

欧州における自動翻訳・音声認識・音声合成技術に係る
研究開発の動向（概要）

平成 28 年 1 月

国立研究開発法人 情報通信研究機構
(欧州連携センター)

第一部 欧州連合の自動翻訳・音声認識・音声合成技術の研究開発助成の基本的方針

欧州連合の第七次枠組計画及びホライゾン2020における自動翻訳及び音声技術研究開発への助成動向

- ・ 特定の言語（英語や仏語等）の自動翻訳技術の開発は進んでおり、翻訳の品質が高くなってきているものの、幾つかの欧州言語に関しては言語技術のサポートが進んでいないことが指摘されており、ホライゾン2020では、EUの全言語間の組み合わせの翻訳品質を向上させることが特に目標とされている。
- ・ 多数の言語が利用されている欧州では、言葉の壁がEUデジタル単一市場形成の障害となっているという認識がある一方で、EUの言語技術への研究開発助成の基本的方針に若干の変更が見られる。ホライゾン2020の2014-2015年度ICT作業プログラムでは、「ICT-17: 言語の壁の打破」という枠組みで、自動翻訳技術の開発支援に対して、1500万ユーロの予算が組まれているものの、2016-2017年度ICT作業プログラムでは「ICT-19: メディアとコンテンツの融合」という枠組みで、メディアとコンテンツにアクセスすることを可能にする技術として、音声転写技術や自動翻訳技術が助成される見込みであり、言語技術へ焦点は置かれていない。このような研究開発助成の基本的方針の変更に対しては、欧州の言語技術分野の研究者から不満の声が上がっている。

欧州連合の機械翻訳サービスMT@ECの動向

- ・ MT@ECは、ISAプログラムの下で開発されたEU諸機関及びEU加盟国の行政機関向けのオンライン共通機械翻訳サービスである。
- ・ MT@ECは、EU関連の文書を最も正確に翻訳する。
- ・ MT@ECは、EU全公用語間での翻訳を直訳で提供する。したがって、より正確な翻訳が必要な場合は、翻訳の専門家が文書を修正する必要がある。
- ・ MT@EC以前は、ECMTという自動翻訳サービスが利用されていた(2010年12月まで)。ECMTでは、規則ベースの機械翻訳技術が利用されていたが、MT@ECでは統計ベースの技術が使われている。
- ・ 欧州のインフラストラクチャの接続を支援するEUの「CEF (Connecting European Facility)」というプログラムでは、公共機関向けの「CEFAT (CEF Automated Translation)」という機械翻訳プラットフォームも開発も支援している。CEFATはMT@EC上に構築される予定であり、EUの24カ国語及びノルウェー語、アイスランド語の全てのペア(両方向)の翻訳を対象とする。

第二部 欧州における自動翻訳・音声認識・音声合成技術に係る研究開発機関の概要と研究開発の動向

英国・エジンバラ大学統計機械翻訳グループ

- ・ エジンバラ大学統計機械翻訳グループは、機械翻訳システム・MOSESの開発で有名である。
- ・ 同研究グループは、数多くのFP7プロジェクト及びホライゾン2020の研究プロジェクトに参加している。
 - FP7プロジェクトの例: Accept, Casmacat, EU-Bridge, Matecat, MosesCore, EuroMatrixPlus, LetsMT!
 - ホライゾン2020の例: QT21, MMT, HimL, TraMOOC, Cracker

アイルランド・ADAPT (旧CNGL)

- ・ アイルランドは、英国と同じく、英語使用国であり、また法人税を軽減していることから、同国に多くの外国企業が欧州本拠地を置いているので、同国では自動翻訳を含めたローカライゼーションの作業が積極的に実施されている。このような社会的背景の中、2007年にCNGL (Centre for Next Generation Localisation) が設立された。
- ・ 2015年1月から、組織名称をCNGLからADAPTに変更した。
- ・ ADAPTは、4つの研究機関、トリニティ・カレッジ・ダブリン、ダブリンシティ大学、ダブリンカレッジ大学、ダブリン技術研究院と産業パートナーの研究者を含む研究センターである。
- ・ ADAPTでは、自動翻訳だけでなく、金融サービス、eコマース、メディア、エンターテインメント・ゲーム、デジタル文化・人文科学、eラーニング等の開発も実施されている。
- ・ ADAPTから、KantantMT.com社とICONIC社という自動翻訳技術関連のスピンオフ企業が2つ設立されている。

ドイツ・カールスルーエ技術研究院インタラクティブシステム研究所 (KIT)

- ・ カールスルーエ技術研究院インタラクティブシステム研究所は、音声及びテキスト機械翻訳、顔追跡、注意追跡、マルチモーダル追跡、ニューラルネットワークの研究開発を行っている。
- ・ 2012年に、同研究所は、大学講義向けの自動同時音声翻訳システムを発表している。
- ・ テキスト機械翻訳技術に関しては、特に欧州言語に焦点を当てて、研究を実施しており、世界的な機械翻訳評価ワークショップであるWMTに数年来参加している。
- ・ テキスト機械翻訳技術の開発を独仏Quaeroプロジェクト、FP7のEU-Bridgeプロジェクト、ホライゾン2020のQT21プロジェクトで進めている。

ドイツ人工知能研究センター言語技術研究所 (DFKI)

- ・ ドイツ人工知能研究センターでは、機械翻訳を含む言語技術の研究開発を言語技術研究所のテキスト分析グループ (ベルリン市) とマルチ言語技術グループ (ザールブリュッケン市) で実施している。
- ・ ドイツ人工知能研究センター言語技術研究所は、FP7のQTLeapプロジェクト、ホライゾン2020のQT21プロジェクトを実施している。
- ・ ドイツ人工知能研究センター言語技術研究所は、言語技術分野の欧州研究機関の提携を強化するMETA-NETのコーディネーターを務めており、欧州で同分野の中心的な研究組織の一つである。また、同研究所は、機械翻訳技術分野の欧州研究機関の提携を強化するホライゾン2020CRACKERプロジェクトのコーディネーターを務めている。
- ・ Acrolinx (言語分析エンジンの提供) とYocoy (機械翻訳モバイルアプリケーションの開発) というスピンオフ企業が設立されている。

フランス・メヌ大学情報学研究所 (LIUM)

- ・ フランスのル・マン市に設立されたメヌ大学情報学研究所 (LIUM) では、音声言語翻訳、ロバスト音声認識、話者ダイアライゼーション、話者認識、話者言語理解の研究が実施されており、これら全てのトピックで、特に深層ニューラルネットワークの利用について研究している。
- ・ LIUMから、Deeplingo社 (企業向け機械翻訳のモデルの提供)、Voxolab社 (企業向け話者特定アプリケーションや音声転写アプリケーションの提供) というスピンオフ企業2社が設立されている。

- ・ LIUMは、EUのCHIST-ERAの枠組みで、M2CRプロジェクト（2015年～2017年）を開始したところである。同プロジェクトでは、深層ニューラルネットワークを利用した音声・画像理解、音声機械翻訳の研究を行う。その他、LIUMはFP7のMateCatプロジェクト（2011年～2014年）で、コンピュータ翻訳支援ツール（CAT）の開発を行った。