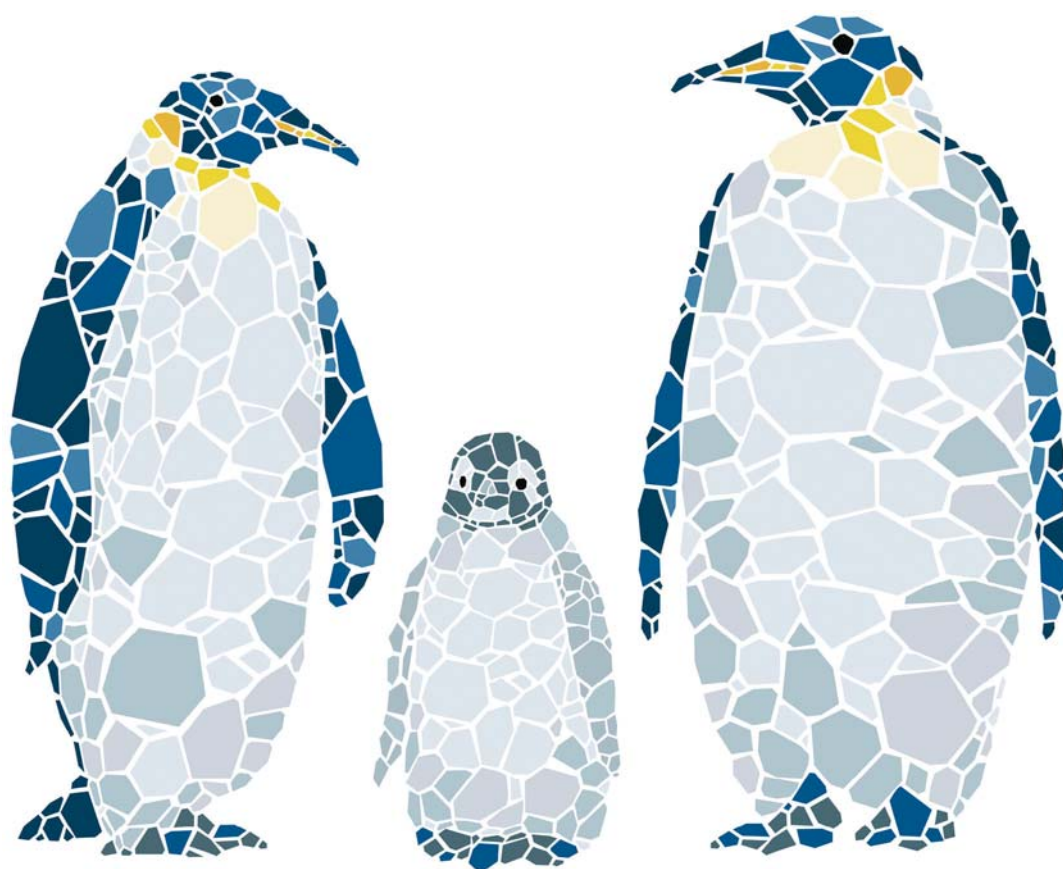


ASCA Bulletin

August 2018
vol. **20**



- プロフェッショナルに聞く

機械翻訳の実用化と翻訳者の役割

- 科学誌 *Science* 日本人著者研究紹介：
CRMP2に結合する化合物 edonerpic maleate は脳損傷による運動機能障害の回復を促進する
- ASCA Letter：
遺伝研で、花見をしながら遺伝学を体験“心に咲いたサクラは見つかりましたか？”

機械翻訳の実用化と翻訳者の役割

すみ た えい ちろう

隅田 英一郎 氏

情報通信研究機構 (NICT) フェロー
アジア太平洋機械翻訳協会 (AAMT) 会長
日本翻訳連盟 (JTF) 理事

近年の技術的進歩を背景に、機械翻訳を商業ベースで実用化する取り組みが急速に進んでいます。今回は公的研究機関で機械翻訳の研究に長年携わるとともに、日本翻訳連盟の理事として機械翻訳の普及推進にも尽力されている隅田英一郎先生に、最近の成果や課題、また翻訳者の役割についてお話を伺いました。



