的な特許文書を用いた定量的評価事例
- (2) 特許文書品質特性モデルの普及に向けた方策の検討・実施
 - ・弁理士会の新人研修での上記研修テキストの使用の打診
 - ・日本知的財産協会等への研修の打診
 - ・パテント、日本知財学会、知財管理、特技懇などへの投稿
 - ・解説書の無料配布(冊子, PDF)

(4) 『特許ライティングマニュアル（改訂版）』

石附 直弥

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



「特許ライティングマニュアル（改訂版）」のご紹介

令和2年3月6日

一般財団法人 日本特許情報機構(Japio)
 特許情報研究所 調査研究部 研究企画課長
 石附 直弥

特許ライティングマニュアル

文
レ
ベ
ル

1. 短文にする

～短くシンプルな文にする～

・一文を短くして複雑な係り受けをなくすことにより、人間の理解が容易になる。

節・句
レ
ベ
ル

2. 省略しない

～隠れている要素がないか注意する～

3. 理解しやすい構成にする

～文の構造に注意する～

4. 横並びの要素の表現を揃える

～対等に並べ意味に注意する～

5. 読点を工夫する

～係り受けや文の構造を明らかにする～

6. 簡潔にする

～シンプルな表現にする～

・不要、冗長な表現は、簡潔な表現に改める。また、日本語特有の表現は、訳しやすい表現に改める。

語
レ
ベ
ル

7. 言い換える

～誰にでも伝わる表現にする～

・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。

ポスターと冊子で
改善の具体例とともに
紹介します！

- 2018年3月に改訂版を発行
- 7つのカテゴリー、27のルールに再構成
- 改善の具体例を豊富に掲載
- 27ページであり、コンパクト
- 特許のみならず、技術文書にも適用可能！

(5) 『システム開発文書品質研究会 (ASDoQ) の活動紹介』

栗田 太郎

ソニー株式会社／

システム開発文書品質研究会 幹事

システム開発文書品質研究会
Association of System Documentation Quality
ASDoQ 「アスドック」
<http://asdoq.jp/>

2020 年 3 月 6 日 (金曜日)
栗田 太郎 (ASDoQ 幹事 / ソニー)
taro.kurita@sony.com

研究会の目的

- ・システム開発, ソフトウェア開発における文書 (要求仕様書・設計書・取扱説明書等) の内容の検討, 記述, 「レビュー」, 参照, 活用, 品質等に関する研究
- ・開発現場における文書にまつわる課題に関する情報交換・意見交換・ネットワーキング
- ・システム開発に限らず, 自然言語の読み書きと文書の管理に関する, 基本的な事項の学習と啓発

ワーキンググループ (WG)

- ・文書品質モデル活用 WG
- ・文書品質測定 WG
- ・人材育成 WG

主な活動

- ・年 2~3 回の研究集会 (名古屋か東京で開催することが多い) や, 毎年恒例の夏合宿 (以前は長野, 最近は京都で開催することが多い) の開催. 主に会員が集まる
- ・「ASDoQ 大会」の開催. 会員・非会員問わず参加を募る (今年は 11 月 6 日 (金曜日) に名古屋大学にある野依記念学術交流館で開催する)
- ・他団体と協働する. 他団体のシンポジウム等に参加し, 交流する

システム／ソフトウェア開発文書の特徴

- ・多種多様な開発対象がある
- ・多種多様な組織，仕事の仕方（プロセス），文書の種類がある
- ・日本語・英語・図記法・表組・形式仕様記述言語等を組み合わせて文書を記述する．図表も含めて様々な固有の文法を持つ「言語」がある

ソフトウェアの作り方のかたち

- ・伝統的なソフトウェア（の作り方のかたちの一つ）：

仕様に基づいて作るプログラム	→	コンパイル	→	マシン語プログラム	→	テストによる検証	。	入力	→	マシン語プログラムの実行	→	出力	。
----------------	---	-------	---	-----------	---	----------	---	----	---	--------------	---	----	---

（演繹的）
- ・現代的なソフトウェア（の作り方のかたちの一つ）：開発と運用と改善（仕様変更）が統合される．文書よりもプログラムによるコミュニケーションを基本として，素早く開発する（アジャイル開発）
- ・AI（人工知能）はソフトウェアか？：

学習用データの準備	→	学習	→	学習済みモデル	→	テスト用データによる検証	。	入力データ	→	学習済みモデルを用いた判断・予測	→	結果	。
-----------	---	----	---	---------	---	--------------	---	-------	---	------------------	---	----	---

（帰納的）

開発現場の課題

- ・私の日本語・英語の読み書き聞き（聴き）話す能力は絶望的に低い
- ・「取説」がないプロダクト・ソフトウェア・「アプリ」・サービスの素早いリリースと，ステークホルダーとの会話・対話・議論に基づく不断のアップデート
- ・AI の仕様策定とその動作の説明責任

皆さまへのお願い

- ・ASDoQ 大会（11/6（金曜日）@名古屋大学）へのご参加，ポスター発表
- ・ASDoQ とのコラボレーションのご提案
- ・ASDoQ へのご入会 → <http://asdoq.jp/>

以上

(6) 『文章読解・作成能力検定（「文章検」）』

山田 乃理子

公益財団法人日本漢字能力検定協会 普及部 部長

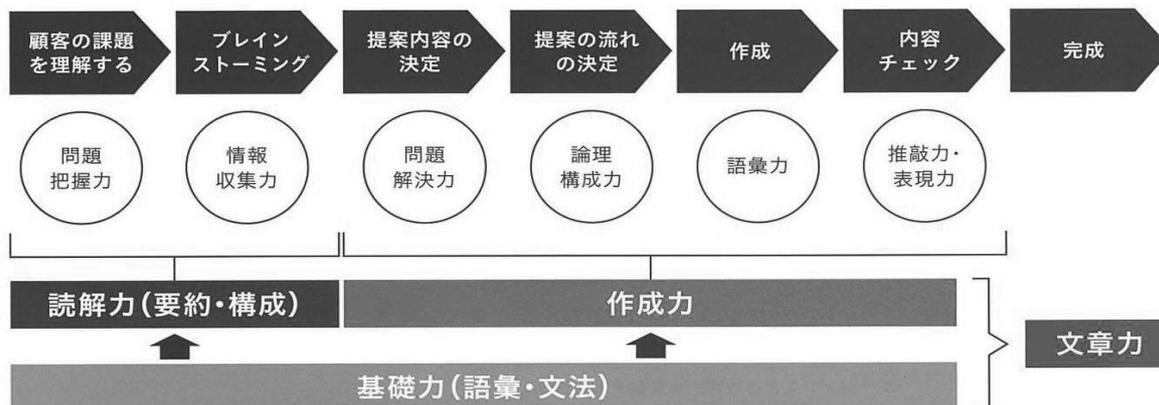
●文章検とは

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



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

●部門ごと・階層ごとの企業課題

部門ごとに異なるお悩み

営業部門

商談や案件の報告書を読んでも、個別に確認しなければボトルネックがわからず、上司が支援できない。顧客ニーズが多様化し、複雑な提案になる。

企画部門

さまざまなデータから分析した結果や考察を、社内外の関係者にわかりやすく説明しなければならない。

管理部門

全社への通達文が形式的になる。多様な職級・立場の社員を動かす説得力のある説明・対応ができない。

技術部門

専門分野を知らない相手にわかりやすく説明しなければならない。製品・システムの質だけでなく、ドキュメントの質も求められる。

階層ごとに異なる課題

①幹部層

経営会議では、担当事業の異なる各関係者が理解し納得できる報告・提案が必要。方針・戦略の意図を適切かつスピーディに管理職へ浸透させる。

②管理層

自分の若手時代とは異なる環境で、多様な価値観の若手社員を育成しなければならない。暗黙知の形式知化が求められる。

③若手層

わかりやすく的確な報告書やメールを書けるようになる。ビジネスにふさわしく失礼のない表現が求められる。

漢検

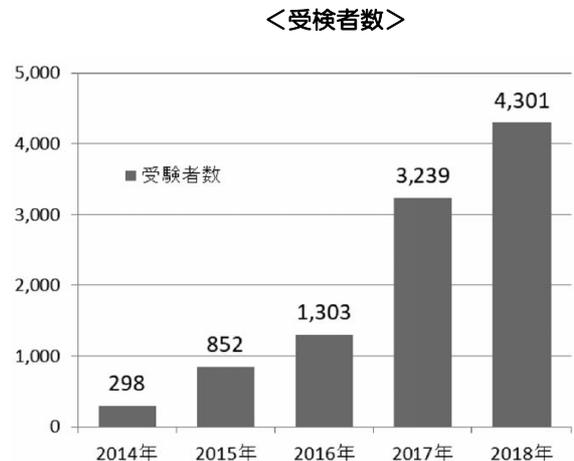
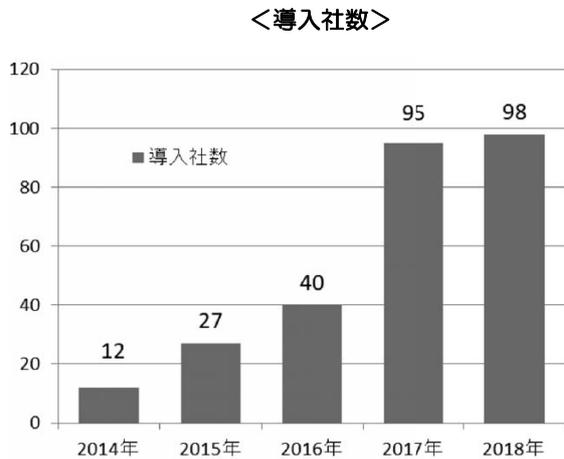
公益財団法人 日本漢字能力検定協会

THE JAPAN KANJI APTITUDE TESTING FOUNDATION

●企業 導入社数・受検者数推移

文章力への重要性や課題感の高まりを背景に、

直近5年で導入社数は8.2倍、受検者数は14.4倍と大幅に増加しています。



文章検
文章読解・字彙能力検定

漢検

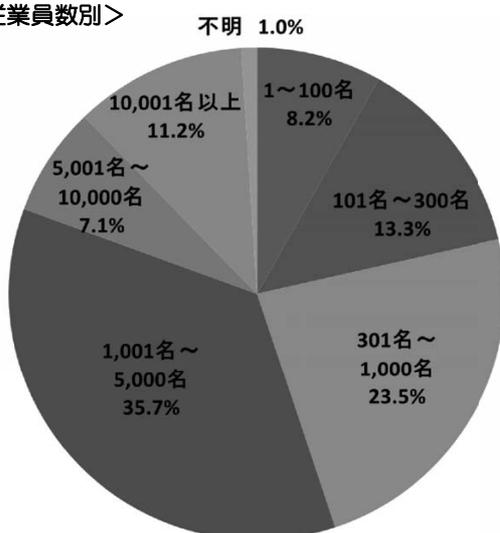
公益財団法人 日本漢字能力検定協会

THE JAPAN KANJI APTITUDE
TESTING FOUNDATION

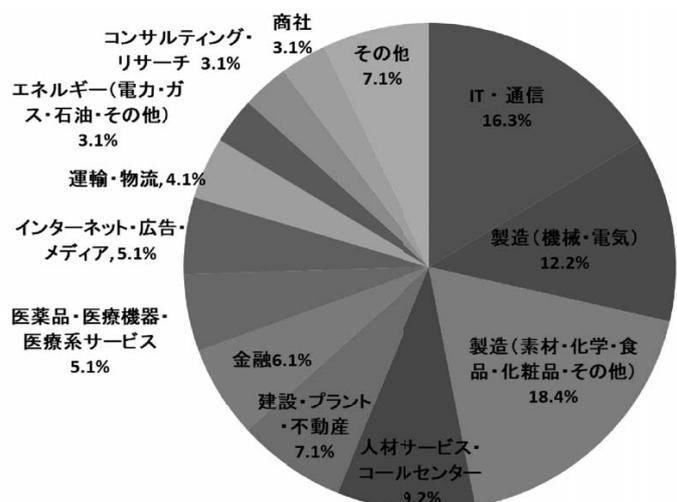
●直近のご導入企業の内訳

従業員規模・業種を問わず、様々な企業様に文章力向上の機会としてご活用頂いています。

＜従業員数別＞



＜業種別＞



※従業員数別・業種別円グラフはともに2018年度実績

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

(7) 『翻訳バンクでニューラル機械翻訳を多分野化する』

隅田 英一郎

国立研究開発法人情報通信研究機構 フェロー

翻訳バンクで ニューラル機械翻訳を 多分野化する

NICT
隅田 英一郎

2020/3/6

© NICT

1

Big Dataから深層学習で構築する ニューラル機械翻訳は高精度

Big Data

人類は、国家や地域、民族や世代など、あらゆる境界を越えて、相互の理解を深め、知恵を交わすなかで、発展してきました。コミュニケーションは人類社会を支えるもつとも重要な活動であり、情報通信技術はそのコミュニケーションを支える基礎であります。情報通信技術はまた、人類の高度な知的活動と経済活動を支える基盤でもあります。

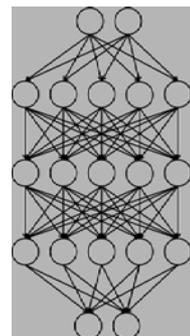
<http://www.nict.go.jp/およびhttp://www.nict.go.jp/en/index.html>

2020/3/6

Humanity has achieved progress as it has deepened its mutual understanding and shared its wisdom, overcoming barriers due to national, regional, ethnic, generational, and other differences. Communication is the most critical activity in human society, and information and communications technology (ICT) is the basis of that communication. ICT is also the infrastructure that supports humanity's advanced intellectual and economic activities.

深層学習

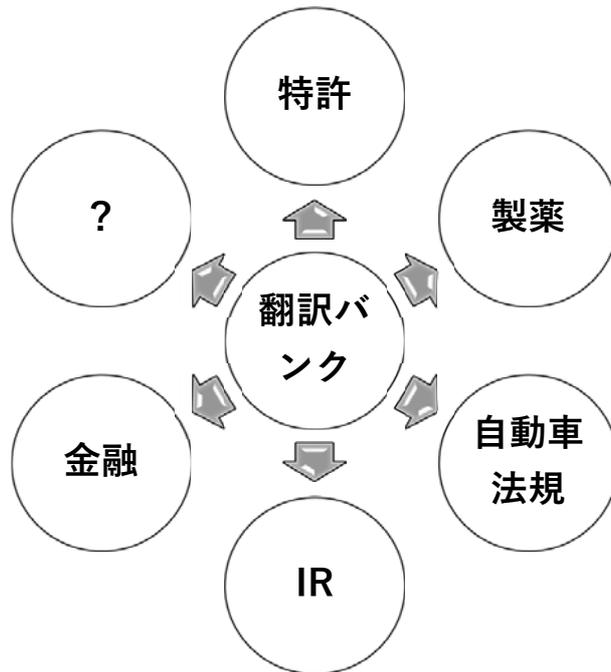
ニューラルネットワーク



© NICT

2

多分野化が加速



2020/3/6

© NICT

3

日興証券と金融特化型を開発

BLEU スコア	解釈
> 60	人が翻訳した場合よりも高品質であることが多い
50~60	非常に高品質で、適切かつ流暢な翻訳
40~50	高品質な翻訳
30~40	理解できる、適度な品質の翻訳
20~29	主旨は明白であるが、文法上の重大なエラーがある
10~19	主旨を理解するのが困難である
< 10	ほとんど役に立たない

13.1 点から
42.8 点に
改善!

入力	H形鋼と熱延鋼板に関しては、鉄屑市況の上昇分をカバーする水準まで上昇しそうだ。
汎用訳	The prices of H-shaped steel and hot-rolled steel are expected to rise to levels that will cover the rise in the steel scrap market.
金融特化訳	For H-beams and hot-rolled sheet, we expect prices to rise to levels that offset the rise in steel scrap prices.