機械翻訳の商用化

深層学習を取り入れたニューラル機械翻訳 (NMT) の登場によって、機械翻訳の精度は大きく向上しました。私が所属する情報通信研究機構 (NICT) では、このNMTのアルゴリズムを改善する研究を推進すると同時に、機械翻訳をいかに社会で活用していくかという枠組みづくりにも取り組んでいます。そのためには、機械翻訳を企業が商用ベースで導入し、産業構造に組み込んでいくことが不可欠だと考えています。

最近の事例として、この7月、私たちは東芝デジタルソリューションズ株式会社と共同で、特許庁の外部向けサービスの一環として、特許文書専用の高精度翻訳システムを構築し、2019年5月から稼働を開始することを発表しました (図1)。産業のグローバル化によって特許翻訳の需要は増大しているのですが、人手不足や高コストといった問題を解決するために機械翻訳によるソリューションが期待されていました。

このシステムの特徴は、NMTだけでなく、他のタイプの機械翻訳をも使い分けていることです。NMTは非常に流暢な翻訳文を出力できるのですが、訳抜けや湧き出しが起りやすいという問題があります。この弱点をカバーするために、特許文書を部分毎にタイプ分けし、NMT、ルールベース機械翻訳、統計的機械翻訳の中から適したエンジンに自動的に振り分けるようにしています。さらに、事前に文章を整形したり、対訳辞書を充実させたりなどもしています。

今回のシステムが選ばれた理由の一つにセキュリティがあります。特許文書には公開済みの公知情報と、出願中の秘密情報が存在します。後者の情報が漏洩することになれば、出願者がひいては日本の産業が大きなダメージを受けます。そのため本システムではデータセンターを国内に限定し、物理的なデータの持ち出しを防止しています。さらに前述のとおりソフトウェア部分も、NICTがNMTのエンジンを提供し、東芝デジタルソリューションズが言語処理のフレームワークを開発するという座組で、純国産技術で構築されます。

図1



特許庁の「機械翻訳システム」の概要

(引用: <http://www.nict.go.jp/press/2018/07/10-1.html>)

医薬分野での機械翻訳の導入

NMTのエンジンを構築するには、対訳データ (コーパスとも呼ばれる) が大量に必要です。現代の機械翻訳のアルゴリズムは、NMTが登場しただけでなく、更なる改良もどんどん進んでいます。コーパスの存在が前提条件であり、これがなければ高精度翻訳は期待できません。特許文書で高精度NMTシステムが稼働できるようになったのも、特許庁との連携で

数億文の対訳データを用意できたためです。

コーパスの収集は容易なことではありません。特に製薬分野では、各製薬企業が大半の文書のデータを保有しており、個別にアプローチを試みてもその意義を理解してもらうことは難しく、データの収集は難航しました。そこで私たちは、対訳データの集積のために「翻訳バンク」と呼ばれるプロジェクトを立ち上げ、総務省とともに運用しています。これは、様々な企業が保有している文書の多言語データを収集し、多様な分野で精度の高い機械翻訳エンジンを構築し公共の利用に資することを旨とする取り組みです (図2)。

図2



高精度翻訳の分野

(引用: <http://h-bank.nict.go.jp/about.html>)

その結果、製薬分野では今年になって2つの企業との提携に成功しました。まずはアストラゼネカ株式会社との間で機械翻訳システムの共同研究を実施することとなり、同社のデータを用いて翻訳システムの実用性を検討する研究がスタートしています。医薬分野に特化した精度の高い機械翻訳システムが開発されることで、翻訳作業が効率化し、新薬をはじめとする薬事申請をより迅速にできることが期待されます。

もうひとつは6月にMSD株式会社より、医学事典「MSDマニュアル」の多言語データの提供を受けることとなり、NICTの旅行用音声翻訳アプリ「VoiceTra」に早速実装する運びとなりました。旅先などでアプリを使用する一般ユーザーに恩恵をもたらすこととなります。