2020/3/6

© NICT

4

(8) 『Japio 世界特許情報全文検索サービス紹介』

長谷川 雅昭

一般財団法人日本特許情報機構 営業推進部 主幹

世界の特許情報の日本語データベース (Japio-GPG/FX) と AI 翻訳

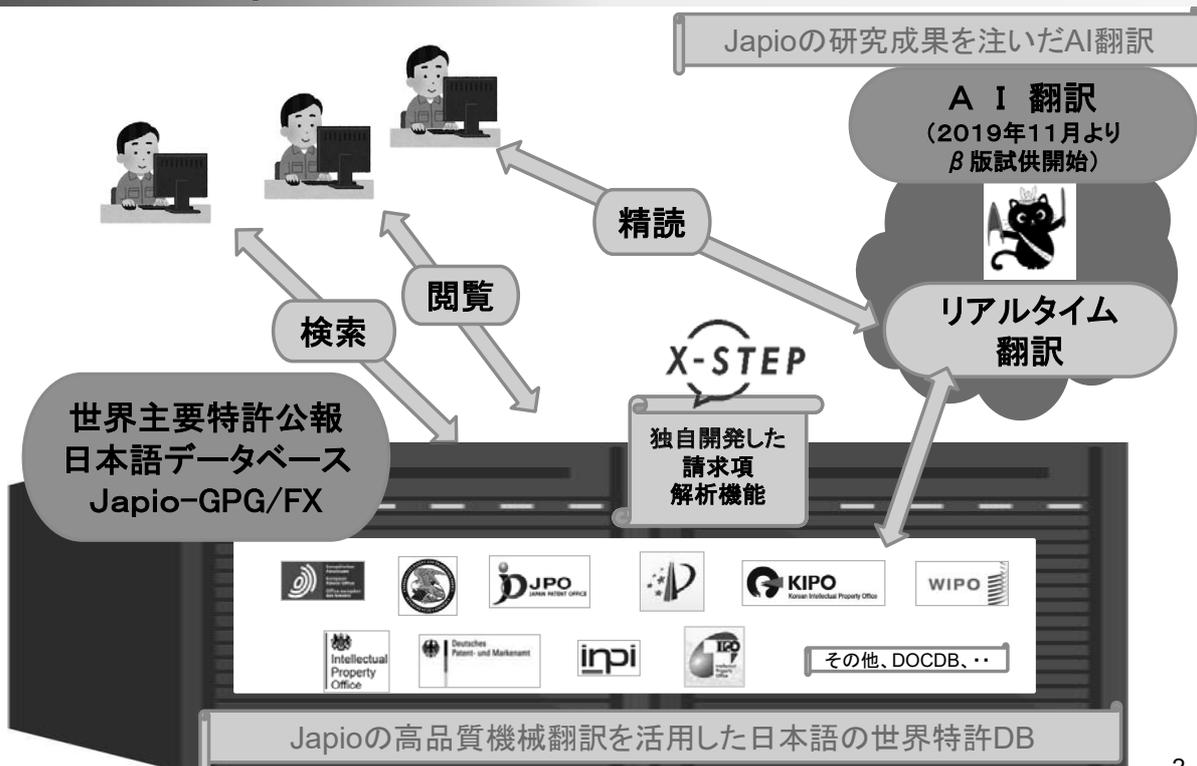
2020年3月6日



1

Japio-GPG/FX資料

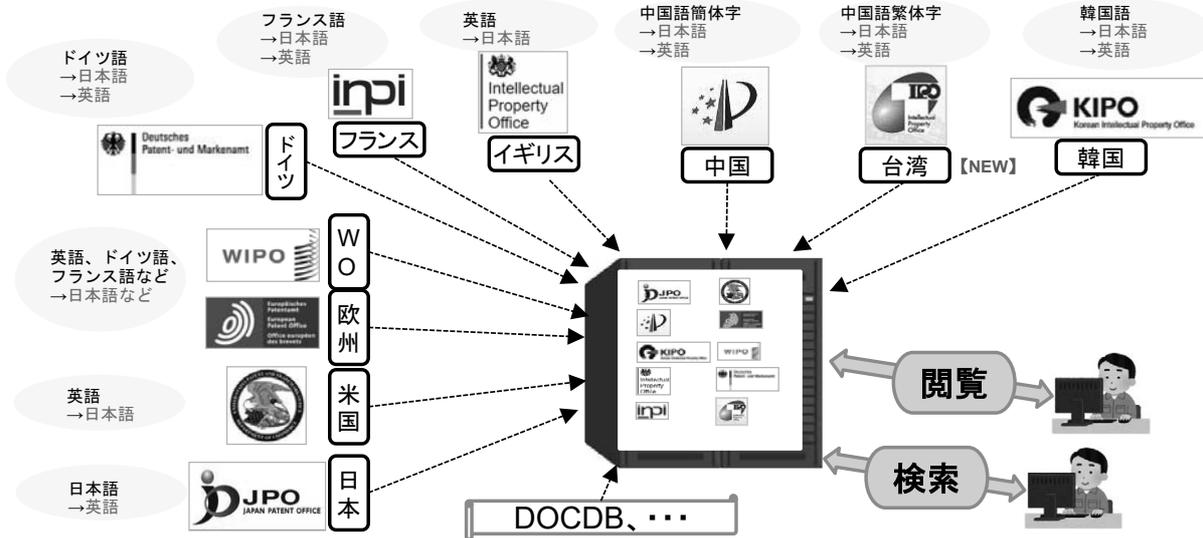
0. Japio-GPG/FXとAI翻訳 イメージ



2

1. Japio-GPG/FXとは

Japio世界特許情報全文検索サービス (Japio-GPG/FX) は、
 ・世界の国・地域/機関(日米欧、中韓、WO、ドイツ、フランス、イギリス、台湾)の
 ・特許/実用新案公報の全文を
 ・日本語(機械翻訳文)で蓄積・表示するDB



Japioの長年の研究成果等を活かし、高品質の機械翻訳文をご提供しています。

2. 対訳表示 ~原文に照らし合わせて日本語を確認~

世界の主要庁/機関の公報の、外国語文と日本語機械翻訳文とが対訳表示形式で参照可能です

➡外国文献の全文を日本語で参照でき、原文の箇所を容易に確認できます

各公報の機械翻訳文は、予め蓄積されています

➡その場で翻訳するシステムと比べ、翻訳文表示が速いです

英語/中国語/原語	日本語
<invention-title>	【C】 発明の名称
用于血球采集的抗凝剂控制系统	血球を採取するための抗凝剤制御システム
<abstract>	【C】 要約
本発明公開了一种用于血球采集的抗凝剂控制系统及方法。其包括抗凝剂泵、血泵、检测装置、第一计算装置和控制装置；利用检测装置检测血液中血球和红细胞比例，利用第一计算装置计算抗凝剂与全血的混合比，再利用控制装置控制抗凝剂泵和血泵的转动。本发明通过对每个献浆员的血球与红细胞比例进行检测，计算得到个性化的抗凝剂使用量，实现了抗凝剂的最低使用目的，同时通过对血球采集过程中使用的耗材特征值进行实时测量，消除了因为不同耗材差异造成的抗凝剂使用增加的问题，提高了采集血球的质量。	血球を採取するための抗凝剤制御システムおよび方法を開示している。その抗凝剤ポンプ、血液ポンプと、検出装置と、第1の計算装置と制御装置を含む検出装置を用いて血液中の血球と赤血球の割合を検出し、第1のコンピューティング装置を抗凝剤と全血計算の混合比により、更に制御装置を制御する抗凝剤ポンプおよび血液ポンプの回転を利用する。本発明各献漿員の血球と血液細胞の割合を検出することにより計算個別化された抗凝剤使用量を得る、抗凝剤の実現した最低使用目的である。同時に血球採取手順に使用される消耗品特徴量をリアルタイムで測定する、異なる消耗品差に起因する抗凝剤を用いるが増加してしまうという問題が解消され血球採取の品質を向上させる。

※ドイツ語やフランス語の公報の場合、英語機械翻訳文と日本語機械翻訳文の対訳表示です
 原文は、ページ下部に一括表示されます
 (なお、フランス国公報の場合、フランス語原文が表示されるのは概ね2019年以降に発行された公報です。)

3. GPG/FXでの翻訳方式

2019年7月現在

統計翻訳(SMT) + 前処理(Japio独自) + 後処理(Japio独自)

	サービス初期の方式	現在の方式
米国 ⇒ 日本語	RBMT※1	SMT (NICTエンジン 英日) ※3
日本 ⇒ 英語	RBMT※1	RBMT※4⇒※8
EP/PCT ⇒ 日本語	RBMT※1	SMT (NICTエンジン 英日) ※3
中国 ⇒ 日本語	SMT (NICTエンジン 中日) ※2	SMT (NICTエンジン 中日v2) ※5⇒※9
台湾 ⇒ 日本語	—	SMT (NICTエンジン 中日v2) ※9 + 台日データ
韓国 ⇒ 日本語	RBMT ^{韓日} ※特許庁の中韓文献翻訳システム由来	SMT (moses 韓日) ※7
独国 ⇒ 日本語	独国の英語データを RBMT※1	SMT (moses 独英) ※6⇒※10 + SMT (NICTエンジン 英日) ※3
仏国 ⇒ 日本語	仏国の英語データを RBMT ※1	SMT (moses 仏英) ※11 + SMT (NICTエンジン 英日) ※3

※1 … 250万語 (36分野) ⇒

※2 … 1,600万文対(中日) ⇒

※3 … 1億文対(英日)

※4 … 517万語 (36分野) ⇒

※5 … 5,200万文対(中日) ⇒

※6 … 2,000万文対(独英) ⇒

※7 … 1,200万文対(韓日)

※8 … 732万語 (36分野) ⇒

※9 … 8,800万文対(中日) ⇒

※10 … 4,000万文対(独英) ⇒

※11 … 2,500万文対(仏英)

補足: RBMT: ルールベース翻訳 SMT: 統計翻訳

5

4. 高品質の日本語機械翻訳文を蓄積

1. 技術用語の翻訳精度が高い

機械翻訳が苦手とする専門用語に加え、韓国語の同音異義語、ドイツ語の複合語などにも対応。

原文	1,4-dihydronaphthyl	열박리	Distraktionsvorrichtung
旧世代の翻訳	1,4つのジヒドロナフチル	劣薄里	Distraktionsvorrichtung (翻訳できず)
Japio-GPG/FX	1,4-ジヒドロナフチル	熱剥離	伸延装置

2. 訳文の自然性が高い

特許分野で使われている書き方に沿った自然で読みやすい翻訳文。

3. 請求項が理解しやすい (Japio独自の特許翻訳向け新フレームワーク)

独自開発した請求項解析機能などにより、請求項が更に読みやすくなった翻訳文。

X-STEP

原文	4. Verfahren nach Anspruch 1, dadurch gekennzeichnet, dass die Erhöhung auf den vorgegebenen dritten Strom-Sollwert (I3) kontinuierlich geschieht.
旧世代の翻訳	4. 請求項1に記載の方法、特徴をなされた、所定の3番目の電流の公称値(i3)への増加が連続的に生じる。
Japio-GPG/FX	【請求項4】所定の第3の電流目標値(I 3)に増加が連続的に実行されることを特徴とする請求項1に記載の方法。

※請求項解析機能は、米国公報、韓国公報、ドイツ国公報、EP/PCTの英語及びドイツ語公報などが対象です。
※旧世代の翻訳とは、ルールベース機械翻訳であり、Japio-GPG/FXにおけるバージョンアップ前の翻訳結果です。

6

5. 翻訳Webサービス (AI翻訳版) (β版)

特許公報に特化した高精度なAI翻訳機能を、β版として2019年11月1日より試供開始。

※画面は開発中のものです。実際の画面とは異なる場合もあります



Japio-GPG /FX **翻訳Webサービス (AI翻訳版)**

Excel 出力 中断
最初 前へ 1 次へ 最後

書誌情報

出願番号 CN201780087724.6
 文献番号 CN110352627A
 発明の名称 显示装置及其制造方法
 出願人 夏替株式会社
 IPC H05B33/04[20060101]; G09F9/30[20060101]; H01L27/32[20060101]; H01L51/50[20060101]; H05B33/06[20060101]; H05B33/10[20060101]; H05B33/12[20060101]; H05B33/22[20060101]

発明の名称	発明の名称 (訳文)
显示装置及其制造方法	表示装置及其制造方法
要約	要約 (訳文)
有機EL显示装置(1)はTFT基板(10)上の表示区域(5)の外側、沿断設置有窓封膜(30)の第一无机层(31)及第二无机层(33)の境界、設置具有突出部(41a)の有機体(41)、第一无机层(31)及第二无机层(33)層叠突出部(41a)、并且在突出部(41a)の下方、第2无机层(41)の境界而设置。	有機EL表示装置1は、TFT基板10上の表示領域5の外側において、封止膜30が設けられた第1無機層31及び第2無機層33の境界に沿って、突出部41aを有する有機体41が設けられ、第1無機層31及び第2無機層33が突出部41aを覆うとともに、突出部41aの下方において基板10の境界に集積して設けられている。
請求項	請求項 (訳文)
1. 一种显示装置，其包括支撑体、以及设置于所述支撑体上的多个发光元	1. 表示装置であって、支撑体と、前記支撑体上に設けられた複数の発

画面右上の翻訳WEBサービス(AI翻訳版) (右図) をクリックすると別Windowで照会画面を開き利用できます。詳細画面の「AI翻訳」ボタン(左図)では、表示中の文献に対してAI翻訳が行われません。(2019年11月1日ご試供開始)

6. 翻訳Webサービス (AI翻訳版) (β版)

(ポイント1) 更に向上した翻訳品質

構文の正確性や流暢さが飛躍的に向上

独自開発の言語資源とノウハウにより、特許特有表現も更に読みやすく

(ポイント2) 多言語対応

外国語から日本語へ: 米国公報、中国公報、台湾公報、ドイツ公報など

日本語から英語へ: 日本公報など

翻訳例

【英語】 2. The AR computing device of claim 1, wherein the AR computing device is further configured to assign an identifier to the first consumer.

【AI翻訳】 2. ARコンピューティングデバイスは、第1の消費者に識別子を割り当てるようにさらに構成されることを特徴とする請求項1に記載のARコンピューティングデバイス。

【中国語】 并在真空下浓缩，得到产物，用柱层析法(10% MeOH / CHCl₃)纯化产物(89%)。

【AI翻訳】 真空下で濃縮して生成物を得、これをカラムクロマトグラフィー(10% MeOH/CHCl₃)で精製した(89%)。

7. 翻訳Webサービス (AI翻訳版) (β版)

The screenshot displays a web interface for patent translation. On the left, there is a search filter menu with 'リチウム' (Lithium) and '二次' (Secondary) selected. The main content area is titled '書誌情報' (Bibliographic Information) and contains the following details:

- 出願番号: CN201880014258.3
- 文献番号: CN110352525A
- 発明の名称: 二次電池及其製造方法
- 出願人: 日本電気株式会社
- IPC: H01M10/0565[20060101]; H01M2/02[20060101]; H01M4/02[20060101]; H01M10/04[20060101]; H01M10/0567[20060101]

Below this, there are two columns for '発明の名称' (Invention Name) and '発明の名称 (訳文)' (Translated Invention Name), both showing '二次電池及其製造方法'. The '要約' (Abstract) section contains Japanese text on the left and its English translation on the right. The '請求項' (Claims) section also shows Japanese text on the left and its English translation on the right. The interface includes a 'ハイライト' (Highlight) button and a search bar at the top right.

翻訳後の画面例

2020年1月以降、AI翻訳β版は有料オプションユーザー様へのご提供。
(オプション料金: Japio-GPG/FX提供料金の10%)

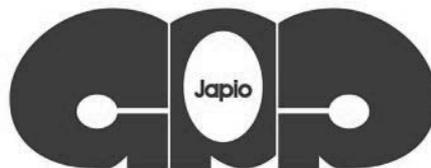
※ AI翻訳β版は、正式提供時にサービス内容、料金等が変更される事があります。

ありがとうございました

一般財団法人 日本特許情報機構

お問合せ先

一般財団法人日本特許情報機構 サービス窓口 03-3615-5510 service@japio.or.jp
株式会社発明通信社 事業企画部 03-5281-5511 info@hypatweb.jp



Japio 世界特許情報全文検索サービス
(Japio-GPG/FX)

【第三部】

講演

『開発現場における暗黙知・属人知の AI 技術活用に向けた取り組みのご紹介』

自動車産業をはじめとする製造業の開発現場においては、製品の複雑化・高度化に伴いエンジニアの専門分野の細分化・分業化が進み、設計品質のレビュー依存度・ベテラン指摘への依存度が日々高くなっていく傾向にあります。このことは設計工程における生産性のネックになっており、熟練エンジニアの高齢化・定年等が進む日本において、部署・組織を跨いだ知識・知見の活用が重要な課題となります。

本研究開発では、ベテランのノウハウを人工知能技術を適用したしくみで再現し、設計業務におけるエンジニアの判断支援に活用することで、この課題解決に貢献すると共に、企業・分野を跨いで展開可能な基盤技術を獲得することを目的としています。

西田 公祐

SOLIZE 株式会社 SOLIZE Innovations カンパニー Senior Manager

西鳥羽 二郎

株式会社レトリバ 取締役/CRO

西村 拓一

産業技術総合研究所 人工知能研究センター

サービスインテリジェンス研究チーム長

開発現場における暗黙知・属人知のAI技術活用に向けた 取り組みのご紹介

次世代人工知能・ロボットの中核となるインテグレート技術開発／
②人工知能技術の適用領域を広げる研究開発／
熟練者観点に基づき、設計リスク評価業務における判断支援を行う人工知能適用技術の開発

2020.3.6

SOLIZE株式会社
株式会社レトリバ
国立研究開発法人 産業技術総合研究所



研究開発の背景・狙い

設計領域の現状と課題

自動車産業をはじめとする量産産業では、
商品ニーズの変化・多様化に即時対応を可能とする「商品開発スピード」と「柔軟な調整力」の獲得が
重要であるが、開発現場は下記課題に直面している。

- ✓ エンジニアの専門分野の細分化・分業化が進み、
各技術者の視野・技術・経験は限定的にならざるを得ない
- ✓ 商品開発全体を俯瞰して見渡せるような有識者が育ちづらい
- ✓ 高齢化・定年等により国内の有識者が減少

品質担保のためのレビュー依存度・ベテラン指摘依存度は増々高くなる傾向にあり、
「レビュー時間増加」「検討漏れによる手戻り発生」「開発期間の長期化」
といった問題が発生

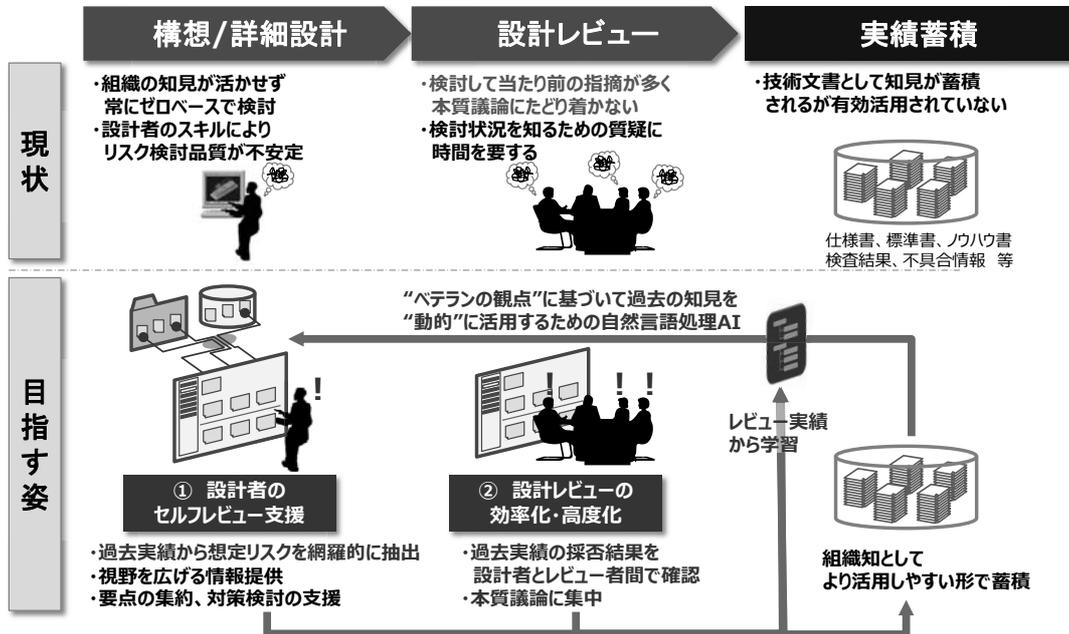
本研究開発の狙い

- 製品開発全体を見渡しながら開発リスクを抽出し、
部門をつないで問題解決を行ってきたベテランの暗黙知を技術・しくみにより再現
- 先人のノウハウは、設計成果物（図面）の背景にある技術文書等に凝縮（但し、自然言語）
- 担当者毎・企業毎の膨大な知識を自然言語処理を用いて、集合知として活用

研究開発の背景・狙い

設計リスク評価業務における判断支援コンセプト

- ・ 本システムは、自動化ではなく、設計者に判断材料となる知識情報を提案する
- ・ 大量の過去知見の有効活用を支援し、設計段階における検討品質を向上させ、過去同様の当たり前指摘に割いていたレビュー時間を効率化し、本質的な課題議論を行う時間へと高度化

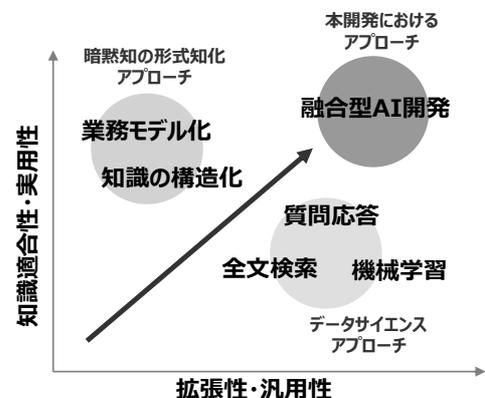


研究開発の概要

研究開発の方向性

ベテランの思考パターンや観点を形式知化した知識ベースに、自然言語処理・機械学習などのデータ分析的な検索技術によって大量のデータを紐付けるという、二つの推論の融合型AI開発によって適切な情報提供を実現する技術開発を行う

本開発	融合型人工知能適用技術	<ul style="list-style-type: none"> ・従来型の2つのアプローチを組み合わせる ・熟練者の判断に着目して業務を紐解き、その観点を知識ベースとしてデータ化、推論ロジックを人工知能技術で学習モデル化し、熟練者の判断を再現するAIを構築 ・知識適合性と拡張性の両立を狙う
従来型	暗黙知の形式知化アプローチ	<ul style="list-style-type: none"> ・熟練設計者の暗黙的な判断や思考パターンを知識構造として形式知化・モデル化し、根拠や関連知識を紐づける。業務への知識適合性が高い。 ・知識構築のコストが高く、更新が困難で陳腐化しやすい課題あり
	データサイエンスアプローチ	<ul style="list-style-type: none"> ・自然言語処理などで大量情報の再利用性向上や知識獲得をする ・高度な意味解釈が困難であり、必ずしも有効な知見は得られない