これら2つの事例は、医薬分野でのNMTを発展させる上で翻訳バンクにとって非常に有意義な取り組みです。同時に協力体制を取ることとなった両者にとってもメリットのあることだと考えます。その最も大きな理由は、ビジネスにおいて機械翻訳を活用するには、機械翻訳がどのようなシチュエーションで有効なのか、またはそうでないかを検証する必要があり、それは各企業が自らの手で行わなければならないからです。機械翻訳は、導入すれば誰にでも使いこなせることができ、すぐに役に立つような単純なシステムではありません。自社の課題がどこにあるのか、例えばプロセスの迅速化なのか、多言語ドキュメントの品質を上げたいのか、まずはそれを把握した上で、ニーズに合わせた評価を行う必要があります。

ポストエディットの意義

機械翻訳の導入によって、翻訳者をはじめとする人手での作業が不要になるかという、まったくそうではありません。NMTを含めた機械翻訳は、原文の意味を理解する用途には非常に有効ですが、そうでなく、パブリケーションレベルの文章を生成するにはまだ不十分な技術です。品質保証の観点から、人間が確認するプロセスを省略することはできないでしょう。その意味で、ポストエディターの役割は非常に重要となっており、こうした人材に対する需要は増すばかりでしょう。逆に、機械翻訳の普及は、どれだけこうした人材を育成できるかにかかっているととも言えます。

NMTの登場によって、ポストエディットという仕事の持つ意味は大きく変わりました。1つ目は、これまでの統計的機械翻訳とは異なり、NMTのポストエディットでは作業量が大幅に減るだろうということです。統計的機械翻訳で出力される文章はぎこちないものでしたから全面的な修正が必要で、ポスト

エディットは普通に翻訳するより、かえって手間がかかるという問題がありました。それが、流暢な翻訳を生成するNMTであればその負荷が改善されるだろうということです。2つ目はその逆の側面で、NMTが流暢な訳を出すことによって誤訳がむしろ見つけにくくなり、その検出にスキルが求められるようになったということです。一見問題なさそうな訳文から素早く誤りを見つけて適切に修正するには、言語と専門性の両方を兼ね備えている必要があると考えています。

したがって、ポストエディットは大変で退屈な嫌われがちな「お仕事」ではなくなっており、翻訳者が効率性を意識しながら、自分の経験や知識を動員して行う高度な知的行為に変わったといえます。このワークスタイルの転換には、情報提供や教育システムの整備など、業界からの後押しも必要です。日本翻訳連盟などの業界団体や翻訳会社が翻訳者を支え、翻訳という仕事の新しいモデルを作り上げることを期待しています。



すみ た えい ちろう
隅田英一郎 氏

30年以上翻訳に携わる技術屋。IBM、ATR、NICTの研究所を渡り歩きつー貫して機械翻訳に関わり、規則、用例、統計、ニューラルネットの全技術を熟知。音声翻訳アプリVoiceTra、テキスト翻訳サイトTexTraを公開して、「対訳データ」と「高度なアルゴリズム」によって、本質的に翻訳が困難な日本語と英語の間でさえも高精度機械翻訳が実現可能であることを証明してきた。機械翻訳への熱い期待を間違いなく結実させるために注力中。現在、2020年をゴールとする音声翻訳の国家プロジェクト「グローバルコミュニケーション計画」を推進しながら、翻訳イノベーションの種蒔き・草抜き・水やり等のお世話に精を出している(^^)。

インタビューを終えて

隅田先生から、機械翻訳は自動洗濯機と同じだとお話しいただいて妙に納得しました。洗剤の入れ方、種類、洗濯物の選別の仕方によって、洗い上がりは違うし、洗濯機から出した後の干し方、たたみ方によって最終の成果物は大きく異なる、と。もちろん大切な衣服はクリーニング屋さんに出すし、物によっては時間や費用をいとわないはず。自分が洗濯するより結果的には安くきれいに上がるなら、遠慮なくクリーニング屋さんを使うはずだとも。洗濯機の性能がいくら上がっても、クリーニング屋さんはなくなりません。私たちは、自動洗濯機をうまく使いこなし、最強のクリーニングサービスを提供したい、と再認識した次第です。



VoiceTra 紹介!

言葉の壁で困らない社会実現を目指してNICTが開発した多言語音声翻訳アプリです。AI技術を用いて、観光地での旅行会話はもとより、病院、商業施設といったさまざまなシーンで精度の高い音声翻訳を無料で提供しています。世界31の言語に対応しています。
<http://voicetra.nict.go.jp/>





科学誌 Science に論文が掲載された日本人著者に、論文の内容、普段の研究内容についてわかりやすく解説していただきました。

脳卒中後も
麻痺のないカラダに戻る!?