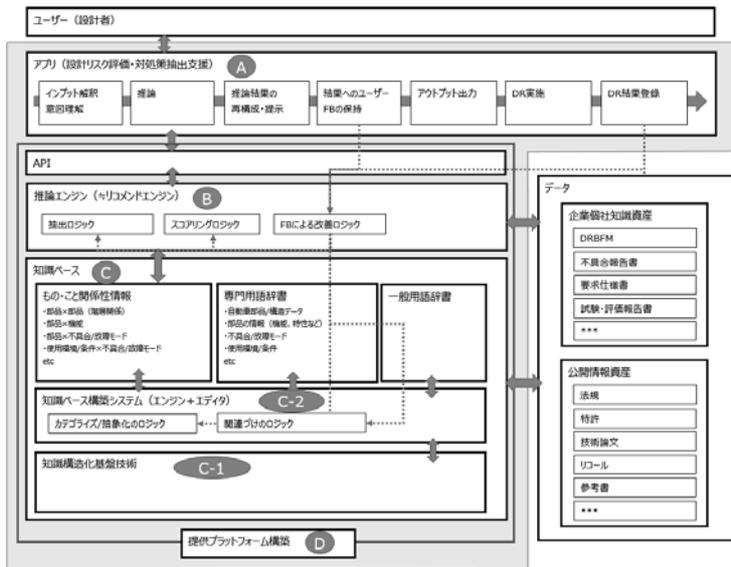


- ・「熟練者の業務における判断に着目した知識ベース」と「人工知能技術」を融合させる研究
- ・実用性と汎用性を兼ね備え、企業・分野を跨いで展開可能な基盤技術獲得をねらう

研究開発の概要

システム構成の全体像

- 設計業務は、より多岐に渡る判断や連続的な判断が求められるなど、複雑な思考を要する業務
- 設計に関するノウハウを「ベテラン観点」で構造化した知識ベースを用いて、推論エンジンによる問い合わせ意図の解釈や関連情報の抽出を強化することでこれを解決する方針



各研究開発テーマと重点ポイント

A: 「設計リスク評価支援のアプリケーション」構築

推論エンジンおよび知識ベースに基づき、設計者の問合せに応じて、リスク対処策抽出の判断支援を行うためのアプリケーションの構築

B: 「推論エンジン」の開発

知識ベースの情報を利用して、問合せ内容に対し、特徴量（キーワード）の抽出と推論補充、関連情報の抽出とスコアリング、フィードバックによるロジック改善を行う推論エンジンの構築

C: 「知識ベース」の開発・構築

設計者の知識・経験を、知識工学等を活用して形式化・構造化し、人工知能を適用したシステムで活用できる形式に変換するための方法論及びこれに基づく知識ベースの構築

C-1: 「知識構造化基盤技術」の開発

知識ベース構築の基盤となる、構造的な知識表現、オントロジーや辞書構造の持ち方、データ間の関連性のつけ方について、産業分野で広く適用可能な基本アーキテクチャ・仕様としての定義・設計

C-2: 「知識ベース構築システム」の開発

知識構造化基盤技術に基づいた知識ベースの構築手法の開発と知識ベース構築コスト低減を図るための知識ベース構築システムの構築

D: 「提供プラットフォーム」構築

先導開発において実証された技術を結合して、アプリケーションとしての実用化・実装および、BおよびCパートの成果をAPIもしくはライブラリとして提供するプラットフォームの構築

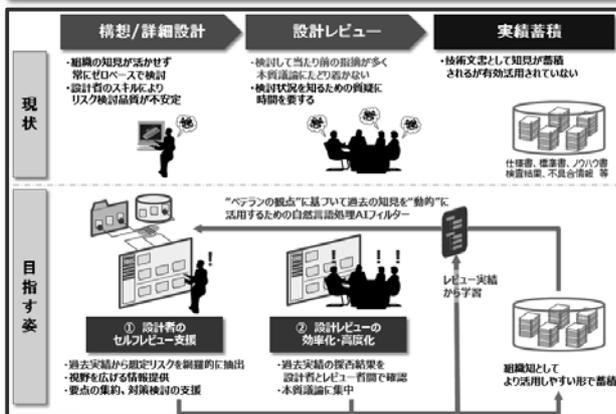
研究開発フェーズ

FY19-20 : 先導開発フェーズ

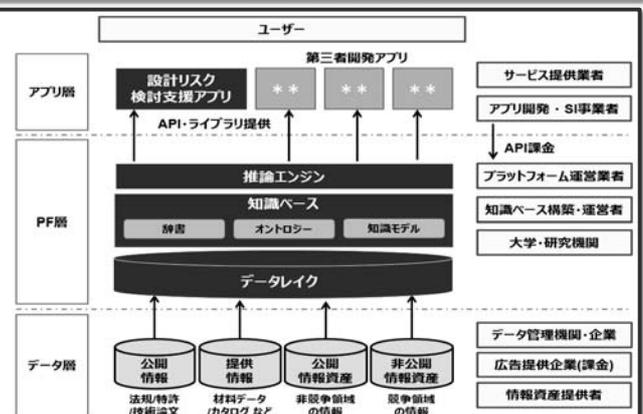
設計不具合の未然防止手法として製造業に広く浸透しているDRBFMの実績資料を中心とした、設計リスク評価支援システムの先導開発・有効性実証

FY21-23 : 実用化・拡張開発フェーズ

・支援アプリ実用化・提供プラットフォーム構築
 ・企業・産業分野を跨いだ適用領域拡大の開発
 推論エンジンや知識ベースを企業を跨いで活用してもらうための汎用化開発を行う



設計リスク評価業務における判断支援コンセプト



実用化・事業化後のイメージ

付記

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の委託業務「次世代人工知能・ロボットの中核となるインテグレート技術開発」の結果得られたものです

講演

『ニューラル機械翻訳と特許ライティングマニュアルを用いた日本語原稿のプリエディット』

日本知的財産翻訳協会 (NIPTA) は、2017 年から、特許機械翻訳研究会を作って、ユーザーの立場から現在のニューラル機械翻訳 (NMT) は使えるのか、どう使えばいいのかを特許ライティングマニュアルの活用を含めて検討してきた。2 年弱にわたるその活動の結果得られた成果を報告したい。2016 年 11 月に Google 翻訳が一気に改善され、機械翻訳がそれまでとは異なるステージに乗ってきたが、その後も改良は続いている。その流れの中で、日本の翻訳文化の在り方も試されている。今後を考えるための基礎となる成果を提供したい。

奥山 尚一

日本知的財産翻訳協会 理事長

『ニューラル機械翻訳と特許ライティングマニュアルを用いた日本語原稿のプリエディット』

発表者：NPO 法人日本知的財産翻訳協会 (NIPTA) 理事長 奥山尚一

ニューラルネットの活用により機械翻訳の精度は大きく向上しました。それをユーザーの立場から検証・検討するために、NIPTA は特許機械翻訳研究会を立ち上げました。

一年目は直球勝負で、機械翻訳の実情と向き合いました。そのとき、これだけ機械翻訳の精度が高まってきたなら、産業日本語もその実力を発揮できるようになってきたのではないかと直感しました。

そこで、二年目の課題として、定点観測（同じブランドのアルゴリズムを使って、性能の変化を計測する）のほかに、産業日本語を使ってプリエディットしてみて、それが有効であるか、翻訳効率が全体として上がるのかを試してみました。直接のかかわりを持ったことはありませんでしたが、以前より産業日本語研究会の活動には関心を持っており、第1版を大きく改良した「特許ライティングマニュアル（第2版）」を利用すると面白い結果が出るのではないかと考えました。

その結果を報告します。

本報告は2部からなります。第1部のとりまとめは、株式会社翻訳センターの渡部孝明氏にお願いしました。第2部のとりまとめは、ブログ「みんなのワードマクロ」管理人の新田順也氏と（一社）日本特許情報機構（JAPIO）の知財 AI センター主幹の清藤弘晃氏にお願いしました。

第1部として NMT の定点観測の結果を説明します。つまり、特許翻訳分野における現在の翻訳エンジン性能を1年前の性能と比較してみて、変化を把握しました。

第2部は、機械翻訳出力結果の利用方法の提案として、ポストエディットとプリエディットの有効性を確認しました。

新田氏には、翻訳者の視点から検討していただき、清藤氏には、明細書作成者の視点から、特許ライティングマニュアル（第2版）の活用を踏まえて、検討していただきました。

検討の結果を振り返ると、最近もトランスフォーマーといった NMT の技術革新があったものの、日本語と英語の間には、越えがたいギャップがあって、定点観測の結果を改良された部分と改悪されてしまった部分とが混在する状況です。

日英翻訳を縦のものを横にするとよくいいますが、それは韜晦であって、主語や目的語の語順が、動詞の位置が違うので相対的にまったく反転する場合があるなど、なかなか機械翻訳だけでは処理できないものがあります。それは、純粋に言語構造の違いなのか、ある対象をどうとらえるかという文化的な違いなのか、日本語と英語の間ではヨーロッパ言語の同士の間にあるのとは異なる難しい課題が横たわっています。

日本語は類縁言語がありません。1万3千年ほども続いた縄文時代に日本語は完全に独自の発展をとげたのでしょう。それを基礎としている文化と精神世界は重要というか、我々の宝です。世界中の言語を見ても類のない、完璧な表音文字である平仮名とカタカナ（最近の悲しい事件が物語るように5歳児でも喋ることをそのまま文字にできる）、2000以上の漢字、ローマ字まで駆使する世界的に比肩するものがまったくない圧倒的な優位性を持つ日本語の魅力を維持しつつ、世界に発信できる普遍性のある日本語というのも考えなければいけないときに来ていると思います。

ある意味、江戸後期から明治にかけて、西洋の言葉と概念を須らく日本語化し、言文一致の日本語を構想して実現した人たちに近い岐路に我々は立っているのではないのでしょうか。

面白い時代になったと思いますが、ここで間違ふと今後何十年にもわたって取り返しのつかない禍根を後世に残してしまうかもしれません。真剣に取り組まなければいけないと思います。

講演

『自然言語処理によるニュース記事執筆の現状と課題』

自然言語処理では、古くからニュース記事をコーパスとして活用してきたが、近年では記事の制作に自然言語処理を活用する事例が増えてきている。本講演では、自然言語処理技術でソーシャルメディアの投稿や国会議員の発言を分析する事例を紹介する。続いて、自然言語処理で記事の執筆を支援する取り組みとして、データから記事を自動的に生成する自動化報道、記事本文からの見出し・要約の生成、記事の校正履歴に基づく自動テキスト校正などの事例を紹介する。

岡崎 直観

東京工業大学 教授

第11回産業日本語研究会・シンポジウム

自然言語処理による ニュース記事執筆の現状と課題



東京工業大学
Tokyo Institute of Technology

岡崎 直観

東京工業大学 情報理工学院

okazaki at c.titech.ac.jp

<https://www.nlp.c.titech.ac.jp/>



OKAZAKILAB

自然言語処理：人間のように言葉を理解・生成できるロボット



機械翻訳

ある言語の文章を
別の言語に翻訳する



質問応答

自然言語で与えら
れた質問に答える



対話エージェント

コンピュータと人間
の間で会話をする



自動要約

情報を集約して
文章を生成する



情報検索

大量の文書の中から
必要なものを探す



情報抽出

文章から事実や
知識を抽出する



意見・感情分析

文章から人間の主観
的評価を抽出する



教育支援

人間が良い文章を
書くように支援する

ニュース記事と自然言語処理

- ニュース記事をコーパスとして利用
 - 正しい日本語の「お手本」に統語解析・意味解析の「お手本」を付与
 - 例: 京都大学テキストコーパス (1995年版毎日新聞)
- **データジャーナリズム**: テキストを分析してニュースの種を得る
 - 様々なテキストデータをマイニングして, 社会を観測する
- **ロボットジャーナリズム**: 自動もしくは半自動で記事を生成する
 - 記事の自動生成
 - 記事の自動校正
 - 見出しの自動生成

2

データジャーナリズム

3

データジャーナリズム (data-driven journalism, computational journalism)

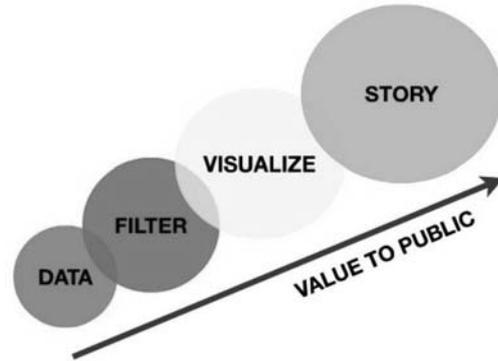
データを収集, 整形, 整理, 分析, 可視化, 公表することで, 新しいニュースを発見したり, ニュースを分かりやすく伝えようとする報道

データ処理による記事化

- データ: データを収集して構造化
- フィルタ: データから知識を抽出
- 可視化: グラフィックなどで表現
- ストーリー: 分析結果から記事のストーリーを作る

株価のデータを載せただけの記事はデータジャーナリズムとは呼ばない

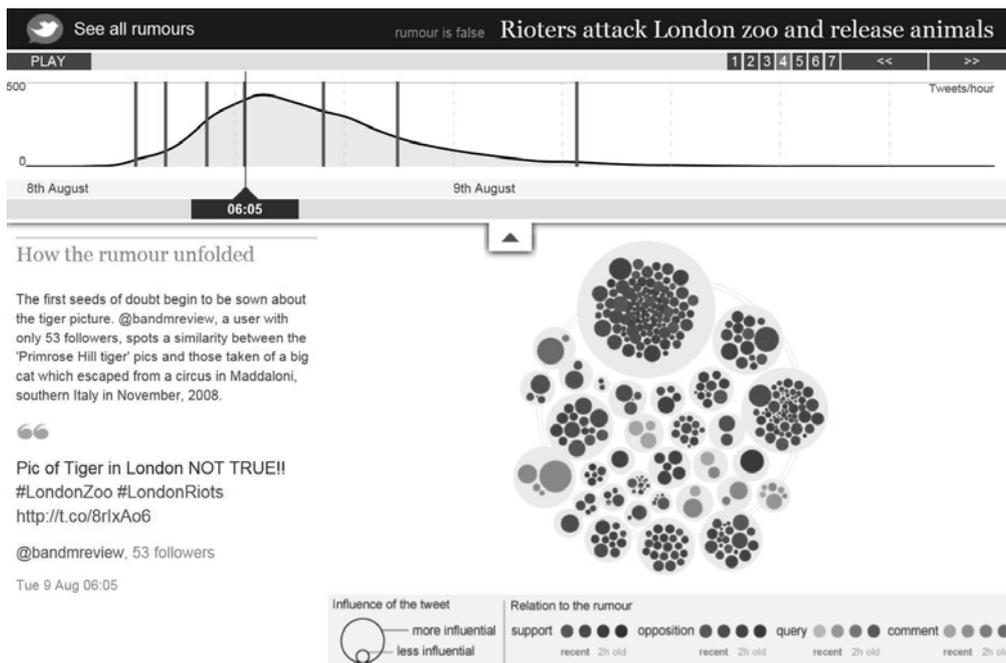
DATA-DRIVEN JOURNALISM = A PROCESS



データジャーナリズム (Lorenz, 2010)

Mirko Lorenz. 2010. Data driven journalism: What is there to learn?
http://mediapusher.eu/datadrivenjournalism/pdf/ddj_paper_final.pdf

Reading the Riots (Guardian)



<https://www.theguardian.com/uk/interactive/2011/dec/07/london-riots-twitter>

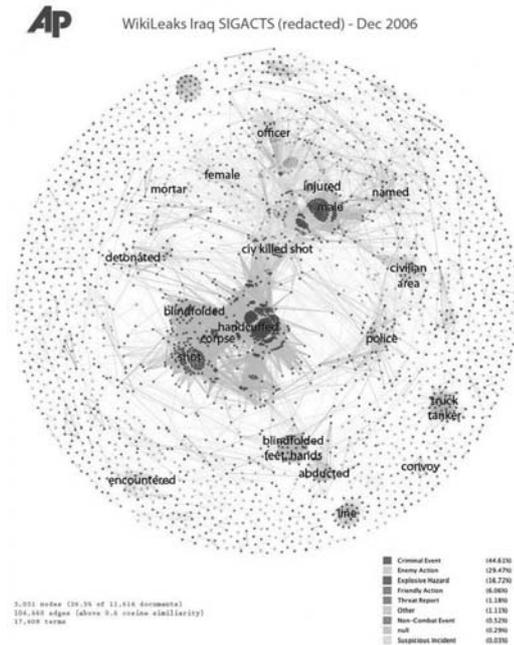
Full-text visualization of the Iraqi War Logs (AP)

分析の狙い

- WikiLeaksが公開したイラク戦争に関する機密文書をネットワーク可視化

分析方法

- 各文書の単語をTF*IDFで重み付け・ベクトル化
- 各文書を重みの高い3つの単語で表現
- 文書をノード、文書間のコサイン類似度でエッジを作成
- ノードの色は事件タイプ
- Gephiでネットワーク可視化



<http://jonathanstray.com/a-full-text-visualization-of-the-iraq-war-logs>

6

Message Machine (ProPublica)

7 People in our sample got this email vs. EMAIL 1 *****

EMAIL 3

About This Group

Gender: F M

Age: 17 74

Previous Donor? YES NO

Requested Donation: \$3 \$300

Subject Line	# Emails
Janet just found out	6
You must be at least a little curious	1

The content of this message is similar to Email 1, but the phrasing is different. Some recipients were asked for \$3, others \$25, and one was asked for \$190.

EMAIL 1 EMAIL 2 EMAIL 3 EMAIL 4 EMAIL 5 EMAIL 6

"You're going to have dinner with the President."

There are only a handful of people who will. You must be at least a little curious, probably don't know Janet from Accokeek, Maryland, Maryland.

But you two other supporters for dinner could be having with four supporters. We're dinner with President Obama, Janet, and Obama together sometime soon.

Janet was the first guest to be selected for the upcoming dinner the President's next Dinner with Barack. And we're counting down the hours until we pick the next winner. Will you be the next person who will hear that sentence?

Pitch-draw the next name.

The second seat could be yours -- you'll be automatically entered for the chance to join when you donate \$XX or whatever you can to help today.

Janet told us that when she donated \$XX last week she was just became trying to show her support this campaign today, and for the President.

No one of them:

She learned she's over hear that sentence, thinks they'll be picked until they are.

Take the chance. Chip in \$X-\$XX or more today to be automatically entered:

<https://donate.barackobama.com/The-Second-Guest>

Thanks,

Julianna

Julianna Smoot Deputy Campaign Manager Obama for America

<http://www.propublica.org/special/message-machine-you-probably-dont-know-janet>

7

東日本大震災ビッグデータワークショップ Project 311

- 概要
- 参加申し込み
- 提供データ
- スケジュール
- ディスカッション
- FAQ
- 規約一覧
- プロジェクト一覧
- 報道関係者向け資料
- イベント
- 報告会
- 経過報告

概要

2011年3月11日の東日本大震災発生時、ソーシャルメディアとマスメディアを通じて大量の情報が広がりました。正しい情報があれば、様々な噂やデマも飛び交い、情報がいかに重要かを改めて知る良い機会でもありました。では、こうした情報はいったいどのように伝えられたのでしょうか？ また、本当に伝えなかった情報がなぜ伝えられなかったのでしょうか？ 当時はデータを振り返った時に、本当に必要なサービスは何だったのか、次の災害に備えるために、今我々にはどんな準備をすればよいのでしょうか？

「東日本大震災ビッグデータワークショップ - Project 311 -」では震災発生から1週間の間に実際に発生したデータを参加者に提供いたします。参加者はそのデータを改めて分析することによって、今後起こりうる災害に備えて、どのようなことができるかを議論し、サービスを開発することができます。

提供データ

- 3月11日から1週間の朝日新聞記事 (提供: 株式会社朝日新聞社)
- Google Trends (提供: グーグル株式会社)
- 東日本大震災直後のテレビ放送テキスト要約データ (提供: JCC株式会社)
- 3月11日から1週間のツイート (提供: Twitter Japan 株式会社)
- NHK総合テレビ大震災発生直後から24時間の放送音声書き起こし及び頻出ワードランキング (提供: 日本放送協会)
- Honda インターナビ通行実績マップデータ (提供: 本田技研工業株式会社)
- レスキューナウの鉄道運行情報/緊急情報/被害状況のまとめ情報 (提供: 株式会社レスキューナウ)
- 混雑統計データ (提供: 株式会社ゼンリンデータコム)