CRMP2 に結合する化合物 edonerpic maleate は脳損傷による運動機能障害の回復を促進する

CRMP2-binding compound, edonerpic maleate, accelerates motor function recovery from brain damage

Science 6 Apr 2018: Vol. 360, Issue 6384



横浜市立大学大学院
医学研究科 生理学
高橋 琢哉 先生

脳卒中をはじめとした脳損傷後のリハビリテーションの効果を劇的に促進する化合物の特定に成功しました。最も重要なことはこの薬剤の治験を成功させ、脳卒中後の麻痺に苦しむ患者さんの元に一刻も早く届けることだと思います。

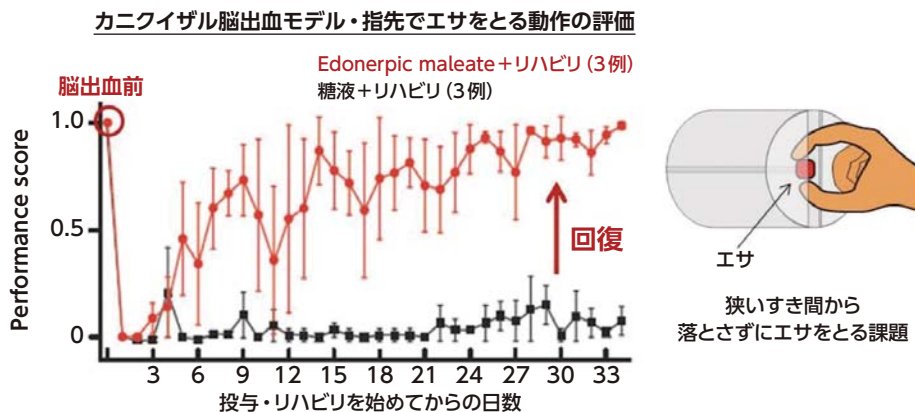
脳卒中は全世界で年間1,700万人、本邦で約30万人が発症しています。脳卒中急性期の治療はかなり進歩し、お亡くなりになる患者さんの数は減少している一方で、しばしば重篤な麻痺を引き起こし、患者さんの生活の質を大きく低下させます。現在の回復期における運動機能の回復を目的とした治療は、地道なトレーニングによるリハビリテーションが主体となっています。しかしながらその効果は限定的であり、より効果的な治療法が望まれています。特に、回復期リハビリテーションに介入する効果的な薬剤は残念ながらありません。

運動機能回復のメカニズムには、リハビリテーションなどの外部からの刺激にตอบสนองした脳の変化(脳の可塑性)が関与していることが知られています。生体における記憶学習といった可塑的变化に伴ってシナプス応答の増強が見られるとき、神経伝達物質であるグルタミン酸の受容体のひとつであるAMPA受容体がシナプスの膜上で増加することが明らかにされています。

本研究では、富士フィルムグループの富山化学工業株式会社、産業技術総合研究所、医薬基盤・健康・栄養研究所との共同研究により、脳卒中後のリハビリテーション効果を大きく促進する新薬の候補化合物(edonerpic maleate)を特定しました。我々は本化合物がCRMP2(collapsing-response-mediator protein 2)に結合することによりAMPA受容体シナプス移行を促進し、脳損傷後の運動機能回復をトレーニング依存的に劇的に促進することを、げっ歯類、カニクイザルを用いて明らかにしました(図)。興味深い点は、本薬剤がトレーニング依存的に効果を発揮するという点です。このようなコンセプトの薬剤はこれまでになく、入力依存的にシナプスに変化がおきる「シナプス可塑性」の概念を体現した極めてユニークな薬剤です。また、特筆すべきは、霊長類の指先の巧緻性の機能回復において最も顕著な回復効果が観察されたことです。カニクイザルの指先の巧緻性はヒトに非常に近く、この機能は脳卒中患者さんの生活の質に直結する非常に重要なものです。