<https://sites.google.com/site/prj311/>

東日本大震災時の誤情報の拡散・終息過程の可視化 (鍋島+, 2013)



鍋島 啓太, 渡邊 研斗, 水野 淳太, 岡崎 直観, 乾 健太郎. 2013. 訂正パターンに基づく誤情報の収集と拡散状況の分析. 自然言語処理, 20(3):461-484. (言語処理学会 2013年度論文賞)

参院選公示日前のツイートの分析

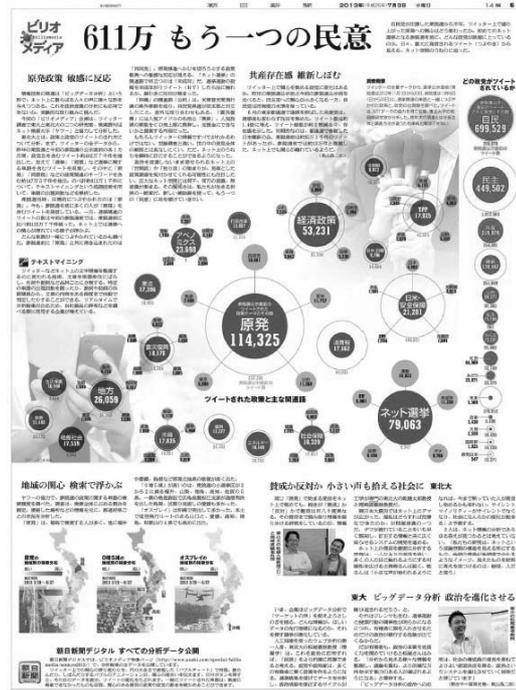
分析の狙い

- 政策・政党名でツイートを検索し、日ごとのツイート数、関連語を分析

分析方法

- ツイート総数が衆院選前と比べ3割減（参院選への関心低い）
- 共産党に関するツイートが増加
- 維新の会のツイートは激減
- ネット選挙に関連して、未成年のリツイート問題が話題に
- 原発を含むツイートが依然多い

2013年7月3日 朝日新聞朝刊（6面）「611万 もう一つの民意」



国会議事録からのキーワード抽出

発言分析（上位20／予測回数との差）

1	日本	(529.8)	11	世界	(179.4)
2	北朝鮮	(356.6)	12	問題	(169.7)
3	話	(345.2)	13	中	(163.2)
4	元気	(327.7)	14	ブラジル	(153.0)
5	辺	(313.0)	15	次	(132.3)
6	中国	(247.1)	16	質問	(126.3)
7	大統領	(207.5)	17	テレビ	(104.8)
8	アメリカ	(200.2)	18	国連	(95.2)
9	人	(195.3)	19	報道	(92.0)
10	ロシア	(183.6)	20	対話	(88.7)

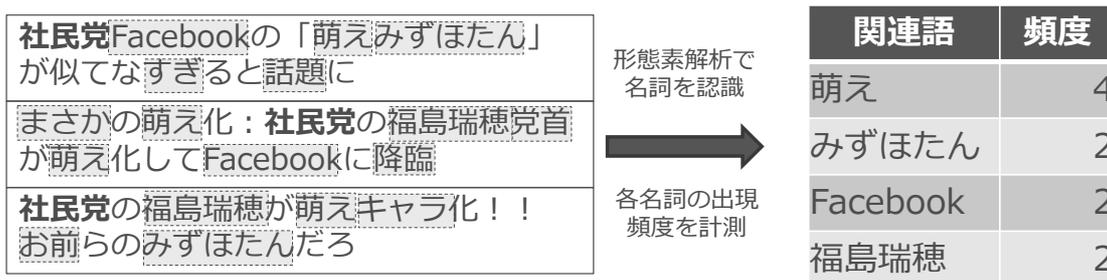
アントニオ猪木氏のキーワード

2019年7月2日 朝日新聞デジタル公開
<https://www.asahi.com/senkyo/senkyo2019/special/politipedia/>

ツイート分析の流れ



ツイート分析の流れ



関連語抽出の例（名詞の頻度計測）

12

報道において大切なポイント

- 分析結果の正確性に責任を持つ
 - 自然言語処理の意味解析の精度は低い
 - 分析方法に対して100%の精度を保証する必要がある
- 分析手法の透明性を確保する
 - 一般読者を想定し、分かりやすい分析・結果が望まれる
 - 「頻度」よりも性能のよい統計尺度（例えばTF*IDF）が使えない
 - 特定の団体・政党への利益・不利益にならないように、中立性を保つ
 - 報道だけでなく、様々な局面で手法の透明性の確保は重要
- 可視化も重要
 - 一般読者の目を引くようなデータの見せ方を考える

13

ロボットジャーナリズム（自動化報道）

14

記事の自動生成（AP通信, Automated Insights社）



Aug. 29, 2017

   [https://ww](https://www.apnews.com/32a904607c112a15524439fda93ede63/Bank-of-Nova-Scotia-reports-3Q-profit)

TORONTO (AP) — Bank of Nova Scotia on Tuesday reported fiscal third-quarter profit of \$2.1 billion Canadian.

The Toronto-based bank said it had earnings of \$1.66 per share Canadian.

The bank posted total revenue of \$6.89 billion in the period.

Bank of Nova Scotia shares have risen 11 percent since the beginning of the year. The stock has climbed 17 percent in the last 12 months.

Elements of this story were generated by Automated Insights (<http://automatedinsights.com/ap>) using data from Zacks Investment Research. Access a Zacks stock report on BNS at <https://www.zacks.com/ap/BNS>

この記事の要素はZacks Investment Researchから提供されたデータを用い、**Automated Insights**により生成された

<https://www.apnews.com/32a904607c112a15524439fda93ede63/Bank-of-Nova-Scotia-reports-3Q-profit>

15

深層ニューラルネットワークによる記事生成: RotoWireコーパス

NBA バasketボールの試合のスタッツとそのダイジェストの組が収録されているコーパス (約4,900事例)。ニューラル文章生成モデルの性能はBLEUスコアで10~15くらいで、実応用には程遠い

TEAM	WIN	LOSS	PTS	FG_PCT	RB	AS ...
Heat	11	12	103	49	47	27
Hawks	7	15	95	43	33	20

PLAYER	AS	RB	PT	FG	FGA	CITY ...
Tyler Johnson	5	2	27	8	16	Miami
Dwight Howard	4	17	23	9	11	Atlanta
Paul Millsap	2	9	21	8	12	Atlanta
Goran Dragic	4	2	21	8	17	Miami
Wayne Ellington	2	3	19	7	15	Miami
Dennis Schroder	7	4	17	8	15	Atlanta
Rodney McGruder	5	5	11	3	8	Miami
Thabo Sefolosa	5	5	10	5	11	Atlanta
Kyle Korver	5	3	9	3	9	Atlanta
...						

The Atlanta Hawks defeated the Miami Heat , 103 - 95 , at Philips Arena on Wednesday . Atlanta was in desperate need of a win and they were able to take care of a shorthanded Miami team here . Defense was key for the Hawks , as they held the Heat to 42 percent shooting and forced them to commit 16 turnovers . Atlanta also dominated in the paint , winning the rebounding battle , 47 - 34 , and outscoring them in the paint 58 - 26.The Hawks shot 49 percent from the field and assisted on 27 of their 43 made baskets . This was a near wire - to - wire win for the Hawks , as Miami held just one lead in the first five minutes . Miami (7 - 15) are as beat - up as anyone right now and it 's taking a toll on the heavily used starters . Hassan Whiteside really struggled in this game , as he amassed eight points , 12 rebounds and one blocks on 4 - of - 12 shooting ...

RotoWireデータセットのデータと記事の例 (Wiseman+ 2017)

Sam Wiseman, Stuart M. Shieber, and Alexander M. Rush. 2017. Challenges in data-to-document generation. In Proc. of EMNLP, pp.2243-2253.

16

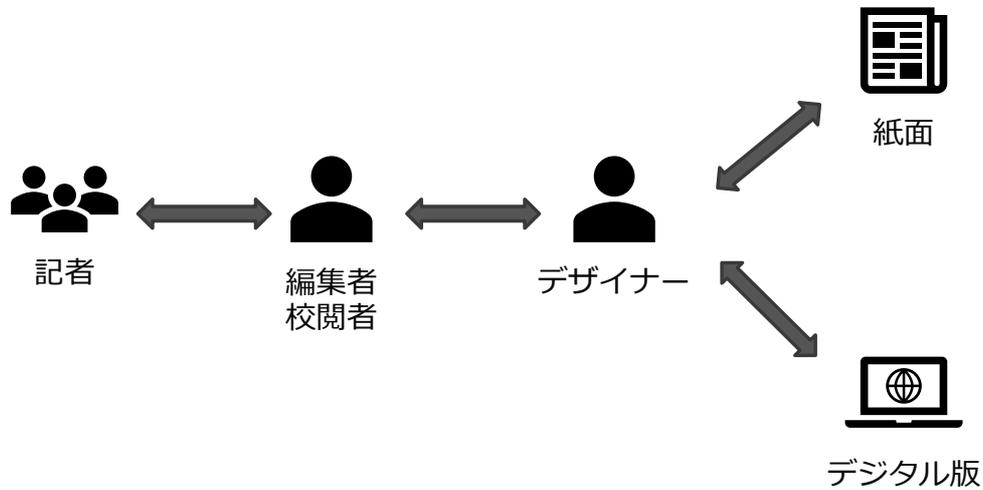
言語生成技術が報道の現場で使われるために (Leppänen+ 2017)

- **透明性** : 報道機関の内部・外部の人が、報道のプロセスを監視、確認、批判、および干渉できる
- **正確性** : 報道内容が客観的な事実に基づき、読者を誤解させない
- **修正可能性とドメイン適応性** : 新しいドメインの記事を生成するときに、システムを容易に修正できる
- **流暢さ** : 一貫性があって流暢な文章を生成できて、読者の満足度を損なわない
- **データの入手性** : 構造化されたデータベースが入手できる
- **ニュースの話題性** : 読者が興味を持つような話題のニュースを生成できる

Leo Leppänen, Myriam Munezero, Mark Granroth-Wilding, and Hannu Toivonen. 2017. Data-Driven News Generation for Automated Journalism. In Proc. of INLG, pp.188-197.