現在臨床治験の準備を進めており、脳卒中医療の革新につながるものと期待されます。

図: カニクイザル脳出血モデルにおいて edonerpic maleate はリハビリテーション依存的に指の精密な運動機能を回復させる



Performance score: 課題の成功率をもとにした運動機能回復の指標(脳出血前の水準を1.0, 脳出血後で投与・リハビリ前の水準を0.0とした相対値)

ASCA x Science

ASCAは、Scienceを発行する米国科学振興協会(AAAS)からの委託により、翻訳、広告代理店業務、カスタマーサービスなど多岐にわたるサービスを提供しています。

ASCAが翻訳しているサイエンス日本語ホームページはこちら! www.sciencemag.jp

Science, Science Signaling, Science Translational Medicineに掲載された最新の研究論文を日本語タイトルから簡単にご確認いただけます。



ASCA Academy

アカデミー特別編!

遺伝研で、花見をしながら遺伝学を体験

“心に咲いたサクラは見つかりましたか?”

4月のASCA Academyでは、静岡県三島市にある国立遺伝学研究所を訪問しました。

遺伝学を中心とした生命科学の先端研究を行っている機関で、大腸菌からヒトまで、分子レベルから生物集団レベルまで、理論から実験まで、遺伝学に関わる幅広い分野の研究が行われています。

所内には研究用の桜が約200種類植えられており、毎年桜の季節に合わせて研究所の一般公開が行われています。当日は研究者の講演会や研究体験に参加することができ、桜を楽しみながら、遺伝学の研究に触れることができます。

今年は桜の開花が例年より早かったため、お花見は難しいかも…とと思っていましたが、当日は遅咲きの種類の桜を楽しむことができました。



参加しました!

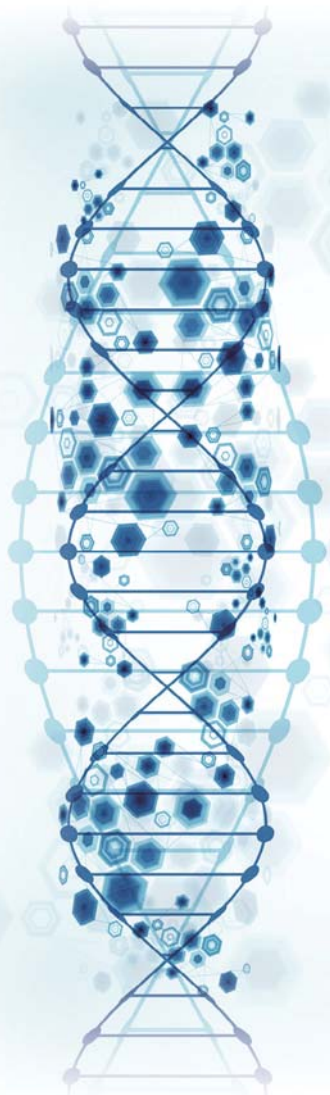
遺伝学とは、生命の根源、生命現象の仕組みを遺伝情報という観点から解明する学問だそうで、研究所の体験コーナーでは様々な生物のメカニズムを見ることができました。

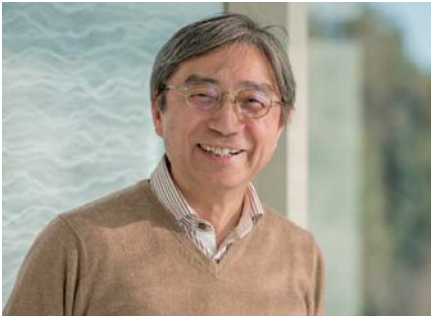
ヒトのもととなる精子と卵子がどう形成され、受精卵がどのようにできるのか? 菌は体の中でどのように働くのか? ショウジョウバエとヒトの遺伝子の違いは? ヒトが死んでもなお生き続けるがん細胞の仕組みとは? など。

ヒトひとりの中にも細胞や菌レベルで様々な生物が存在して作用しあっている様子を、実際に顕微鏡を覗きながら研究者の方々に解説していただくことができました。まるで学校の理科の実験に参加したようで、とてもワクワクしました。

今回のアカデミーを通して、遺伝は単に「親から受け継いだもの」だけではないことに気づかされました。生命と名の付くあらゆるものの根源を知ること、その生命一つ一つの長年のつながりを解明することであり、人間の枠を超えた広がりのある世界なのだと、とても奥深く壮大なものを感じることが。

まさに、今回のイベントのメッセージの通り、
“心に咲いたサクラが見つかりました”。





国立遺伝学研究所 副所長・教授

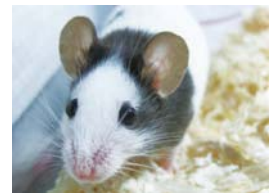
城石 俊彦

当日、「日本産マウスのはるかな旅路」というタイトルでご講演下さった国立遺伝学研究所教授の城石俊彦先生にお話を伺いました。

先生の講演から、江戸時代に日本の外へ持ち出されたマウスが、巡り巡って今日世界中の実験で大活躍している実験用マウスの起源にかかわったことが分かり、遺伝学のパワーや歴史の不思議さを感じることができました。

“野生マウスには大きく分けて3種類のグループが存在する”

野生マウスの遺伝子を解析すると、大きく3種類（西ヨーロッパ、ユーラシア大陸北部、東南アジアに分布）に分類されるようで、日本産野生マウスは大陸と東南アジアの雑種だそうです。種類が異なる場合、兄妹交配を続けると生殖能力が減少し、やがて子供が生まれなくなるものの、同一種類内であれば兄妹交配を20回以上行うことができるらしいです。今使われている実験用マウスは、西ヨーロッパと日本産マウスの遺伝子が混じっているとのことで、この日本産マウスの遺伝子は江戸時代に海外に持ち出されたパンダマウス（JF1/Ms系統）の祖先由来だったそうです。



“遺伝的背景を考慮すべき”

実験で使われているマウスにもさまざまな系統があり、遺伝的背景は全く異なります。なので、同じ変異遺伝子でもどの系統の中に存在するかによって症状の出方が異なるそうです。パンダマウスでは、エンドセリン・レセプター B 遺伝子の piebald 変異がホモとなっているため、白黒がち（白斑）が生じます。しかし、別の系統でこの変異遺伝子をホモで持つと生存できなくなります。系統によって性格や行動パターン、病気のなりやすさなども異なります。系統の違いがパンダマウスの表現型（性質や症状）に影響を与えるのです。



“実験用マウスのゲノム構成を知った上で実験デザインを考える必要がある”

現在、2万以上ある遺伝子をそれぞれノックアウトした2万系統のミュータントマウスを作製し、その表現型を細かく調べようという国際プロジェクトが進んでいます。先ほど述べたように、死なずに白斑を生じるのは日本産マウス（JF1/Ms）だけ。系統のバリエーションによって表現型が変わるので、遺伝子の機能のある特定の系統の中だけで議論することは危険とのことです。

人間においても同様で、同じ遺伝子の病気でも、他にどのような遺伝子のバリエーションを持っているかによって症状が変わります。中には非常に強い変異もあり、背景にかかわらず症状が出てしまうこともあるのですが、多くは、他の遺伝子のバリエーションによる影響が考えられるそうです。体質や性格なども、1つの遺伝子によってではなく、複数の遺伝子のバリエーションによって決定されるとのことです。

「マウス1系統で遺伝子を1つずつ調べるだけでも大変な労力と費用がかかります。それでもまずは最初のステップとして取り組み、将来は他の遺伝子との関係でどう表現型が変わっていくのか研究することが重要です」とお話し下さいました。

城石先生のご講演では美術のお話を交えながら、とても生き生きとユーモアあふれる口調でパンダマウスができるまでの歴史を分かりやすく解説してくださいました。また取材を通し、改めて、遺伝学の面白さ、重要さを認識した次第です。

詳しくは https://www.youtube.com/watch?v=o8xpc_Tieq8（講演の動画）をぜひご覧下さい。

終わりに

遺伝学というあまり馴染みのない学問について、先生にわかりやすく説明していただけた貴重な機会でした。

医療翻訳に携わる者として、今後も科学への関心・興味を持ち続け、よりお客様のニーズに応えられるよう精進していきたいと思いました。

引き続きAcademyの活動にも乞うご期待ください！