17

記事の自動校正：新聞記事ができるまで



18

記事の自動校正：編集前後の記事の活用 (Tamori+ 2017)

- 記事の編集操作の履歴データは残っていない
 - 記事をどのような編集操作で書き換えたのか分からない
- 類似度を用い、編集前後の記事の文の対応関係を自動的に分析

A 裁判員経験者の意見を生かそうと宇都宮地裁は12日、裁判員経験者と裁判官、検察官、弁護士による意見交換会を開いた。**B** 今年1～8月にかけて裁判員として審理に参加した9人が出席し、審理日程や評議についての感想や改善点などを話した。**C** 参加した裁判員経験者は「社会の中で責任を果たす自覚を持った」、「貴重な経験ができた」などという前向きな感想が多かった一方、「弁護士の冒頭陳述がわかりにくかった」などの課題も指摘された。…

編集・校閲前の記事

a 宇都宮地裁が裁判員経験者と裁判官、検察官、弁護士による意見交換会を開いた。**b** 裁判員経験者からは「社会の中で責任を果たす自覚を持った」、「貴重な経験ができた」などという前向きな声が出た一方、「弁護士の冒頭陳述がわかりにくかった」といった課題も指摘された。**c** 意見交換会は12日に開かれ、今年1～8月にかけて裁判員として審理に参加した9人が出席。**d** 審理日程や評議についての感想や改善点などを話した。…

最終版の記事

cとdの2文に分割

Hideaki Tamori, Yuta Hitomi, Naoaki Okazaki, and Kentaro Inui. 2017. Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment. In Proc. of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 46–50.

19

記事の自動校正：記事の編集操作の推定 (Tamori+ 2017)

- 編集距離と同様のアルゴリズムで編集操作を推定
- 2015年10月から2016年9月までの新聞記事の編集前後の記事から
 - 文アライメントを求めることで2,197,739文対を獲得
 - 1,108,750文対は編集なし, 1,088,989文対は何らかの編集あり

裁判員経験者の意見を生かそうと宇都宮地裁は12日、裁判員経験者と裁判官、検察官、弁護士による意見交換会を開いた。 (文 A から a への修正)

参加した裁判員経験者からは「社会の中で責任を果たす自覚を持った」、「貴重な経験ができた」などという前向きな声が出た感想が多かった一方、「弁護士の冒頭陳述がわかりにくかった」などのといった課題も指摘された。 (文 C から b への修正)

意見交換会は12日に開かれ、今年1～8月にかけて裁判員として審理に参加した9人が出席し、審理日程や評議についての感想や改善点などを話した。 (文 B から c,d への修正)

編集操作の推定

Hideaki Tamori, Yuta Hitomi, Naoaki Okazaki, and Kentaro Inui. 2017. Analyzing the Revision Logs of a Japanese Newspaper for Article Quality Assessment. In Proc. of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism, pages 46–50.

20

深層学習による記事の自動校正 (Hitomi+ 2017)

- 提案モデル
 - エンコーダ: 双方向LSTM
 - デコーダ: 文生成 (注意機構付きLSTM) と編集操作予測のマルチタスク

- 評価結果

評価対象	GLEU	WER	BLEU
すべての文対	70.68	23.90	74.48
編集された文対	68.63	35.55	65.31
変更されなかった文対	87.44	16.17	82.51

- 多くの個所は変更されずに残るので、BLEUスコアは高めの値となる
- 変更箇所に対する再現率は6.28%と低い

Yuta Hitomi, Hideaki Tamori, Naoaki Okazaki, and Kentaro Inui. 2017. Proofread Sentence Generation as Multi-Task Learning with Editing Operation Prediction. In Proc. of IJCNLP, pages 436–441.

21

見出し生成デモ

ウェブサービスとして
一般公開中

国内2例目のコロナウイルス感染=武漢の外国人男性-厚労省

サンプル記事



記事から自動生成

記事を入力する

厚生労働省は24日、国内で2例目の新型コロナウイルスによる肺炎の患者が確認されたと発表した。同省などによると、発症したのは旅行で訪れた武漢市在住の40代の外国人男性。東京都内の病院に入院し、快方に向かっているという。男性は14日から発熱があり、渡航前の15、17両日に医療機関を受診。その際は肺炎と診断されなかった。

多言語音声翻訳コンテストのオプションAPIとして提供中



TOP IDEA CONTEST PoC CONTEST NEWS FAQ INQUIRY NICT
トップ アイデアコンテスト 試作編コンテスト ニュース よくある質問 お問い合わせ

岡崎研究室 見出し生成API

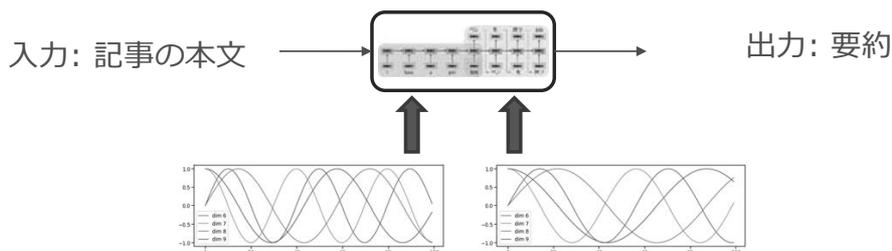
ニュース記事を入力として受け取り、見出しを生成して返します。

<https://headline.okazakilab.org/>

22

出力長を制御できる見出し生成モデル (Takase+ 2019)

- エンコーダ・デコーダによる自動要約
 - 翻訳元を記事本文、翻訳先を要約文とした「機械翻訳」
 - 日本語の新聞記事からの見出し生成の研究は少ない
- 機械翻訳と異なるのは出力長の制御
 - 機械翻訳では入力の情報をすべて翻訳したら終了
 - エンコーダ・デコーダに出力長を制御する機構が必要



Sho Takase and Naoaki Okazaki. 2019. Positional Encoding to Control Output Sequence Length. In Proc. of NAACL, pp. 3999-4004.

23

出力長制御の効果 (Takase+ 2019)

提案手法は出力長を正確に制御できる

(出力長を指定して) 要約を生成した際の長さの分散

手法	10文字	13文字	26文字
ベースライン (Transformer)	8.989	5.581	2.510
先行研究 (Kikuchi+ 16)	0.005	0.002	0.002
提案手法	0.000	0.000	0.000

要約の質も向上する

生成された要約と人間の要約の類似度 (ROUGE-L)

手法	10文字	13文字	26文字
ベースライン (Transformer)	31.31	39.12	36.18
先行研究 (Kikuchi+ 16)	35.76	38.47	37.54
提案手法	40.83	43.30	41.21

Sho Takase and Naoaki Okazaki. 2019. Positional Encoding to Control Output Sequence Length. In Proc. of NAACL, pp. 3999-4004.

24

記事から逸脱した見出しの削減

仮説: 本文から逸脱した見出しが生成されるのはタスクや訓練データの問題？

- 記事が見出しを含意しない記事が学習データに含まれているのではないか？

検証結果: 学習データにも問題あり

- 本文の一部だけを見出し生成器の入力として与えると、情報が不足する
- 入力文章に含まれない情報が見出しに現れる事例が全体の30-40%ある
- 含意関係認識器で学習データを精錬

記事が見出しを含意する割合 (%)

データセット	先頭1文	先頭3文	全文
Gigaword	70.3	—	92.8
JAMUL	—	61.4	94.2

実験結果: 学習データから無理な事例を取り除くと見出しの忠実性が改善した

- ROUGEの向上は確認できないが、記事と見出しの含意率は向上 (自動・人手評価)

学習データ	ROUGE-1	ROUGE-2	ROUGE-L	含意率	人手評価
無理な事例を除去しない	48.08	22.21	40.02	90.29%	89.91%
無理な事例を除去する	45.62	20.55	38.10	92.26%	92.66%

松丸 和樹, 高瀬 翔, 岡崎 直観. 含意関係に基づく見出し生成タスクの見直し. 第240回自然言語処理研究会, 2019年6月.
松丸 和樹, 高瀬 翔, 岡崎 直観. 見出し生成の忠実性の改善. 言語処理学会第26回年次大会 (NLP2020), 2020年3月.

26

自然言語処理によるニュース記事執筆の課題

- 日々生み出されている貴重なデータを保存していない
 - 記事の編集履歴を保存していない
 - 記事と短文要約との対応関係を保存していない
 - ……
- 人材不足
 - 新聞社の売上高は減少傾向が続いている
 - データがあっても分析・活用できる人材が社内にはいない
- 正確性に関する高いハードル
 - 新聞記事の生成を完全自動にすることは困難
 - 記者や編集者の作業を軽減できる性能・仕組みが必要

27

おわりに

謝辞

- 東京工業大学: 高瀬翔氏, 松丸和樹氏
- 朝日新聞社: 田森秀明氏, 人見雄太氏, 田口雄哉氏, 浦川通氏
- 時事通信社: 朝賀英裕氏, 川上貴之氏

参考文献

- 岡崎 直観. データジャーナリズムとデータ科学. 電子情報通信学会誌, 99(4):339-346, 2016年.
- 岡崎 直観. ロボットジャーナリズムの現状と課題. 映像情報メディア学会誌, 72(2):70-75, 2018年.

解説

データジャーナリズムとデータ科学

Data Journalism and Data Science

岡崎直観

Topics
話題

ロボットジャーナリズムの現状と課題

岡崎直観

岡崎直観'

28

不許複製・禁無断転載

主催者	高度言語情報融合フォーラム (ALAGIN) 一般財団法人 日本特許情報機構 (Japio)
発行日	令和2年3月6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 E-mail: info@alagin.jp TEL: 03-3351-8151

